Magnitude estimation reveals Poisson-like noise underlying perception

Cristina Rodriguez-Arribas[1,*], Joan López-Moliner[1], Daniel Linares[1,2,*]

[1] Vision and Control of Action Group, Departament de Cognició, Desenvolupament i Psicologia de l'Educació, Institut de Neurociències, Universitat de Barcelona, Barcelona, Spain.
[2] Serra-Húnter Fellow
* Corresponding authors: cristinarodriguezarribas@gmail.com, linares@ub.edu

Abstract

A fundamental goal of sensory neuroscience is to understand how stimuli map into perceptual experience. Insights into this process traditionally rely on measurements of discrimination sensitivity, but these cannot identify the internal noise distribution underlying perception because they conflate it with transducer gain. Here, using visual contrast, we show that rating variability from magnitude estimation provides access to this noise. The analysis of this variability alongside mean responses reveals an internal representation comprising a sigmoidal transducer and Poisson-like noise that quantitatively predicts discrimination sensitivity without free parameters. These predictions include the classic pedestal effect and Weber-like behavior, which now we can attribute to the transducer nonlinearities and signal-dependent noise, respectively. We empirically establish that a single stochastic representation underlies subjective ratings and objective discrimination.

Introduction

Understanding how the brain transforms sensory information from physical stimuli into perceptual experiences is a fundamental question in neuroscience. Key insights into the core principles of this process come from behavioral experiments measuring sensitivity $D(x)$, defined as the capacity to discriminate between a baseline stimulus intensity, x, and that baseline plus an increment, $x + \Delta x$. One hallmark of this process is Weber-like behavior, in which sensitivity decreases over most of the perceivable intensity range as intensity increases following a function $D(x) \propto x^w$, with w typically close to −1 (Fechner, 1860; Gescheider, 1997). Yet, at very low intensities, Weber-like behavior breaks down and sensitivity instead increases with intensity, a phenomenon known as the pedestal effect (Arabzadeh et al., 2008; Green, 1960; Hanna et al., 1986; Legge & Foley, 1980; Nachmias & Sansbury, 1974). Consequently, sensitivity is best characterized by a non-linear, inverted U-shaped function that peaks at very low intensities (Arabzadeh et al., 2008; Hanna et al., 1986; Legge & Foley, 1980; Nachmias & Sansbury, 1974).

Sensitivity measurements alone, however, are insufficient to fully constrain the underlying mechanisms of sensory processing (Zhou et al., 2024). Let the perceptual response R to a stimulus of intensity x be modeled by a probability density function $f(r, x)$, which constitutes the internal representation of the stimulus. The mean of the distribution, $\mu(x)$, defines the transducer function (the mapping of physical intensity to mean internal response), and the variance, $\sigma^2(x)$, reflects the internal noise. The difficulty in identifying the internal representation arises because the sensitivity depends only on the signal-to-noise ratio, the slope of the transducer divided by the noise (Eq. 1) (Seriès et al., 2009; Zhou et al., 2024).

$$D(x) = \frac{|\mu'(x)|}{\sigma(x)} \quad (1)$$

Consequently, any combination of transducer gain and noise level that preserves this ratio yields an identical empirical sensitivity.

Recently, two approaches have begun to overcome this non-identifiable problem by complementing sensitivity data with additional behavioral metrics. The first approach, from Pardo-Vazquez et al. (2019), showed that sensitivity and reaction times could be jointly and parsimoniously explained with a bounded accumulation model that incorporates a power function transducer and Poisson noise ($\sigma(x) = \mu(x)^{0.5}$).

The second approach, introduced by Zhou et al. (2024), addressed the signal-to-noise ambiguity by combining sensitivity measurements and subjective ratings from previous studies. The subjective ratings were derived from magnitude estimation (ME), a task in which participants assign numerical values to match stimulus intensities. They showed that these two types of judgments could be compatible assuming an internal representation characterized by modulated Poisson noise, $\sigma(x) = \sqrt{\mu(x) + g^2\mu(x)^2}$ (Goris et al., 2014). This noise structure transitions from a Poisson-like regime at low intensities ($\sigma(x) = \mu(x)^{0.5}$), to a multiplicative regime at higher intensities ($\sigma(x) = g\mu(x)$).

While both approaches converge on the presence of Poisson-like noise, direct measurements of the internal noise distribution remain limited. In this context, the framework proposed by Zhou et al. (2024) offers a particularly compelling starting point. Their approach relies on the assumption of direct psychophysics that ME responses are proportional to perceived intensity (Gescheider, 1997; Stevens, 1957), proposing that the mean response reflects the transducer function $\mu(x)$. A natural extension of this framework is that the variability of ME responses—a quantity that has received little attention (Green & Luce, 1974)—might directly reflect the internal noise $\sigma(x)$. This raises the possibility that ME data could fully constrain the probability density function of the internal representation. If valid, this hypothesis implies that discrimination sensitivity, including hallmark phenomena such as Weber-like behavior and the pedestal effect, should be predictable based solely on ME responses.

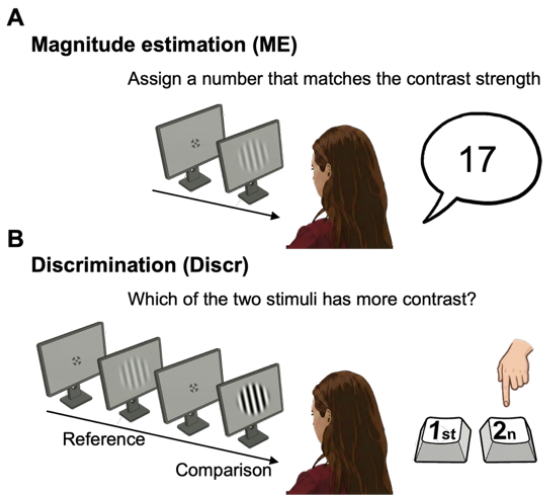

Fig. 1. Illustration of the tasks. *(**A**) In the magnitude estimation task (ME), participants rate stimuli contrasts using unconstrained verbal numerical reports. (**B**) In the discrimination task, participants compare references and comparison stimuli with both their presentation order and contrast values randomized across trials.*

ME reveals a sigmoidal transducer and Poisson-like noise

To test this hypothesis, we presented gratings with different contrasts ranging from 0.05% to 14% and asked 11 participants to report their perceived intensity using any number of their choice (ME; Fig. 1a). We modeled the response distributions using a normal probability density function $N(\mu(x), \sigma(x)^2)$, systematically testing different functional forms for the transducer and noise (fig. S1 to S4). We found that for most participants the data were best described by a model combining a non-saturating sigmoidal transducer (Fig. 2. A and B, Eq. 2) with power-law noise including an additive component (Fig. 2C, Eq. 3, and table S1).

$$\mu(x) = k \frac{x^{\beta}}{v^{\beta-\alpha}+x^{\beta-\alpha}} \quad (2)$$

$$\sigma(x) = \sigma_0 + q\,\mu(x)^{\gamma} \quad (3)$$

At high contrasts ($x \gg v$, where v is the semisaturation constant with a mean across participants of 2.7%; 95% CI: [1.7, 3.7]), the transducer exhibits a compressive nonlinearity approximating a power law $\mu(x) \propto x^{\alpha}$, with a mean exponent of $\alpha = 0.74$ (95% CI: [0.54, 0.94]). This is consistent with earlier reports of compressive power-law behavior in contrast perception obtained using ME (Gottesman et al., 1981). Conversely, at low contrasts ($x \ll v$), the transducer exhibits an expansive nonlinearity $\mu(x) \propto x^{\beta}$ with exponent of $\beta = 4.68$ (95% CI: [3.49, 5.87]). This large exponent captures a rapid steepening of ME ratings at very low intensities. This departure from the simple global power-law behavior has been documented for many sensory attributes (Gescheider, 1997), including contrast (Gottesman et al., 1981).

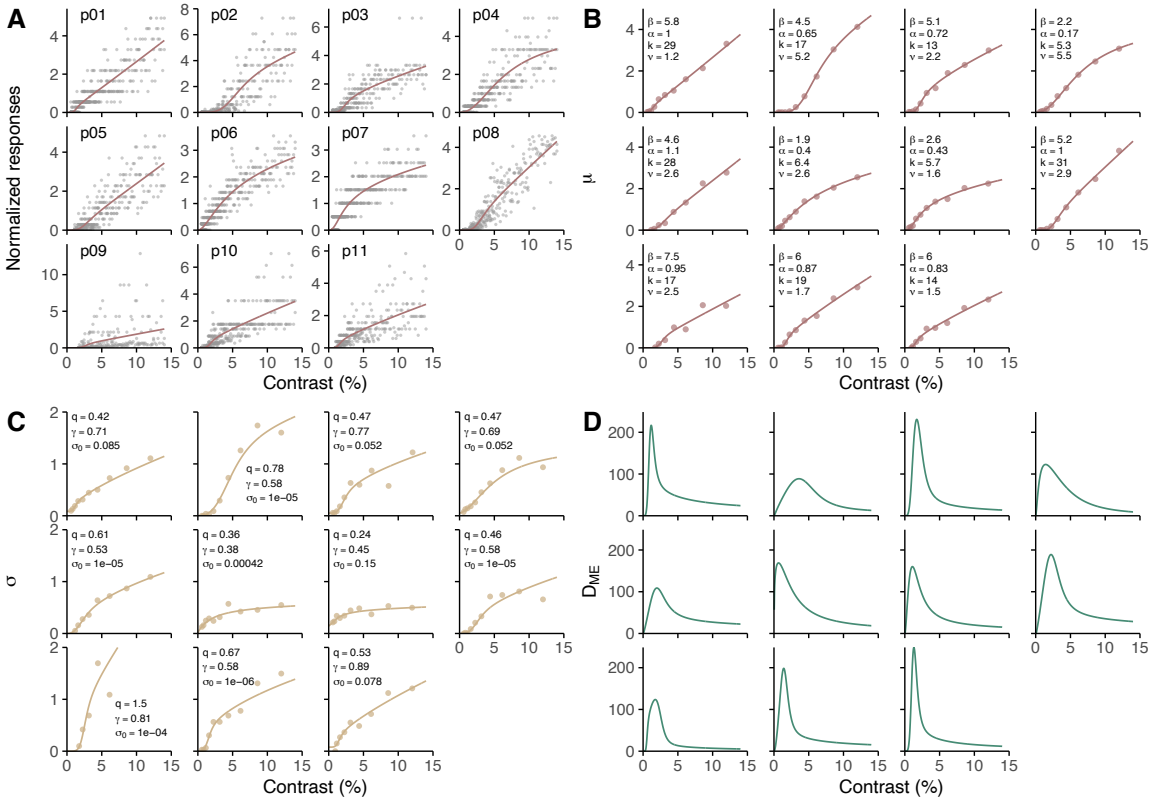

Fig. 2. ME predicts hallmark discrimination phenomena. *(**A**) ME responses (dots) normalized by each participant's mean, plotted against contrast for each participant. Data for each participant are shown in separate sub-panels, and this participant order is maintained in panels **B** to **D**. The solid line represents the best-fitting transducer function (Eq. 2). (**B**) Mean response for each contrast bin (dots) and the corresponding best-fitting transducer function (line). Insets display the optimal parameters of the model. (**C**) Standard deviation of responses for each contrast bin (dots) plotted against contrast. The solid line represents the best-fitting noise function (Eq. 3). Insets display the optimal parameters. Note that the data for participant p09 were truncated for visualization. (**D**) Discrimination sensitivity profiles predicted solely from ME data (transducer and noise) using Eq. 1. The predicted curves recover both the Weber-like behavior (decline at higher contrast) and the pedestal effect (initial rise).*

Internal noise increased as a compressive power law with a mean exponent of $\gamma = 0.64$ (95% CI: [0.53, 0.74]). Thus, we identified a signal-dependent noise structure qualitatively similar to the Poisson-like behavior postulated by Pardo-Vazquez et al. (2019) and Zhou et al. (2024), but with an exponent slightly above the strict Poisson prediction (t(10) = 2.9, P = 0.016; BF_{10} = 4.19). Model comparisons revealed that this power-law parametrization provided a superior description of the data for most participants

compared to the Poisson model and the modulated Poisson model (table S1). However, when the Poisson models were relaxed to include a free scaling parameter ($\sigma(x) = q\,\mu(x)^{0.5}$ and $\sigma(x) = q\sqrt{\mu(x) + g^2\mu(x)}$), the resulting fits were highly comparable to the power-law model (fig. S5 and table S2). Finally, while the additive baseline noise term σ_0 statistically improved the fit across the sample, its estimated values were numerically small for most participants (represented by the y-intercept in Fig. 2C).

Having fully characterized the internal response function via ME, we could predict the expected discrimination sensitivity using Eq. 1 without any free parameters. We found that for all participants, the predicted sensitivity reproduced both Weber-like behavior and the pedestal effect (Fig. 2D).

Discrimination sensitivity quantitatively matches predictions from ME

Having successfully reproduced the qualitative structure of these hallmark discrimination phenomena, we next tested whether the internal representation derived from ME could quantitatively predict empirical discrimination sensitivity. Using the same stimuli, we measured sensitivity in the same participants through a two-interval forced-choice task where they judged which of two successive stimuli had higher contrast (Fig. 1B). To obtain a continuous measure of sensitivity, we modeled the proportion of correct responses (Fig. 3A) using psychometric functions arising explicitly from a transducer $\mu(x)$ and internal noise $\sigma(x)$ (García-Pérez & Alcalá-Quintana, 2007). Because discrimination data cannot uniquely constrain both functions simultaneously (Eq. 1), we used the functional forms that provided the best fit for the ME task (Eq. 2 and 3), but independently refit their parameters to this new dataset. Importantly, this choice was not critical, as conventional fitting procedures, calculating thresholds from independent logistic functions at each reference level, yielded similar sensitivities (fig. S6).

As expected, the contrast discrimination task successfully reproduced Weber-like behavior and the pedestal effect in all participants (Fig. 3B). To quantify the agreement between the sensitivity predicted from ME (D_{ME}) and the empirically observed discrimination sensitivity (D_{Discr}), we compared three defining features of the sensitivity profiles: the peak amplitude, the peak position, and the power-law exponent characterizing the Weber-like regime.

For all three metrics, both tasks yielded values of comparable overall magnitude (Fig. 3, C to E). At the group level, peak amplitudes (Fig. 3C) and Weber-like exponents (Fig. 3E) showed no systematic differences, though D_{ME} peaked at slightly lower contrasts than D_{Discr} (Fig. 3D). To assess predictive accuracy at the individual level, we evaluated whether ME could account for variations in discrimination performance (Fig. 3. F to H). We found that, while individual peak amplitudes did not correlate (Fig. 3F), D_{ME} successfully captured individual differences in peak positions (Fig. 3G) and in Weber-like exponents (Fig. 3H).

When evaluating the full sensitivity profiles at the group level, we observed that D_{ME} underestimated D_{Discr} over a broad range of contrasts after the peak (Fig. 3I). This underestimation might stem from known biases in ME (Petzschner et al., 2015; Poulton, 1968). Overall, however, the quantitative agreement of the sensitivity profiles derived from these two methodologies, which have historically been viewed as fundamentally different, is remarkable.

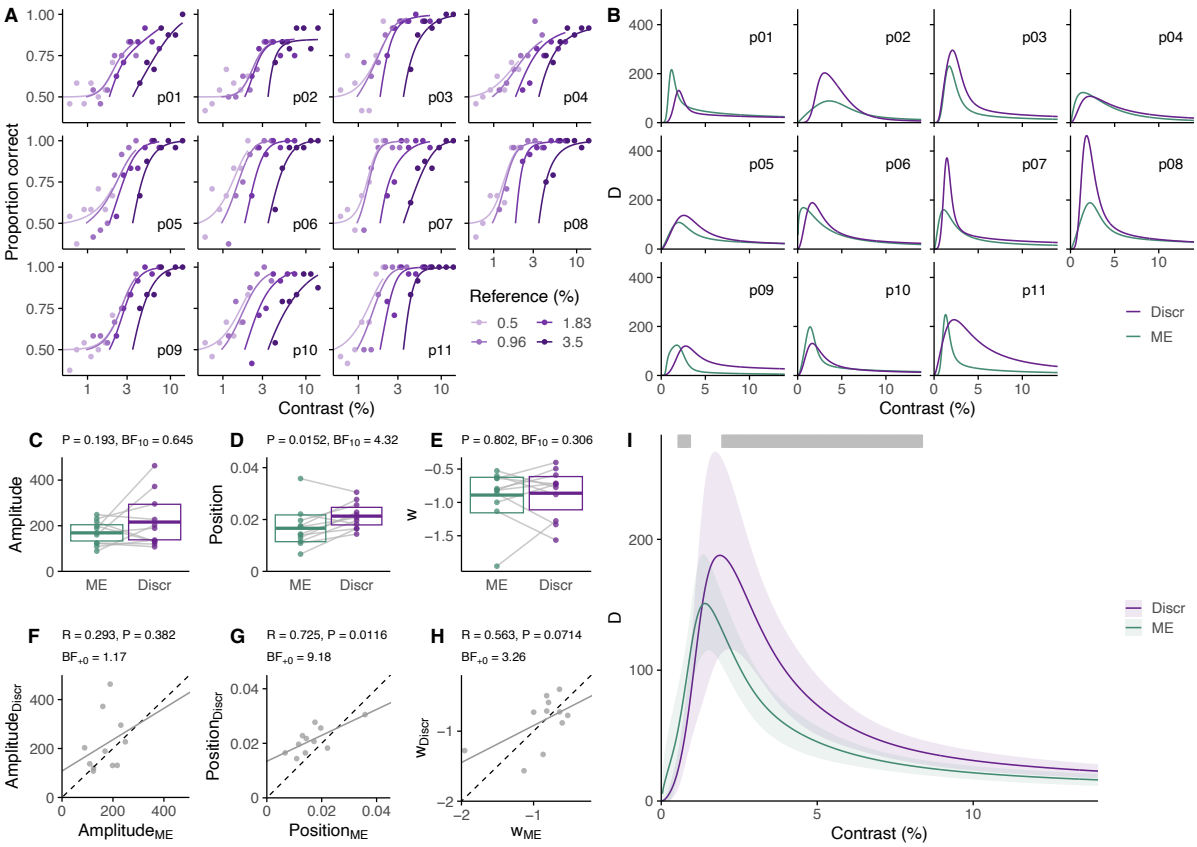

Fig 3. Discrimination sensitivity quantitatively matches predictions from ME. *(**A**) Proportion of correct responses (dots) plotted against the log contrast of the comparison stimulus for four reference contrasts for each participant. Solid lines represent the psychometric functions that best fit this discrimination data. (**B**) Sensitivity profiles derived from the psychometric functions in (**A**) compared with predictions from the ME task. (**C** to **E**) Comparison of the peak amplitude, peak position, and Weber-like exponent between the discrimination and ME sensitivity profiles. The boxes indicate group means and t-based 95% confidence intervals. Annotations indicate P-values from two-tailed paired t-tests (P) and Bayes Factors (BF_{10}) quantifying the evidence against the null hypothesis. (**F to H**) Correlations across participants for the peak amplitude, peak position, and Weber-like exponent between both tasks. Annotations indicate Pearson's correlation coefficient (R), two-tailed P-value (P), and one-tailed Bayes Factor correlation analysis (BF_{+0}). (**I**) Mean sensitivity profile across participants for ME and discrimination tasks. Gray shaded areas indicate contrast values with significant differences between tasks (P < 0.05, two-tailed paired t-test).*

Disentangling the sources of sensitivity

Having confirmed that the internal representation—defined by evaluating the ME parameters in Eq. 2 and 3 (Fig. 2, B and C)—is compatible with discrimination performance, we now use it to disentangle the sources of sensitivity. Substituting this representation into the general sensitivity equation (Eq. 1) yields a closed-form expression that isolates the respective contributions of signal transduction and internal noise.

At high contrasts, sensitivity scales as $D(x) \propto x^{\alpha(1-\gamma)-1}$. Using the mean parameter estimates across participants ($\alpha = 0.74, \gamma = 0.64$), the resulting exponent is $\alpha(1-\gamma)-1 = 0.74 \cdot (1-0.64) - 1 = -0.73$, which is consistent with Weber-like behavior. This decomposition clarifies the relative contributions of the transducer and internal noise to Weber-like behavior. If the noise were constant ($\gamma = 0$), the transducer's compressive nonlinearity alone ($\alpha = 0.74$) would predict a slope of $0.74 \cdot (1-0) - 1 = -0.26$, which is insufficient to capture the steepness of the observed Weber-like behavior. Conversely, a linear transducer ($\alpha = 1$) combined with our observed noise scaling ($\gamma = 0.64$) would yield a slope of $1 \cdot (1-0.64) - 1 = -0.64$, which is much closer to the observed Weber-like behavior. This

suggests that Weber-like behavior primarily emerges from signal-dependent noise rather than transducer compression.

At very low contrast, internal noise is dominated by the additive component (σ_0), yielding a sensitivity that scales as $D(x) \propto x^{\beta-1}$. Sensitivity is therefore governed by the expansive nonlinearity of the transducer with an exponent of $4.68 - 1 = 3.68$. As the contrast increases, the signal-dependent noise term $g\,\mu(x)^{\gamma}$ grows and progressively attenuates this rapid increase in sensitivity. However, even if this term completely dominates, sensitivity continues to increase according to $D(x) \propto x^{\beta(1-\gamma)-1}$ with an exponent of $4.68 \cdot (1 - 0.64) - 1 = 0.68$. Thus, the pedestal effect arises because the expansive nonlinearity of the transducer amplifies the signal sufficiently fast to outpace the increasing noise.

Discussion

While decades of research in the tradition of Stevens' psychophysics have documented how the mean of subjective ratings scales with intensity (Gescheider, 1997), the variability of these responses has been largely overlooked. Here, we analyzed this variability and found that it follows a lawful, signal-dependent pattern. Importantly, when coupled with the mean response, this variability quantitatively predicts the complex nonlinear phenomena observed in discrimination tasks—specifically Weber-like behavior and the pedestal effect, hallmarks of Fechner's psychophysics that have constrained theories of sensory processing (Legge & Foley, 1980; Pardo-Vazquez et al., 2019). This result not only validates the variability of ME as a behaviorally accessible measure of perceptual noise, but also provides direct empirical support for the foundational assumption that the mean response reflects the sensory transducer function. Ultimately, our findings show that ME effectively characterizes the internal representation of intensities, demonstrating that a single internal response function underlies both subjective ratings and objective discrimination. Thus, while recent work has established that Fechner's and Stevens' psychophysics could be mathematically compatible (Zhou et al., 2024), our results provide strong evidence that they are empirically compatible when the variability of ME is accounted for.

We demonstrated this unification using visual contrast, a canonical dimension for studying sensory processing (DeValois & DeValois, 1990). The alignment of the contrast-dependent sensitivity curves derived from ME and discrimination tasks suggests that these judgments rely on a shared internal representation. This core representation is likely situated at early processing stages, before the divergence of task-specific decision mechanisms. These later mechanisms likely account for the overall underestimation of the sensitivity that we derived from ME (Petzschner et al., 2015; Poulton, 1968). Despite this task-specific offset, recovering the functional form of the internal representation of contrast allows us to directly address the long-standing debate over whether the limits on contrast discrimination arise primarily from nonlinear transduction or from variable internal noise (e.g. Georgeson & Meese, 2006). Our results show that both mechanisms are necessary, but dominate sensitivity at different contrast regimes. At low contrasts, the pedestal effect emerges from the transducer's expansive nonlinearity, which acts as an effective exponentiation threshold that amplifies signal gain sufficiently to overcome internal noise. At higher contrasts, Weber-like behavior arises primarily from the signal-dependent Poisson-like noise, which progressively overtakes signal gain to attenuate sensitivity.

The components of the internal representation that we recover are consistent with known biological mechanisms. The derived sigmoidal transducer matches contrast response functions observed in animal visual cortex (Albrecht & Hamilton, 1982), human fMRI (Boynton et al., 1999), and EEG (Norcia et al., 2015), and aligns naturally with the canonical normalization model of neural coding (Heeger, 1992). Similarly, the slightly supra-Poisson noise structure that we identified mirrors the signal-dependent variability reported in visual cortical neurons across contrast levels (Tolhurst et al., 1981).

Materials and methods

Participants

Eleven participants (7 males; age range 25-45 years), including the first and last authors, with normal or corrected-to-normal visual acuity participated after written informed consent was obtained. Five participants were experienced in psychophysical tasks, two of whom were familiar with discrimination tasks, while none had prior experience with ME tasks. The study was approved by the University of Barcelona Ethics Committee.

Apparatus and Stimuli

Stimuli were generated using PsychoPy (Peirce et al., 2019) and displayed on a gamma-corrected monitor (ASUS ROG PG258Q; 1920 x 1080 pixels) viewed from a distance of 57 cm in a lit room. The stimuli consisted of vertical Gabor patches (standard deviation of the Gaussian envelope: 0.5º; spatial frequency: 1.5 cycles/º) and a fixation cross (height: 0.58º, 101.4 cd/m^2) on which participants were instructed to maintain fixation. All stimuli were presented against a uniform grey background (180 cd/m^2). To achieve high bit-depth resolution for contrast presentation, we implemented the noisy-bit method (Allard & Faubert, 2008).

Experimental procedure

ME task

Each trial began with a fixation point displayed for 500 ms, followed by a 500 ms blank interval, and concluded with a Gabor patch presented for 100 ms. Then, participants provided a verbal numerical estimate of the perceived contrast using an unrestricted numerical scale (allowing both integers and decimals), which the experimenter recorded manually. The Gabor patches were presented at 200 contrast levels, logarithmically spaced between 0.5% and 14%. Each contrast level was presented once per session in a random order across two separate sessions, resulting in a total of 400 trials per participant. Prior to these sessions, participants completed a practice block of 30 trials to stabilize their use of the internal response scale.

Discrimination task

Before conducting the ME task, participants performed a two-interval forced-choice discrimination task. Each trial began with the presentation of a fixation point for 500 ms, followed by a 500 ms blank interval. Then, the reference and comparison stimuli were presented sequentially in random order for 100 ms each, separated by an inter-stimulus interval of 1000 ms. Participants indicated which interval contained the stimulus with higher contrast using a keyboard response.

Four reference contrast levels (0.5%, 0.96%, 1.83%, and 3.5%) were each paired with seven logarithmically spaced comparison contrasts. The comparison ranges were 0.61–2% for the 0.5% reference, 1.17–3.84% for the 0.96% reference, 2.23–7.32% for the 1.83% reference, and 4.27–14% for the 3.5% reference. For every reference-comparison pair, there were eight repetitions, consisting of two presentation orders (reference or comparison first) crossed with four spatial phases ($0, \pi/2, \pi, 3/2\pi$). Participants completed three blocks of trials, resulting in a total of 672 trials (4 references x 7 comparisons x 8 repetitions x 3 blocks). The trial order was fully randomized within each block.

Data Analysis

Analysis and figures were conducted using the R statistical computing environment (R Core Team, 2025).

ME task

To ensure data quality prior to curve fitting, we preprocessed the data to address scale usage, outliers, and zero-inflation. First, to assess the stability of scale usage, we estimated the standard deviation of responses across consecutive trial bins. This analysis revealed unusually large response variability during the initial trials of three participants (p03, p09 and p11), leading to the exclusion of the first 10% of their trials. Next, outliers were identified by fitting a LOESS curve to the raw data. Responses falling outside ± 3 standard deviations from the fitted curve were excluded, resulting in the removal of only 9 trials across the entire dataset. Finally, during preliminary curve fitting, we observed poor model fits at high contrast levels for three participants (p05, p10, p11). This was driven by a heavy concentration of zero responses at low contrasts, which disproportionately biased the maximum likelihood estimation. To restore the balance of the likelihood function, we randomly downsampled the zero responses for these specific participants, removing 21%, 18%, and 35% of their total trials, respectively.

To facilitate comparisons across participants, the ME responses were normalized by dividing each participant's trial responses by their own overall mean. We modeled these normalized responses r using a normal distribution, $N(\mu(x), \sigma(x))$, and fitted its parameters by maximizing the log-likelihood function:

$$log\, L = -\frac{n}{2} log\,(2\pi) - \sum_{i=1}^{n} log\,(\sigma(x_i)) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{r_i - \mu(x_i)}{\sigma(x_i)}\right)^2$$

where n is the number of responses, $\mu(x)$ represents the transducer and $\sigma(x)$ the internal noise. By considering different forms of the transducer and the noise, we evaluated a comprehensive set of candidate models (figs. S1 to S5). We performed the optimization using the R package *DEoptim* (Ardia et al., 2022), which performs global optimization over bounded parameter spaces. Model selection (tables S1 and S2) was performed using the Akaike Information Criterion (Burnham & Anderson, 2013).

Discrimination task

We modeled the proportion of correct responses in the two-interval forced-choice task within a signal detection theory framework (García-Pérez & Alcalá-Quintana, 2007). In this approach, psychometric functions for all reference levels are fitted simultaneously by assuming an underlying transducer function, $\mu(x)$, and an internal noise function, $\sigma(x)$. However, discrimination data alone cannot uniquely identify either the underlying functional forms or their specific parameters (Eq. 1). For the functional forms, adopting the best model established by our magnitude estimation data (Eq. 2 and 3) emerges as the natural choice. As for the parameters, we left them free to vary during the fitting process, but their values remain underdetermined by the signal-to-noise constraint. Consequently, these parameters serve exclusively as a computational tool to extract the continuous empirical sensitivity profile and should not be considered reliable for direct comparison with those obtained from magnitude estimation.

Within this framework, the perceptual responses elicited by the reference and the comparison stimuli were modeled as $\mathrm{R}_r \sim N(\mu(x), \sigma(x)^2)$ and $\mathrm{R}_c \sim N(\mu(x+\Delta x), \sigma(x+\Delta x)^2)$, respectively, where x represents the contrast of the reference stimulus and $x + \Delta x$ represents the contrast of the comparison stimulus. Decisions were assumed to be based on the difference between these two independent perceptual responses, $R_{dif} = R_c - R_r$, which is distributed as $N(\mu(x+\Delta x) - \mu(x), \sigma(x)^2 + \sigma(x+\Delta x)^2)$.

The expected proportion of correct responses corresponds to the probability that this difference is positive, indicating that the comparison stimulus elicited a stronger response than the reference. Evaluated across values of Δx, this probability defines the predicted psychometric functions:

$$P(R_{dif} > 0) = \int_0^{\infty} N(R_{dif})\, dR_{dif}$$

To estimate the best-fitting parameters for the chosen functional forms, we used maximum likelihood estimation, maximizing the binomial log-likelihood across all reference levels simultaneously using the R package *nloptr* (Johnson, 2008). Finally, we calculated sensitivity using Eq. 1.

We used this joint-fitting approach because it naturally provides a continuous measure of sensitivity across all contrast levels. Importantly, fitting independent psychometric functions for each reference level and estimating sensitivity as the inverse of the threshold yielded comparable results (fig. S6).

Statistical analysis

We conducted mean comparisons using paired two-tailed Student's t-tests and Bayes factor analyses with standard Cauchy priors (scale = 0.707). We assessed the strength and direction of linear associations using Pearson correlation coefficients (R) and one-tailed Bayesian correlation analyses (BF_{+0}) to test for positive associations. In accordance with recent guidelines (Keysers et al., 2020), we employed directional priors for hypotheses with a predicted direction to mitigate the bias against the null hypothesis that can arise from two-tailed Bayesian tests.

Acknowledgments: We thank Cristina de la Malla and Albert Compte for comments in an early version of the manuscript.

Funding: Ministry of Science and Innovation fellowship PRE2021-099277 (CRA), PID2023-150081NB-I00 grant MICIU/AEI/10.13039/501100011 (JLM),
PID2023-151752NB-I00 grant MICIU/AEI/10.13039/501100011 (DL)

Competing interests: Authors declare that they have no competing interests

References

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, *48*(1), 217-237. https://doi.org/10.1152/jn.1982.48.1.217

Allard, R., & Faubert, J. (2008). The noisy-bit method for digital displays: Converting a 256 luminance resolution into a continuous resolution. *Behavior Research Methods*, *40*(3), 735-743. https://doi.org/10.3758/brm.40.3.735

Arabzadeh, E., Clifford, C. W. G., & Harris, J. A. (2008). Vision Merges With Touch in a Purely Tactile Discrimination. *Psychological Science*, *19*(7), 635-641. https://doi.org/10.1111/j.1467-9280.2008.02134.x

Ardia, D., Mullen, K., Peterson, B., Ulrich, J., & Boudt, K. (2022). *DEoptim: Global Optimization by Differential Evolution* (Versión 2.2-8) [Software]. https://cran.r-project.org/web/packages/DEoptim/index.html

Boynton, G. M., Demb, J. B., Gloverd, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research*, *39*(2), 257-269. https://doi.org/10.1016/S0042-6989(98)00113-8

Burnham, K. P., & Anderson, D. R. (2013). *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media.

DeValois, R. L., & DeValois, K. K. (1990). *Spatial Vision*. OUP USA.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf u. Härtel.

García-Pérez, M. A., & Alcalá-Quintana, R. (2007). The transducer model for contrast detection and discrimination: Formal relations, implications, and an empirical test. *Spatial Vision*, *20*(1/2), 5. https://doi.org/10.1163/156856807779369724

Georgeson, M. A., & Meese, T. S. (2006). Fixed or variable noise in contrast discrimination? The jury's still out…. *Vision Research*, *46*(25), 4294-4303. https://doi.org/10.1016/j.visres.2005.08.024

Gescheider, G. A. (1997). *Psychophysics: The Fundamentals* (3.ª ed.). Psychology Press. https://doi.org/10.4324/9780203774458

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, *17*(6), 858-865. https://doi.org/10.1038/nn.3711

Gottesman, J., Rubin, G. S., & Legge, G. E. (1981). A power law for perceived contrast in human vision. *Vision Research*, *21*(6), 791-799. https://doi.org/10.1016/0042-6989(81)90176-0

Green, D. M. (1960). Psychoacoustics and Detection Theory. *The Journal of the Acoustical Society of America*, *32*(10), 1189-1203. https://doi.org/10.1121/1.1907882

Green, D. M., & Luce, R. D. (1974). Variability of magnitude estimates: A timing theory analysis. *Perception & Psychophysics*, *15*(2), 291-300. https://doi.org/10.3758/BF03213947

Hanna, T. E., von Gierke, S. M., & Green, D. M. (1986). Detection and intensity discrimination of a sinusoid. *The Journal of the Acoustical Society of America*, *80*(5), 1335-1340. https://doi.org/10.1121/1.394385

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*(2), 181-197. https://doi.org/10.1017/S0952523800009640

Johnson, S. G. (2008). *The NLopt nonlinear-optimization package*. https://github.com/stevengj/nlopt

Keysers, C., Gazzola, V., & Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*(7), 788-799. https://doi.org/10.1038/s41593-020-0660-4

Kingdom, F. A. A., & Prins, N. (2016). *Psychophysics: A Practical Introduction* (2nd ed.). Academic Press. https://www.elsevier.com/books/psychophysics/kingdom/978-0-12-407156-8

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *JOSA, Vol. 70, Issue 12, Pp. 1458-1471*. https://doi.org/10.1364/JOSA.70.001458

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better than detection. *Vision Research*, *14*(10), 1039-1042. https://doi.org/10.1016/0042-6989(74)90175-8

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*(6), 4. https://doi.org/10.1167/15.6.4

Pardo-Vazquez, J. L., Castiñeiras-de Saa, J. R., Valente, M., Damião, I., Costa, T., Vicente, M. I., Mendonça, A. G., Mainen, Z. F., & Renart, A. (2019). The mechanistic foundation of Weber's law. *Nature Neuroscience*, *22*(9), 1493-1502. https://doi.org/10.1038/s41593-019-0439-7

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285-293. https://doi.org/10.1016/j.tics.2015.03.002

Poulton, E. C. (1968). The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, *69*(1), 1-19. https://doi.org/10.1037/h0025267

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Seriès, P., Stocker, A. A., & Simoncelli, E. P. (2009). Is the Homunculus "Aware" of Sensory Adaptation? *Neural Computation*, *21*(12), 3271-3304. https://doi.org/10.1162/neco.2009.09-08-869

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153-181. https://doi.org/10.1037/h0046162

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, *41*(3), 414-419. https://doi.org/10.1007/BF00238900

Zhou, J., Duong, L. R., & Simoncelli, E. P. (2024). A unified framework for perceived magnitude and discriminability of sensory stimuli. *Proceedings of the National Academy of Sciences*, *121*(25), e2312293121. https://doi.org/10.1073/pnas.2312293121

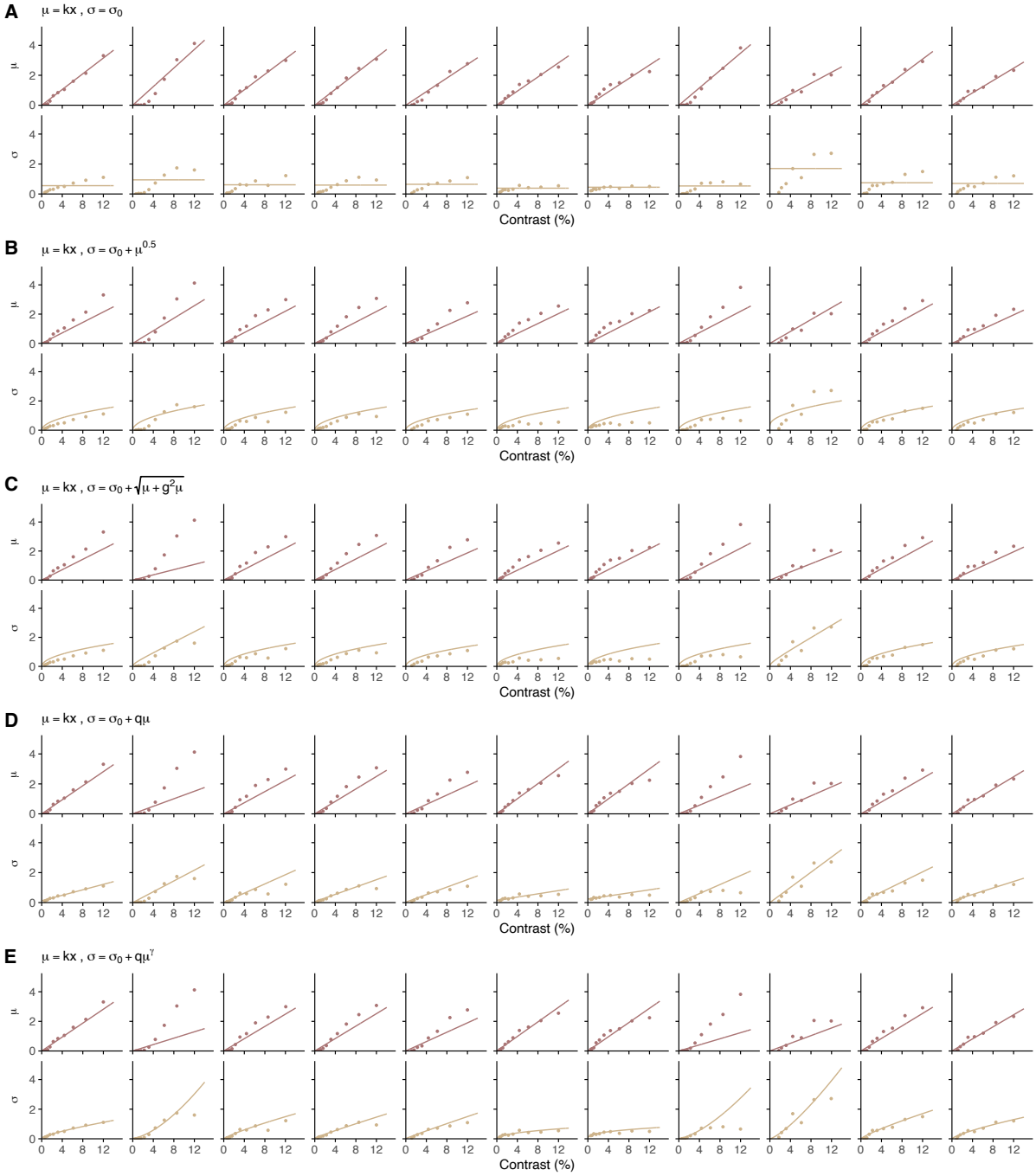

Fig. S1. Model fits for a linear transducer coupled with alternative noise functions. *All models assume a linear transducer but vary in their internal noise assumptions: (**A**) additive, (**B**) Poisson, (**C**) modulated Poisson, (**D**) multiplicative, and (**E**) power. Each column represents an individual participant (ordered from left to right as p01 to p11, matching the labels in the main text). Within each panel, the top row shows the mean response (μ) and the bottom row shows the response standard deviation (σ) as a function of stimulus contrast. Dots indicate the empirical estimates for each contrast bin, while solid lines represent the corresponding model fits. To prevent parameters from increasing indefinitely during the fitting process, we applied an upper bound of 2 to parameter g for participant p02 in the model in (**C**), and to parameter q for participants p02 and p08 in the model in (**E**). In all other cases, the optimized parameters did not hit their boundary constraints.*

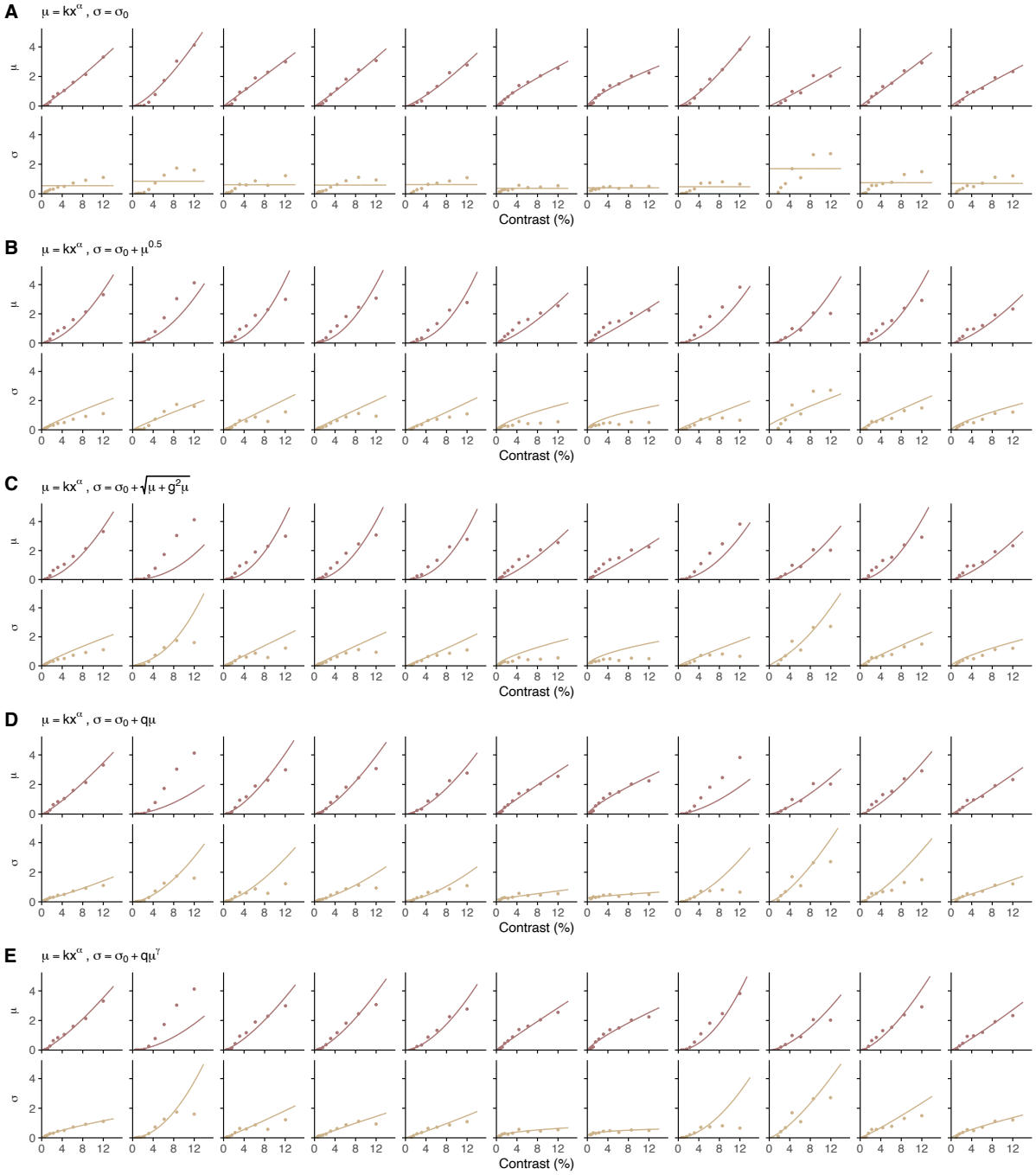

Fig. S2. Model fits for a power law transducer coupled with alternative noise functions. *Figure layout and conventions are identical to those in fig. S1. The panels compare fits for the same five internal noise models ((**A**) additive, (**B**) Poisson, (**C**) modulated Poisson, (**D**) multiplicative, and (**E**) power noise), but assuming a power-law transducer. To prevent parameter k from increasing indefinitely during the fitting process, we applied an upper bound—defined by the mean of the remaining participants—to parameter k for participants p02 and p08 in the models in (**B**) through (**E**). As a result of this restriction, for participant p02, the optimized value of parameter g for the model in (**C**) and parameter q for the models in (**D**) and (**E**) reached their upper bound of 2. In all other cases, the optimized parameters did not reach their boundary constraints.*

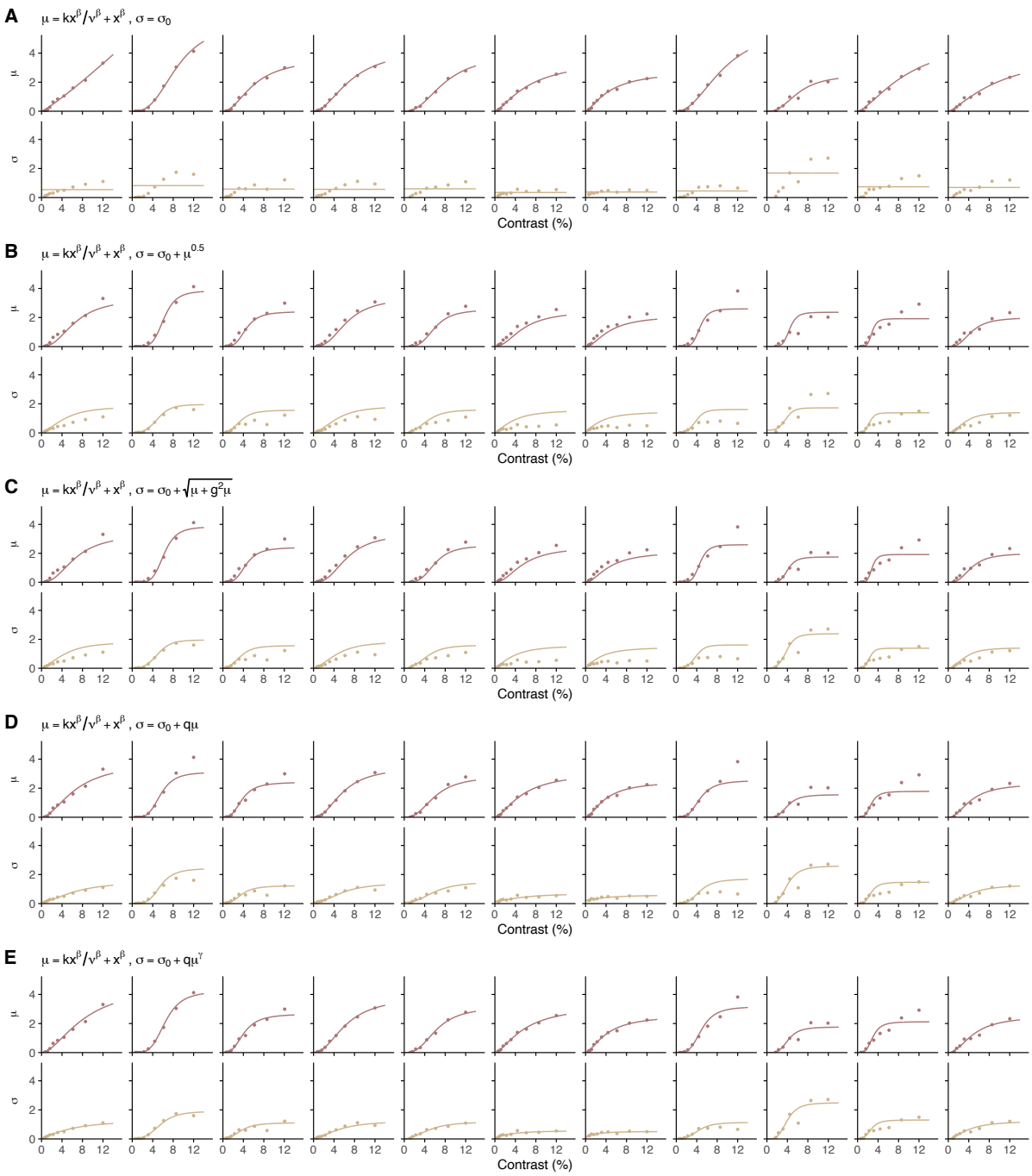

Fig. S3. Model fits for a saturating sigmoidal transducer coupled with alternative noise functions. *Figure layout and conventions are identical to fig. S1. The panels compare fits for the same five internal noise models ((**A**) additive, (**B**) Poisson, (**C**) modulated Poisson, (**D**) multiplicative, and (**E**) power noise), but assuming a saturating sigmoidal transducer. The optimized parameters did not reach their boundary constraints.*

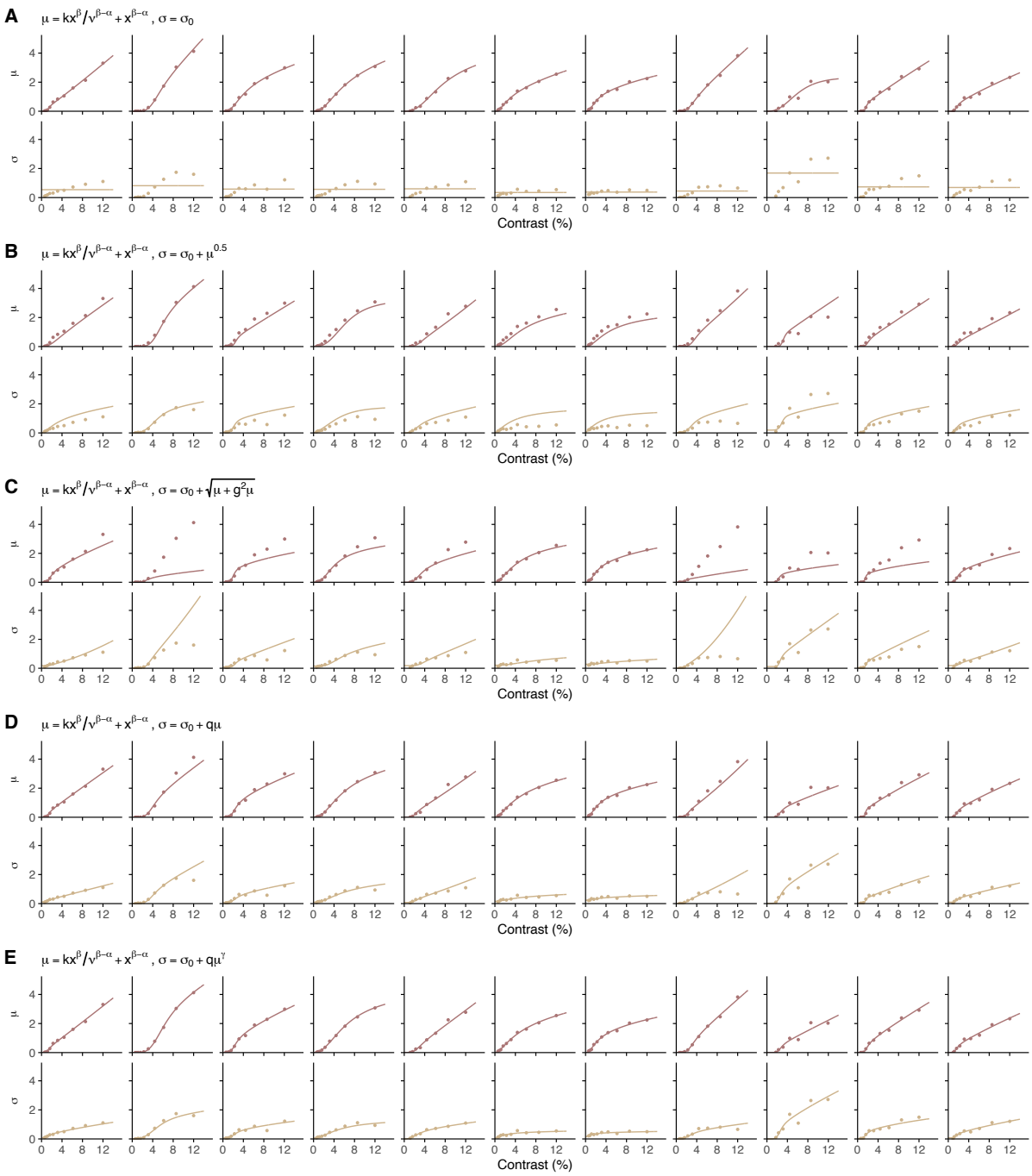

Fig. S4. Model fits for a non-saturating sigmoidal transducer coupled with alternative noise functions. *Figure layout and conventions are identical to Supplementary Figure 1. The panels compare fits for the same five internal noise models ((**A**) additive, (**B**) Poisson, (**C**) modulated Poisson, (**D**) multiplicative, and (**E**) power noise), but here all models assume a non-saturating sigmoidal transducer. The optimized parameters did not reach their boundary constraints.*

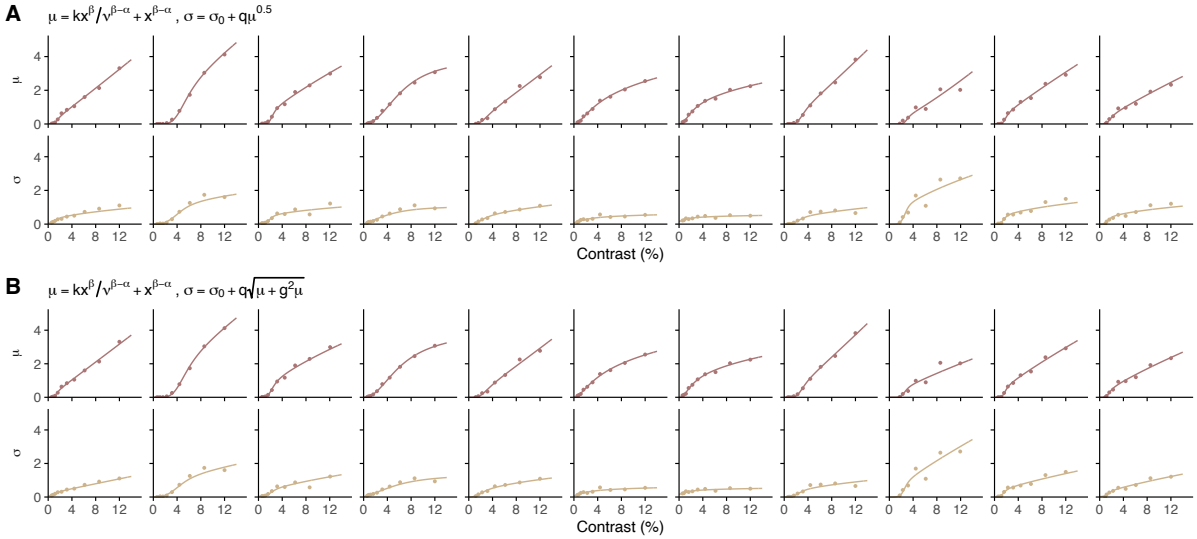

Fig. S5. Model fits for a non-saturating sigmoidal transducer coupled with scaled Poisson models. *Figure layout and conventions are identical to fig S1. The panels compare fits for a Poisson (**A**) and a modulated Poisson (**B**) models that include a free scaling parameter and assume a non-saturating sigmoidal transducer. The optimized parameters did not hit their boundary constraints. The resulting fits were highly comparable to the power-law model (fig. S4E) suggesting that any of these three models could be reasonably chosen. We opted to retain the power-law formulation because it remained the best-fitting model for most participants (table S1).*

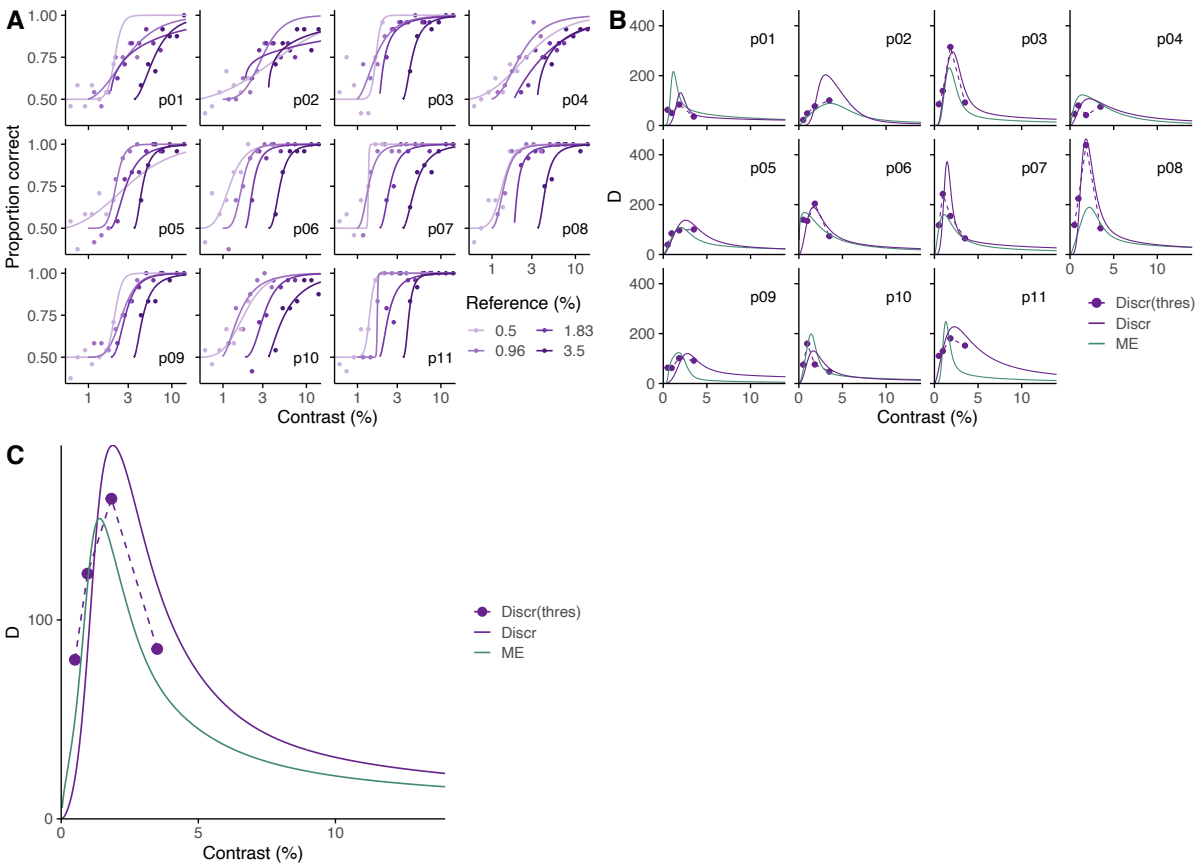

Fig. S6. Discrimination sensitivity estimated from thresholds of independently fitted psychometric functions. *(**A**) Proportion of correct responses (dots) plotted against the log contrast of the comparison stimulus for four reference contrasts for each participant. Solid lines represent the best-fitting psychometric functions fitted independently to each reference level and participant using a logistic function. (**B**) Sensitivity computed as the inverse of the threshold* (Δx^{-1}) *for the four reference levels (dots), derived from the independent fits in (A). For comparison, the sensitivity profiles (lines) from Fig. 3B are replotted. Thresholds were defined at the 76% performance level, which corresponds to a detectability index* d' *of 1 in a two-alternative forced-choice task, where the proportion of correct responses is given by* $\Phi\left(\frac{d'}{\sqrt{2}}\right)$*, and* Φ *is the standard cumulative normal function (Kingdom & Prins, 2016; Seriès et al., 2009). Under this condition, given the general relationship* $\Delta x^{-1}(x) = \frac{D(x)}{d'(x)}$*, sensitivity estimated as the inverse of the threshold coincides with the sensitivity metric defined in Equation 1 (Seriès et al., 2009; Zhou et al., 2024). (**C**) Mean sensitivities across participants for the four reference levels (dots) plotted together with the mean sensitivity profiles (lines) from Fig. 3I.*

Model	N	AIC	ΔAIC	wAIC
$\mu = \frac{kx^{\beta}}{\nu^{\beta-\alpha}+x^{\beta-\alpha}}, \sigma = \sigma_0 + q\mu^{\gamma}$	9	1778	0	0.723
$\mu = \frac{kx^{\beta}}{\nu^{\beta}+x^{\beta}}, \sigma = \sigma_0 + q\mu^{\gamma}$	1	2009	230	0.136
$\mu = \frac{kx^{\beta}}{\nu^{\beta-\alpha}+x^{\beta-\alpha}}, \sigma = \sigma_0 + q\mu$	1	2108	330	0.135

Table S1. Model selection summary for the candidate models from figs. S1-S4 (which exclude the scaled Poisson models) that were the best fit for at least one participant. *The best-fitting model for each participant is identified by the lowest Akaike Information Criterion (AIC) score. N represents the number of participants for whom a given model provided the best fit. To evaluate performance at the group level, the reported AIC corresponds to the sum of individual AIC scores across all participants. ΔAIC represents the difference between a given model's summed AIC and the summed AIC of the best-performing model in the candidate set (where ΔAIC = 0). Finally, wAIC indicates the Akaike weights averaged across all participants, reflecting the mean probability of each model being the best fit among the entire set of candidate models in figs. S1 to S4.*

Model	N	AIC	ΔAIC	wAIC
$\mu = \frac{kx^{\beta}}{\nu^{\beta-\alpha}+x^{\beta-\alpha}}, \sigma = \sigma_0 + q\mu^{\gamma}$	4	1778	0	0.402
$\mu = \frac{kx^{\beta}}{\nu^{\beta-\alpha}+x^{\beta-\alpha}}, \sigma = \sigma_0 + q\sqrt{\mu + g^2\mu}$	2	1790	12	0.259
$\mu = \frac{kx^{\beta}}{\nu^{\beta-\alpha}+x^{\beta-\alpha}}, \sigma = \sigma_0 + q\mu^{0.5}$	3	1825	46	0.165
$\mu = \frac{kx^{\beta}}{\nu^{\beta}+x^{\beta}}, \sigma = \sigma_0 + q\mu^{\gamma}$	1	2009	230	0.088
$\mu = \frac{kx^{\beta}}{\nu^{\beta-\alpha}+x^{\beta-\alpha}}, \sigma = \sigma_0 + q\mu$	1	2108	330	0.083

Table S2. Model selection summary for all candidate models including the scaled Poisson models (figs. S1-S5) that were the best fit for at least one participant. *The best-fitting model for each participant is identified by the lowest Akaike Information Criterion (AIC) score. N represents the number of participants for whom a given model provided the best fit. To evaluate performance at the group level, the reported AIC corresponds to the sum of individual AIC scores across all participants. ΔAIC represents the difference between a given model's summed AIC and the summed AIC of the best-performing model in the candidate set (where ΔAIC = 0). Finally, wAIC indicates the Akaike weights averaged across all participants, reflecting the mean probability of each model being the best fit among the entire set of candidate models in figs. S1 to S5.*