Shared neuronal variability accounts for behavioral variability in count discrimination tasks

Mikio C. Aoi, Benjamin B. Scott, Christine M. Constantinople, Carlos D. Brody, & Jonathan W. Pillow

The psychophysics of numerosity, or number sense, is the study of how discrete numbers of stimuli are perceived. Empirical studies have shown that the number sense conforms to Weber's law of perceptual discriminability. This means that the standard deviation of the perceptual noise on the perceived number of objects scales linearly with the number of objects, indicating perceptual noise is not independent of the percept. This property, observed in both humans and non-human animals, is commonly known as "scalar variability." The generation of large scale behavioral data sets ($\approx 10^6$ trials) from animals trained to perform numerosity tasks, allows us to determine whether or not judgements about count stimuli coincide with deviations from precise scalar variability (Scott & Constantinople, et al. *Elife*, 2015). Here we propose a probabilistic behavioral model to account for variability in count discrimination tasks. Our model, based on a model of shared stochastic gain in neuronal populations, can closely fit the psychometric function and the inferred uncertainty of count perception from behaving animals. We compare the performance of our model to a previously proposed 16-parameter model based on signal detection theory and show that our model fits data better than the previous model with just two parameters. This work draws a direct connection between neurophysiology and behavior by demonstrating that perceptual psychophysics can be explained by the statistical properties of neuronal populations.

Additional Information

Model description: Our model is motivated by recent work on neural response variability in the monkey visual system (Goris & Ziemba et al., COSYNE, 2017; Goris, Movshon, Simoncelli. *Nat Neuro*, 2014). This work showed that over-dispersion in neural spiking could be explained by a stochastic gain that is shared among all neurons. We generalized this model to spiking activity from a population of neurons used to perform a perceptual decision making task and, rather than directly fit this model to neuronal data, we used it to account for behavior in a count discrimination task.

Our model begins with characterizing the neuronal responses on each trial. Ignoring the timing of spikes, we assume that the trial-by-trial response of each neuron is given by a Poisson random variable with a rate that is proportional to the number of stimuli n, i.e. $\eta_i|n \sim \text{Poiss}(\lambda_i)$, where $\lambda_i = c_i n + \delta_i$. This model is consistent with models of neural integration during evidence accumulation (Hanks et al., *Nature*, 2015; Scott et al., *Neuron*, 2017). The parameters c_i and δ_i are neuron-specific sensitivity and baseline noise parameters, respectively. If we sum the total number of spikes over all neurons then the population spike rate will be $\lambda = cn + \delta$, where $c = \sum_i c_i$ and $\delta = \sum_i \delta_i$.

In addition to the noise induced by Poisson firing, on each trial there is a shared, stochastic gain that acts as multiplicative noise. If the gain on a given trial γ is drawn from a gamma distribution then we may model the number of population spikes η on each trial by

$$\eta|\gamma, n, c, \delta \sim \text{Poiss}(\gamma\lambda), \qquad \gamma|\nu \sim \text{Gamma}(1/\nu, \nu),$$

where ν is the variance of γ. These two distributions form a conjugate prior and likelihood, allowing us to derive the marginal distribution of η in closed form, resulting in a negative binomial distribution

$$\eta|\nu, n, c, \delta \sim \text{NB}(1/\nu, p)$$

where $p = \frac{\nu(cn+\delta)}{\nu(cn+\delta)+1}$.

Count discrimination: Suppose a subject is presented with n_L stimuli on the left, and n_R stimuli on the right. A separate population of neurons counts each of the sets of stimuli. Assuming

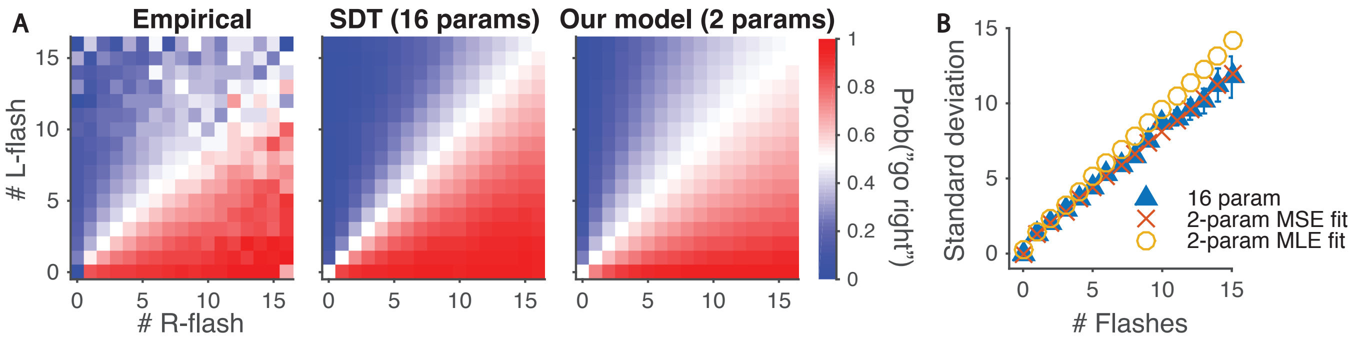

Figure 1: **Comparison of model fits. A)** Estimated psychometric functions from the count discrimination task. The signal detection theory (SDT) estimate is based on the model presented in (Scott & Constantinople, et al. *Elife*, 2015). **B)** Comparison of standard deviations based on maximum likelihood using the SDT model and our 2-parameter model, as well as a direct fit of our variance model to the estimates from the SDT model (MSE).

that the perceptual parameters are the same for both populations, but that the populations are independent. The responses to the number of stimuli can be described by $\eta_j \sim \text{NB}(1/\nu, p_j)$, where $p_j = \nu(cn_j + \delta)/(1 + \nu(cn_j + \delta))$, $j \in \{R, L\}$. The subject's judgement will be made on $Y \equiv \eta_R - \eta_L$, which is distributed according to a *skewed generalized discrete Laplace distribution* (GDL) (Lekshmi & Sebastian. *Int. J of Math. and Stat. Invention*, 2014), which gives the probability mass $P_{GDL}(Y = m|\nu, n_L, n_R)$ that the difference in perceived stimuli will be m. If the subject's decisions are determined completely by the GDL distribution then the probability of the animal choosing "right" is simply

$$P(\text{right}|\theta, n_L, n_R) = P(Y > 0|\theta, n_L, n_R) = \sum_{m=1}^{\infty} P(Y = m|\theta, n_L, n_R), \tag{1}$$

where $\theta = \{\nu, \delta, c\}$. Thus, the discriminative behavior of the animal can conceivably be described by the above distribution with only three parameters.

Experiments: We examined how well our model describes animal behavior during a count discrimination task (Scott & Constantinople, et al. *Elife*, 2015). On each trial a rat was presented with between 0-16 flashes from LED's on both sides of their visual field and were trained to report by nose-poke which side had the larger number of flashes.

We used over 5×10^6 trials of behavioral data to fit (1) with $c = 1$ (i.e. number of spikes scales exactly with number of flashes). We compared the resulting 2-parameter model fits to a previously described 16-parameter model based on signal detection theory (SDT) (Scott & Constantinople, et al. *Elife*, 2015). The estimated psychometric functions are presented in Figure 1A. We found that our 2-parameter model had a larger log-likelihood (LL) than the 16-parameter model ($\text{LL}_{2\text{ param}} = -3.04 \times 10^5$, $\text{LL}_{16\text{ param}} = -3.12 \times 10^5$), indicating that our model better fit the data with far fewer degrees of freedom.

We also compared estimates of perceptual noise for each flash count from our model to estimates from the SDT model. The SDT model estimated a different variance for each flash count. In contrast, our model has a functional form of the variance given by $\text{Var}[\eta|n] = n + \delta + \nu(n + \delta)^2$. As shown in Figure 1B, we find that maximum likelihood fits of our parameters display qualitatively similar fits as the SDT model, but imply slightly larger perceptual noise. This is not due to the inability of our model to fit the estimates from the SDT model however, as we found that directly fitting the model parameters to the SDT model standard deviations (by minimizing mean-squared error (MSE)) resulted in nearly perfect agreement.