Robust averaging during perceptual judgment is not optimal

de Gardelle and Summerfield (1) claimed that, when judging the mean of a set of stimuli, it is optimal to downweight outliers, and that human subjects follow this robust averaging strategy. Here, we show that, in their task, the optimal observer would equally weight all observations. In ref. 1, subjects were presented with a set of eight colors that are denoted by a vector $\mathbf{x} = (x_1, \ldots, x_8)$ and drawn independently from a Gaussian distribution on a red–blue color axis with variance σ^2 and mean of either μ (blue) or $-\mu$ (red). On a given trial, μ was set randomly to one of two values, and σ^2 was set randomly to one of three values. The subject indicated whether the mean was blue ($C = 1$) or red ($C = -1$). When the prior probabilities are equal, the optimal decision is based on the likelihoods of both options [i.e., $p(\mathbf{x}|C = 1)$ and $p(\mathbf{x}|C=-1)$]. Because μ and σ^2 are unknown to the observer, the optimal observer computes these likelihoods by averaging (marginalizing) over all six possibilities (Eq. **1**):

$$\text{Correct: } p(\mathbf{x}\,|\,C) = \frac{1}{6}\sum_{\mu,\sigma} p\left(\mathbf{x}\,|\,\mu,\sigma^2,C\right) = \frac{1}{6}\sum_{\mu,\sigma}\prod_i p\left(x_i\,|\,\mu,\sigma^2,C\right), \quad [1]$$

where in the last step, we have used the conditional independence of the observations given μ and σ^2. The work by de Gardelle and Summerfield (1), however, computed the likelihoods by first factorizing and then marginalizing (Eq. **2**):

$$\text{Incorrect: } p(\mathbf{x}\,|\,C) = \prod_i p(x_i\,|\,C) = \prod_i \frac{1}{6}\sum_{\mu,\sigma} p\left(x_i\,|\,\mu,\sigma^2,C\right). \quad [2]$$

The first step in Eq. **2** is a mathematical mistake, because the observations are only independent when conditioned on μ and σ^2. They strongly covary otherwise, because the same μ and σ^2 are used for all observations on a given trial. de Gardelle and Summerfield recognized this mistake in their *SI Methods* but failed to realize that their main model prediction was a direct consequence of it. Indeed, when we simulate decisions based on the incorrect likelihood (Eq. **2**) and then perform logistic regression as in ref. 1, we find downweighting of observations x_i with a larger magnitude (Fig. 1, dashed), consistent with the model predictions in ref. 1. By contrast, the decision based on the correct likelihood (Eq. **1**) is to report blue whenever $\Sigma_i x_i$ is positive (i.e., a simple averaging rule). Thus, the optimal observer equally weights all observations (Fig. 1, solid). The moral of the story is that not applying marginalization and conditional independence rules in the correct order can have severe consequences.

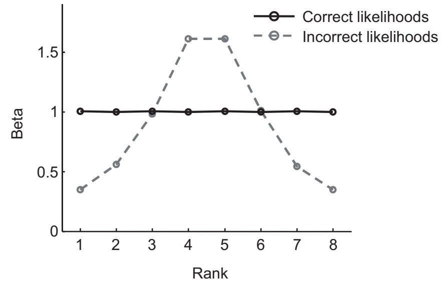

Fig. 1. Weight given to an observation (normalized logistic regression coefficient) as a function of the rank of the observation obtained by simulating the optimal observer. On each trial, observations were ranked by their values. Downweighting of more outlying values occurs only when incorrect likelihoods (grey, dashed line) but not when correct ones (black, solid line) are used.

The experimental data presented in ref. 1, if reliable, should be taken as evidence against the hypothesis that humans accumulate evidence optimally in this task. That being said, robust averaging is a known concept in perception (2) and can be optimal when the observer considers multiple possible generative processes for the data (3). It remains to be seen whether this explanation applies to the current data. An alternative explanation could be that the stimulus spaces used in ref. 1 were not perceptually uniform.

Ronald van den Berg and Wei Ji Ma[1]
Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030

1. de Gardelle V, Summerfield C (2011) Robust averaging during perceptual judgment. *Proc Natl Acad Sci USA* 108:13341–13346.
2. Landy MS, Maloney LT, Johnston EB, Young M (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res* 35:389–412.
3. Knill DC (2003) Mixture models and the probabilistic structure of depth cues. *Vision Res* 43:831–854.

Author contributions: R.v.d.B. and W.J.M. performed research and wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: wjma@bcm.edu.