nature portfolio

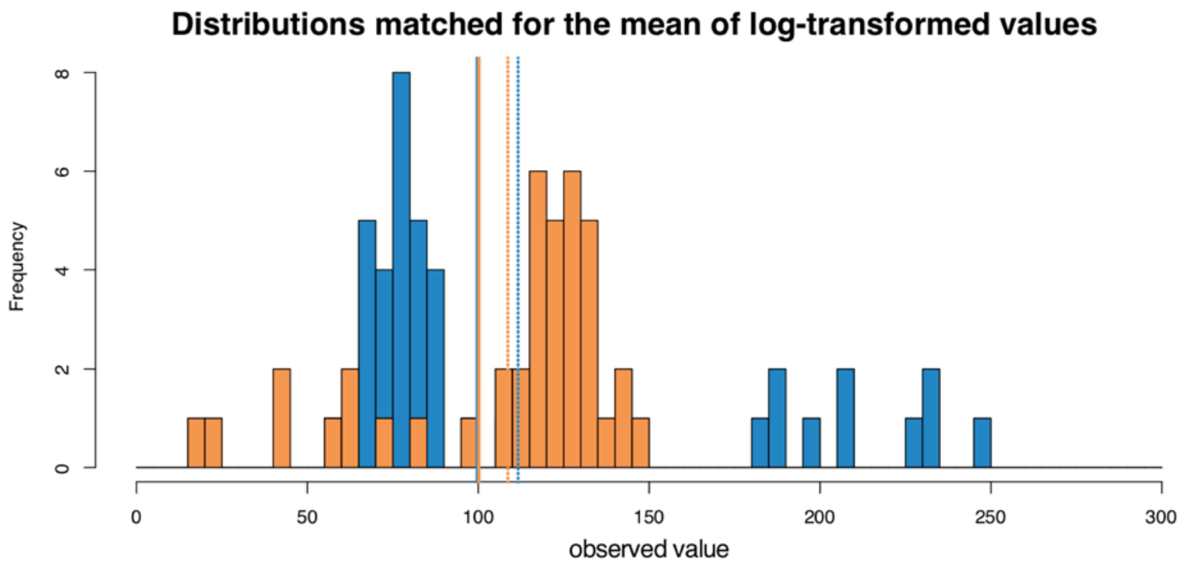
https://doi.org/10.1038/s41562-021-01065-0

Supplementary information

Biased evaluations emerge from inferring hidden causes

In the format provided by the authors and unedited

Supplementary Experiment: Estimation biases in distributions matched for the mean of log-transformed values.



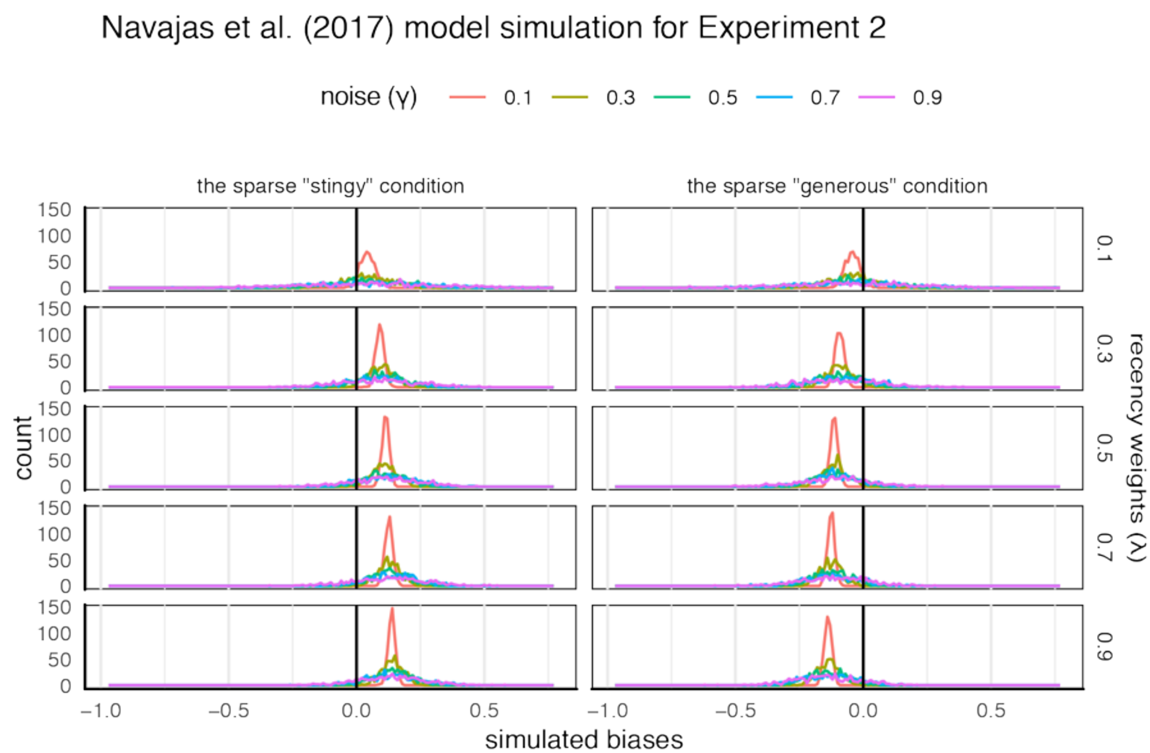
Supplementary Figure 1. Distributions matched for the mean of log-transformed values. In this experiment, the distributions for sparse "stingy" (orange) and sparse "generous" (blue) conditions were determined such that the arithmetic means of *log-transformed observed values* (solid lines) were matched (rather than matching the arithmetic means of the observed values themselves, as in all our other experiments). The true means of linear-scale values (dotted lines) were higher than the means of the log-transformed values in both conditions.

A potential mechanism for a general negativity bias is that each observation is perceived on a logarithmic rather than linear scale. In this case, small numbers would have greater relative significance than larger numbers, leading the overall estimation to be biased to below the true (linear) mean. This would predict negativity biases regardless of sparsity manipulations.

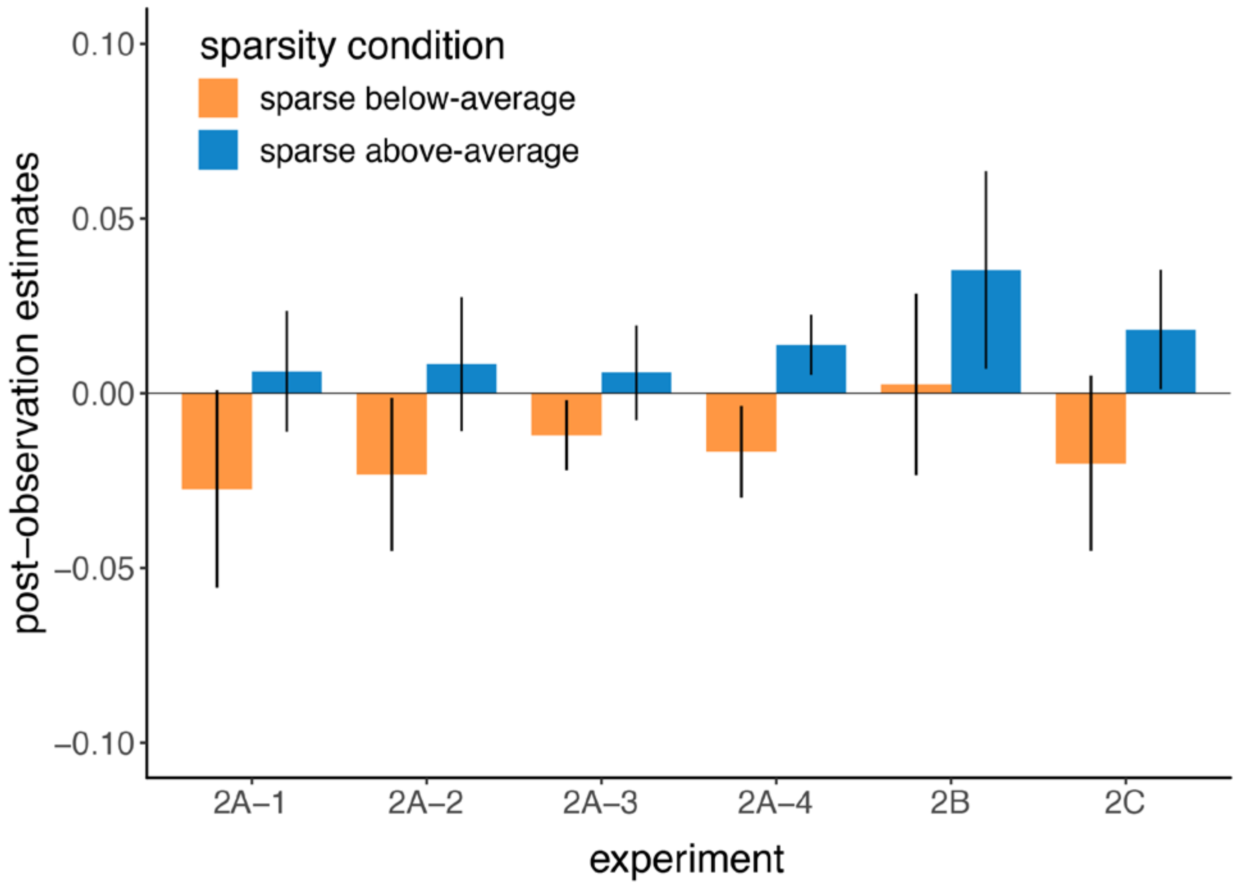
To test the logarithmic hypothesis directly, we ran an experiment (n=39) matching the two conditions (sparse "stingy" and "generous") for the arithmetic mean of log-transformed donation values at 100 (Figure S1 solid lines). The means of log-transformed values were both lower than the true linear-scale mean (Figure S1 dotted lines; sparse "stingy" condition linear-scale mean = 108.65; sparse "generous" condition linear-scale mean = 111.65). Other procedures were identical to Experiment 1, and the values were normalized by the scale (300) for the following analyses.

To test if estimation biases are driven by the logarithmic scale, we first compared the post-observation estimates to the mean of the log-transformed values. The empirical estimates were significantly different from the mean of the log-transformed values (mean difference = 0.060, $t(38) = 4.543$, $p < 0.001$, Cohen's d = 0.727, 95% CI = [0.033, 0.086]). Specifically, the estimates were higher than the mean of the log-transformed values in both the sparse “stingy” condition ($n = 19$; mean difference = 0.034, $t(18) = 2.203$, $p = 0.041$, Cohen's d = 0.505, 95% CI = [0.002, 0.067]) and the sparse “generous” condition ($n = 20$; mean difference = 0.084; $t(19) = 4.232$, $p < 0.001$, Cohen's d = 0.946, 95% CI = [0.042, 0.126]).

A key prediction of the logarithmic hypothesis is that the estimated mean would be *lower* than the true linear-scale mean, regardless of the sparsity manipulation. Both our latent cause model and the logarithmic hypothesis predict a negative bias for the sparse “stingy” condition. However, if estimation biases emerge as a result of latent-cause inference, the sparse “generous” condition should show a positivity bias that the logarithmic hypothesis cannot predict. A two-tailed t-test showed that the post-observation estimate was positively biased in the sparse “generous” condition (mean bias = 0.044, $t(19) = 2.210$, $p = 0.040$, Cohen's d = 0.494, 95% CI = [0.002, 0.085]), supporting the latent cause inference hypothesis.



Supplementary Figure 2. Alternative model (Navajas et al., 2017) simulation results for Experiment 2. In addition to a learning model that scales learning rates by surprise (Pearce & Hall, 1980; see main text), we simulated a model that takes into account the variance of observations, scaling decision noise (γ) by the variance of the observation sequence (Navajas et al., 2017), using the stimulus sequences from Experiment 2. The stimulus sequences were designed such that the values from the dense distribution were observed just prior to average estimation, thereby leading a recency-weighted model to show a density bias due to the enhanced effect of recent experiences, while leaving a latent-cause inference model to show a sparsity bias. Simulation results show the predicted density bias in both the sparse "stingy" (left) and "generous" (right) conditions for different levels of recency-weights (λ) and decision noises (γ). These predictions were opposite to our empirical results, which showed a sparsity bias.



Supplementary Figure 3. Post-observation estimates for social (Experiments 2A-1, 2, 3, 4) and non-social (Experiments 2B, 2C) experiments. In the social domain, negativity biases in the sparse "stingy"/below-average condition were marginally significant (Experiment 2A-1 M = -0.027, t(25) = -2.001, p = 0.056, Cohen's D = -0.392, 95% CI = [-0.056, 0.001]) or significant (Experiment 2A-2 M = -0.023, t(27) = -2.185, p = 0.038, Cohen's D = -0.413, 95% CI = [-0.045, -0.001]; 2A-3 M = -0.012, t(132) = -2.382, p = 0.019, Cohen's D = -0.207, 95% CI = [-0.022, -0.002]; 2A-4 M = -0.017, t(117) = -2.532, p = 0.013, Cohen's D = -0.233, 95% CI = [-0.030, -0.004]), while these biases were not present in the non-social domain (Experiment 2B M = 0.003, t(37) = 0.197, p = 0.845, Cohen's D = 0.032, 95% CI = [-0.023, 0.029]; 2C M = -0.020; t(50) = -1.605, p = 0.115, Cohen's D = -0.225, 95% CI = [-0.045, 0.005]). Positivity biases in the sparse "generous" condition were not significant in three replications (Experiment 2A-1 M = 0.006, t(43) = 0.731, p = 0.469, Cohen's D = 0.110, 95% CI = [-0.011, 0.024]; 2A-2 M = 0.008, t(38) = 0.874, p = 0.388, Cohen's D = 0.140, 95% CI = [-0.011, 0.027]; 2A-3 M = 0.006, t(126) = 0.860, p = 0.391, Cohen's D = 0.076, 95% CI = [-0.008, 0.019]) while significant in one (Experiment 2A-4 M = 0.014, t(110) = 3.196, p = 0.002, Cohen's D = 0.303, 95% CI = [0.005, 0.022]). In the non-social domain, positivity bias was significant in both Experiments 2B ($M = 0.035$, $t(42) = 2.519$, $p = 0.016$, Cohen's D = 0.384, 95% CI = [0.007, 0.063]) and 2C ($M = 0.018$, $t(49) = 2.134$, $p = 0.038$, Cohen's D = 0.302, 95% CI = [0.001, 0.035]).