Data analyses without additional exclusions

Experiment 1

Participants were still excluded as per the participants section: "6 participants were initially removed from the dataset for failing to follow experimental instructions - providing a large range for their best estimate of the giraffe's weight, or they reported estimating the giraffe's weight in tons rather than lbs. 3 additional participants (in line with departmental ethical guidelines) were excluded for not reporting their age, or reporting that they were younger than 18 years old. After these exclusions, there were 857 U.S-based Mechanical Turk workers who participated in this experiment (319 female), aged 18-76 (median = 25)." No further participants were excluded from the data though. We report the analyses exactly as in the main text.

Because responses were positively skewed, all responses were log transformed for inferential analyses, although we report the non-transformed descriptive statistics (we did not repeat the analyses with non-transformed data due to the even greater positive skew from not excluding outliers from the upper end of the distribution). We initially followed Frederick and Mochon (2012), comparing each of the experimental groups with the control condition. The mean estimate of the giraffe's weight in the control condition was 1131 lbs. (SE = 72). We replicated Frederick and Mochon's results, observing an anchoring effect in the 'lbs.-no comparison' condition (mean = 1708, SE = 170), $t(340) = 3.55$, $p < .001$, $d = 0.38$, and no anchoring effect in the 'tons-no comparison' condition (mean = 1222, SE = 153), $t(338) = 1.76$, $p = .08$, $d = -0.19$[1].

[1] The trend for a lower estimate in this condition over the control condition is driven by 6 responses under 10 lbs. (3 x 1 lbs., 1 x 2 lb., 1 x 3 lb.), which were deleted in the primary analysis.

The critical analysis concerned the ‘tons-comparison’ condition. With no exclusions, a significant (cross-scale) anchoring effect was not observed (mean = 1998, SE = 415), $t(342) = 1.20$, $p = .23$, $d = 0.13$. Excluding five participants who gave an estimate lower than 10 lbs. in the tons-comparison condition (1 x 0.5 lbs.; 2 x 1 lb.; 1 x 2.5 lbs.; 1 x 4 lbs.) reinstates the significant anchoring effect reported in the main analysis, $t(337) = 3.86$, $p < .001$, $d = 0.42$. Figure A plots the log transformed estimates in the tons-comparison condition. It is clear that these low responses are outliers, which (as outlined in the main text) we assume result from participants erroneously answering in tons rather than lbs. To complete the pairwise tests with the control condition, the ‘lbs.-comparison’ condition was compared with the control group, and again a significant anchoring effect was observed (mean = 7582, SE = 5772 – this was the condition which included an estimate of 1 million lbs. for the weight of a giraffe!), $t(342) = 4.54$, $p < .001$, $d = 0.49$.

In terms of comparing the results of the experimental conditions with the control group, the data were in line with experimental predictions. To better ascertain the relative contributions of scale consistency and a comparison question, a 2x2 ANOVA was conducted between the 4 experimental conditions. The 2 main effects were significant in this analysis. Higher estimates (i.e., a stronger anchoring effect) were observed in the presence of the comparison question than in its absence, $F(1, 682) = 7.26$, $p = .007$, $MSE = 0.27$, $\text{eta}_p^2 = .011$. There was also a main effect of scale, such that a stronger anchoring effect was observed when the anchor was on the same scale as the target judgment (i.e., lbs.), $F(1, 682) = 19.06$, $p < .001$, $MSE = 0.27$, $\text{eta}_p^2 = .027$. Unlike in the primary analyses, the interaction did not reach significance, $F(1, 682) = 1.78$, $p = .183$, $MSE = 0.27$, $\text{eta}_p^2 = .003$. It is worth noting that

---

The next lowest response after these was 100 lbs. A histogram of the log transformed responses shows these to be genuine outliers.

simply excluding the estimates under 10 lbs., or even those estimates plus those (two) of 70,000 lbs. or greater did not reinstate the significant interaction term.

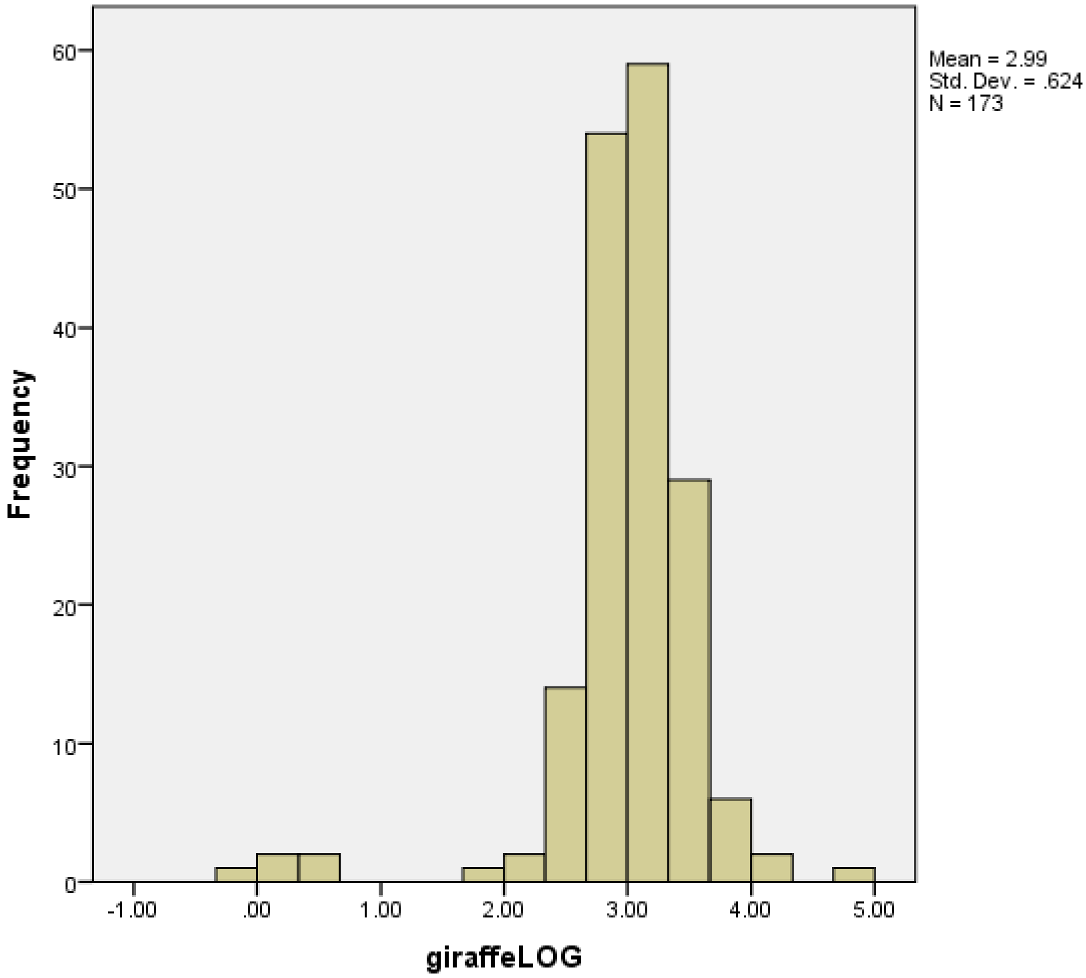

Figure A. Histogram of log-transformed estimates of the giraffe's weight in the 'tons-comparison' condition. The five estimates under 10 lbs. (to the left of the figure) are clearly outliers.

Experiment 2

The analyses (inferential statistics were performed on log-transformed data) are presented below with the impossible result of 'zero' still excluded (log10(0) is undefined). The pattern of results is identical to that reported in the main text.

Estimates of the giraffe's height were lower in the anchor condition (mean = 16.46; SE = 0.60) than the control condition (mean = 22.22; SE = 2.02), $t(322) = 4.15$, $p < .001$, $d = 0.46$.

Experiment 3

There were no exclusions from the dataset in the analyses presented below (inferential statistics were performed on log-transformed data).

A significant effect of anchor condition on participants' estimates of giraffe height was observed, $F(2, 479) = 3.48$, $MSE = 0.026$, $p = .031$, $\text{eta}_p^2 = .014$. Planned pairwise comparisons demonstrated that estimates in the low anchor condition (mean = 15.69; SE = 0.49) were lower than in the control condition (mean = 17.84; SE = 0.63), $t(331) = 2.56$, $p = .011$, $d = 0.28$. Estimates in the high anchor condition (mean = 16.46; SE = 0.61) did not significantly differ from estimates in the control condition, $t(327) = 1.66$, $p = .098$, $d = -0.18$. Unexpectedly, estimates in the high anchor condition were (directionally) lower than those in the control condition. For completeness, estimates in the high and low anchor conditions did not differ from each other, $t(300) < 1$.