When Metrics Matter: How Reasoning in Different Metrics Impacts Judgments of Uncertainty

David Zimmerman* (UCLA), Stephen Spiller (UCLA, Nicholas Reinholtz (CU Boulder), Sam Maglio (U of Toronto)

Zoom: tinyurl.com/bdfz7vdh

INTRO

Predictions come from three empirical regularities in the literature:

1) Subjective risk is positively related to both the standard deviation and the mean of the observed data[1]
2) People are insufficiently sensitive to unit when making evaluations or predictions[2]
3) People tend to assume uncertainty around a value is relatively symmetric[3]

In all studies people are asked for estimates of upper and lower bounds of 80% prediction interval (PI). Outcome metrics are PI width (90^{th} percentile – 10^{th} percentile) and PI skew ((90^{th} percentile – 50^{th} percentile) / (50^{th} percentile – 10^{th} percentile)). All participants are recruited from Amazon Mechanical Turk. Note the distribution of the outcomes were highly skewed in studies 2-3c so a log transformation was applied.

Experiment 1 (n = 167)

People estimated profit or revenue for a business with a fixed cost structure. We predict greater uncertainty (wider PIs) for the more numerous unit (revenue) vs less numerous unit (profit).

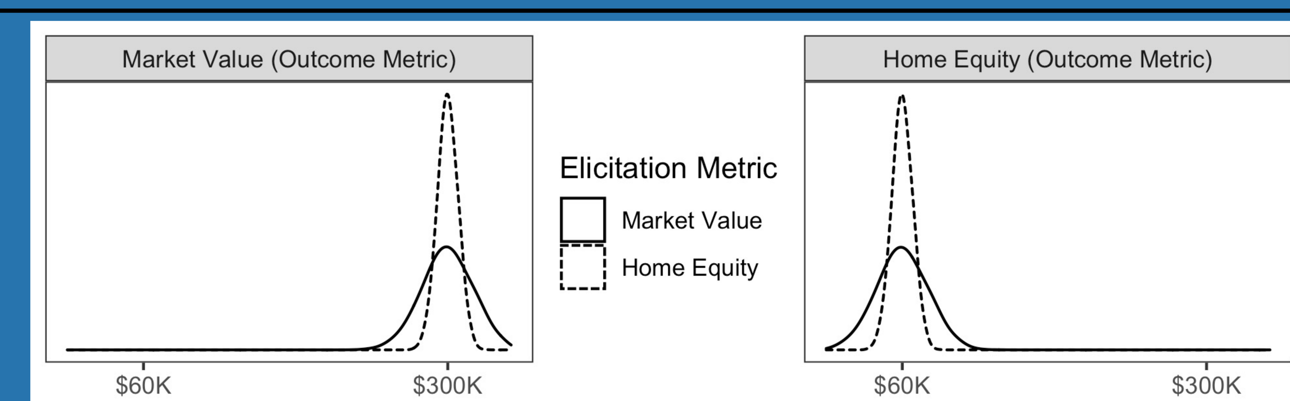
Results: Average PI width wider for revenue (more numerous unit) than profit ($M_{Revenue}$ = 1123.3, M_{Profit} = 900.8; t(165) = 2.00, p = .047, 95% CI = [1.65, 220.91], cohen's d = 0.31). No difference in skew (p = 0.57).

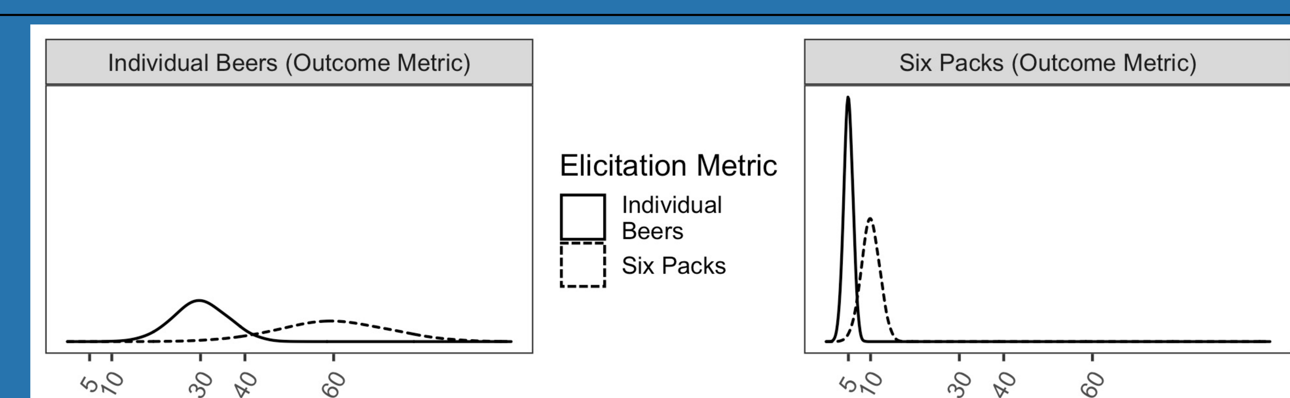
Experiment 2 (n = 239)

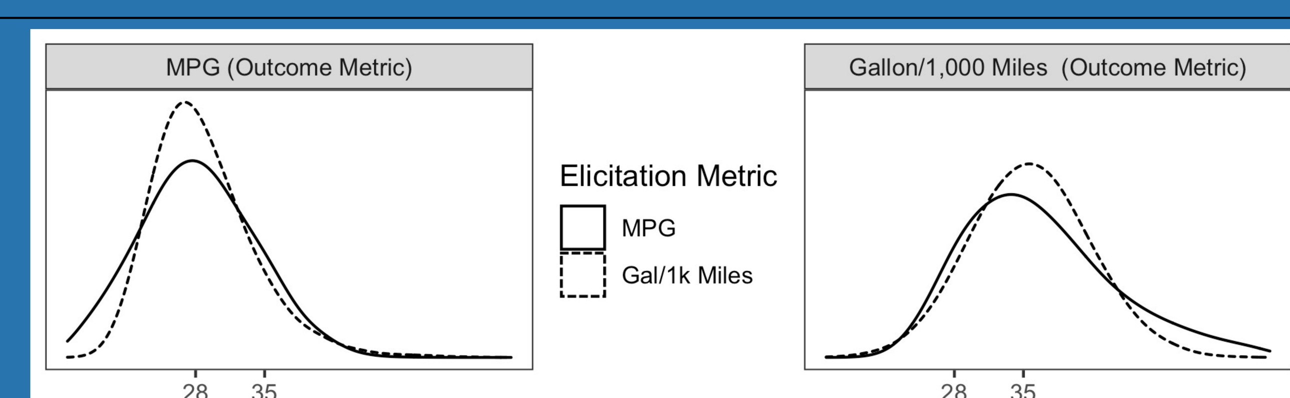
People estimated egg sales in individual eggs or dozens of eggs for a single day at a grocery store. People should rescale predictions from individual eggs to dozens by 12, but we expected their estimates will be rescaled by substantially less than 12. When multiplying the actual estimates by 12, we predict the resulting PIs will be wider in the dozens versus individual eggs condition.

Results: Average PI width wider for dozens (larger unit size) than individual eggs (M_{Dozens} = 649.9, $M_{Individual}$ = 262.8; t(291) = 3.21, p = .001, 95% CI = [138.09, 574.95], cohen's d = .38). No difference in skew (p = 0.14).

Simple **transformations** of **metric** (e.g., adding a constant, multiplying by constant, or inverting a rate) **impact** people's **estimates** of **uncertainty**.

<table>
<tr><th>Metric Relationship</th><th>Outcome Metric A   Outcome Metric B</th><th>Prediction</th></tr>
<tr><td>Units Differ by Constant (Risk Scales with Magnitude)</td><td>

Market Value = Home Equity + $240k (Mortgage Balance)</td><td>H1: Greater uncertainty for more numerous unit</td></tr>
<tr><td>Units Differ by Multiple (Unit Insensitivity)</td><td>

Individual Beers = Six Packs * 6</td><td>H2: Greater for the larger unit (six packs)</td></tr>
<tr><td>Units Inversely Related (Symmetry Assumption)</td><td>

$$\frac{Miles}{Gallon} = \frac{1{,}000}{\frac{Gallon}{1{,}000\ Miles}}$$</td><td>H3: More symmetric uncertainty for elicited metric</td></tr>
</table>

David.Zimmerman.phd@anderson.ucla.edu ResearchBox: https://tinyurl.com/mur3x82k

Experiments 3a-c (n = 99, 169, 294)

3a: People estimated their wage rates for the last 100 HITs in minutes/dollar or cents/minute. We predict more symmetric estimates in the elicited unit that become more positively skewed when inverted.

Results: All estimates converted to minutes/dollar. PIs are more symmetric in the elicited unit than the transformed unit ($M_{cents/minute}$ = 2.1, $M_{minutes/dollar}$ = 3.7, t(95) = −3.57, p < .001, 95% CI = [−0.42, −0.12], cohen's d = .72). Surprisingly, average PI widths are wider in the cents/minute metric ($M_{cents/minute}$ = 10.6, $M_{minutes/dollar}$ = 15.1; t(95) = 2.44, p = .017, 95% CI = [0.03, 0.33])

3b: People estimated the exchange rate between US dollars and Turkish Lira two weeks away. All estimates are converted to lira/dollar.

Results: PIs are more symmetric in the elicited unit than the transformed unit ($M_{cents/lira}$ = 1.6, $M_{lira/dollar}$ = 1.0, t(160) = −3.67, p < .001, 95% CI = [−0.37, −0.11], cohen's d = .58). PI widths did not differ (p = 0.79).

3c: People estimated the distribution of gasoline prices across the US in price/gallon or gallons/$40. We predict more symmetric estimates in the elicited metric than the transformed metric. All estimates are converted to price/gallon.

Results: ($M_{price/gallon}$ = 1.36, $M_{gallons/\$40}$ = 2.2, t(290) = −4.04, p < .001, 95% CI = [−0.35, −0.12], cohen's d = .47). Average PI width for gallons/$40 was greater than price/gallon (p = 0.031).

1. Hogarth, R. M. (1975). Cognitive Processes and the Assessment of Subjective Probability Distributions. Journal of the American Statistical Association, 70(350), 271–289.
2. Burson, K. A., Larrick, R. P., & Lynch, J. G. (2009). Six of One, Half Dozen of the Other. *Psychological Science*, *20*(9), 1074–1078.
3. Flannagan, M. J., Fried, L. S., & Holyoak, K. J. (1986). Distributional Expectations and the Induction of Category Structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(2), 241–256.