

Framing numerical sequences as household bills partially corrects underestimation from intuitive summation

Féidhlim McGowan, Eleanor Denny and Pete Lunn

Zoom Link: <https://tcd-ie.zoom.us/j/97678554117>

Preregistration available at <https://osf.io/87bjg/registrations>

Email fmcgowan@tcd.ie to discuss this research

Motivation

Underestimation from intuitive summation (hereafter UFIS) has been recorded in a lab experiment and an unincentivized field experiment (Scheibehenne, 2019). UFIS is thought to be caused by a compressive scaling of numbers when they are encoded internally (Dehaene et al., 2008). Experiments on intuitive arithmetic usually use abstract stimuli - clouds of dots or numbers without a concrete meaning. Studies on errors in logical reasoning have found that reframing problems into a familiar context can improve both inductive and deductive reasoning (e.g. Fiddick et. al, 2000). We tested whether familiarity would aid accuracy in intuitive summation too. We varied whether rapidly presented numerical sequences were framed as familiar household bills or as unfamiliar, abstract pay-outs from slot machines. If compressive scaling when numerical stimuli are encoded causes UFIS, framing should make no difference.

Further contributions:

1. We tested whether UFIS generalises to an alternative elicitation mechanism by using a **two-alternative forced-choice (2AFC)** task in addition to the previously deployed judgment task in which participants type their estimate for the sequence sum.
2. We manipulated **sequence length** within-subject.

Results

Hypotheses

H1a: In the judgment task underestimation will be recorded in both frames.

H1b: The null hypothesis is no relative difference in UFIS between frames. The alternative hypothesis is that UFIS will be attenuated in the bill frame.

H2a: In the 2AFC task underestimation will be recorded in both frames.

H2b: As in H1b except for 2AFC task.

H3: The magnitude of UFIS will be greater for longer sequences.

Method

Design summary: 2 (frame) x 2 (task order counterbalanced) between-subject manipulation with within-subject variation of sequence length and magnitude. Each task comprised two practice trials and 24 incentivised trials.

Bill Frame (Familiar)



You are going to see **6 bi-monthly bills** for electricity.

After viewing the sequence you will be asked to type in your best guess for the sum of all the bills.

or

Abstract Frame (Unfamiliar Slot Machines)



You are going to see **6 bi-monthly** payouts from slot machine A.

After viewing the sequence you will be asked to type in your best guess for the sum of all the payouts.

Judgment Task Details

- 24 trials in total, 3 for each of eight bills (or slot machines) in randomised order.
- Eight bills: electricity, gas, phone, TV, internet, gym, car insurance, health insurance. Slot Machines labelled A-H.
- Feedback after every eighth trial on level of average absolute accuracy.

Please type in your best guess for the total sum of the bills you just saw.

2AFC Task Details

- Six increment levels for difference between correct answer and referent: 0.15, 0.09, 0.03 and -0.15, -0.09, -0.03.
- Four trials at each increment level.
- No feedback on performance.

This is an estimate for the **sum** of all the bills you saw:

£ 482

Do you think the true sum was more or less than this amount?

Less

More

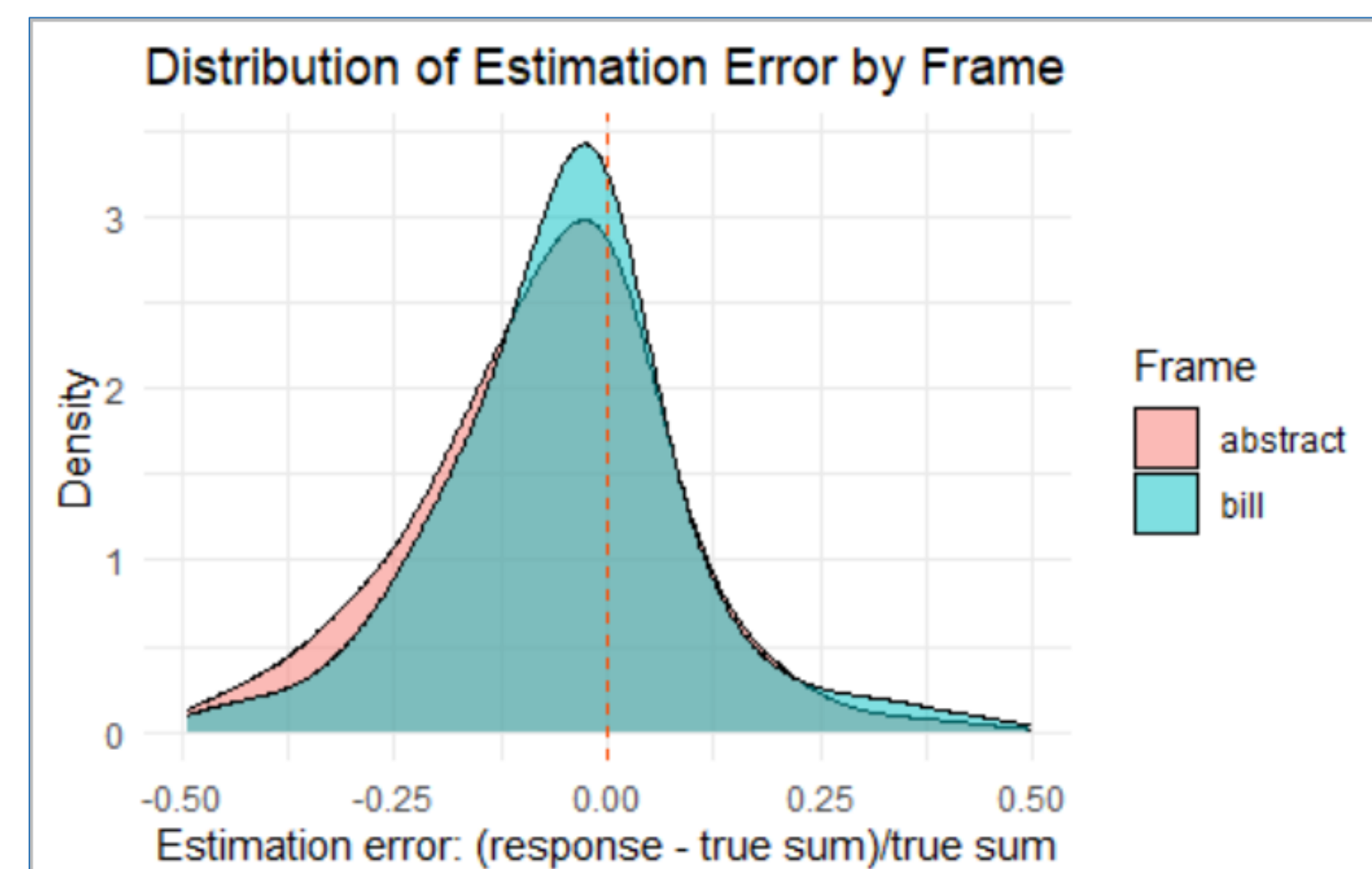


Figure 1 (left) shows the distribution of estimation error by frame. The mean of both distributions is significantly below zero, supporting H1a.

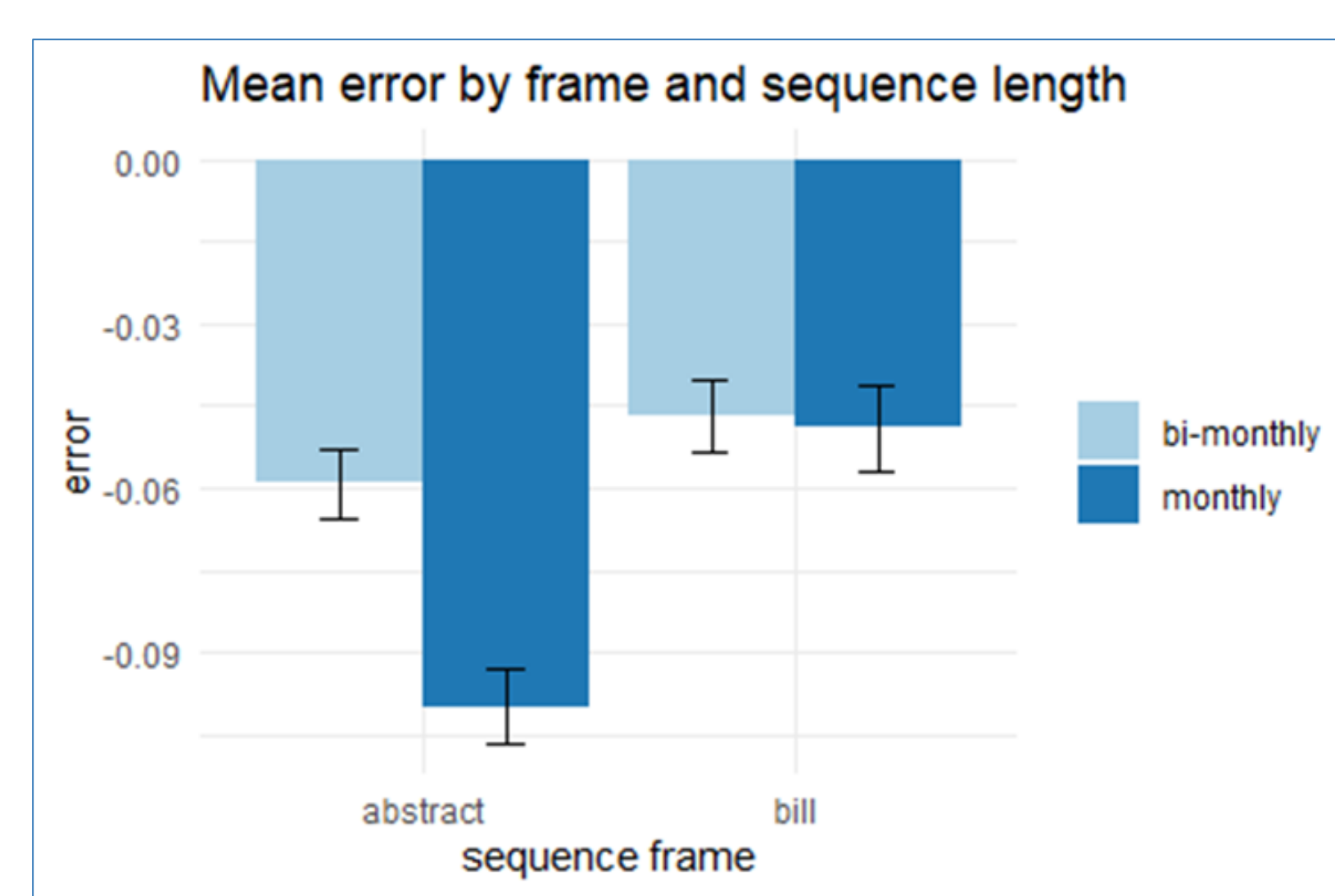


Figure 2 (right) shows the breakdown of mean estimation error by frame and sequence length. The higher UFIS in the abstract condition is driven by the longer sequences.

Figure 3 below shows the proportion of trials that were answered correctly at each of the three increment levels split by whether the correct answer was 'More' or 'Less'. UFIS predicts a higher correct rate on 'Less' trials. The point estimate supports this, but the difference is significant only when the increments are pooled (binomial test of proportions, $z = 2.296$, $p = 0.0108$, one-tailed).

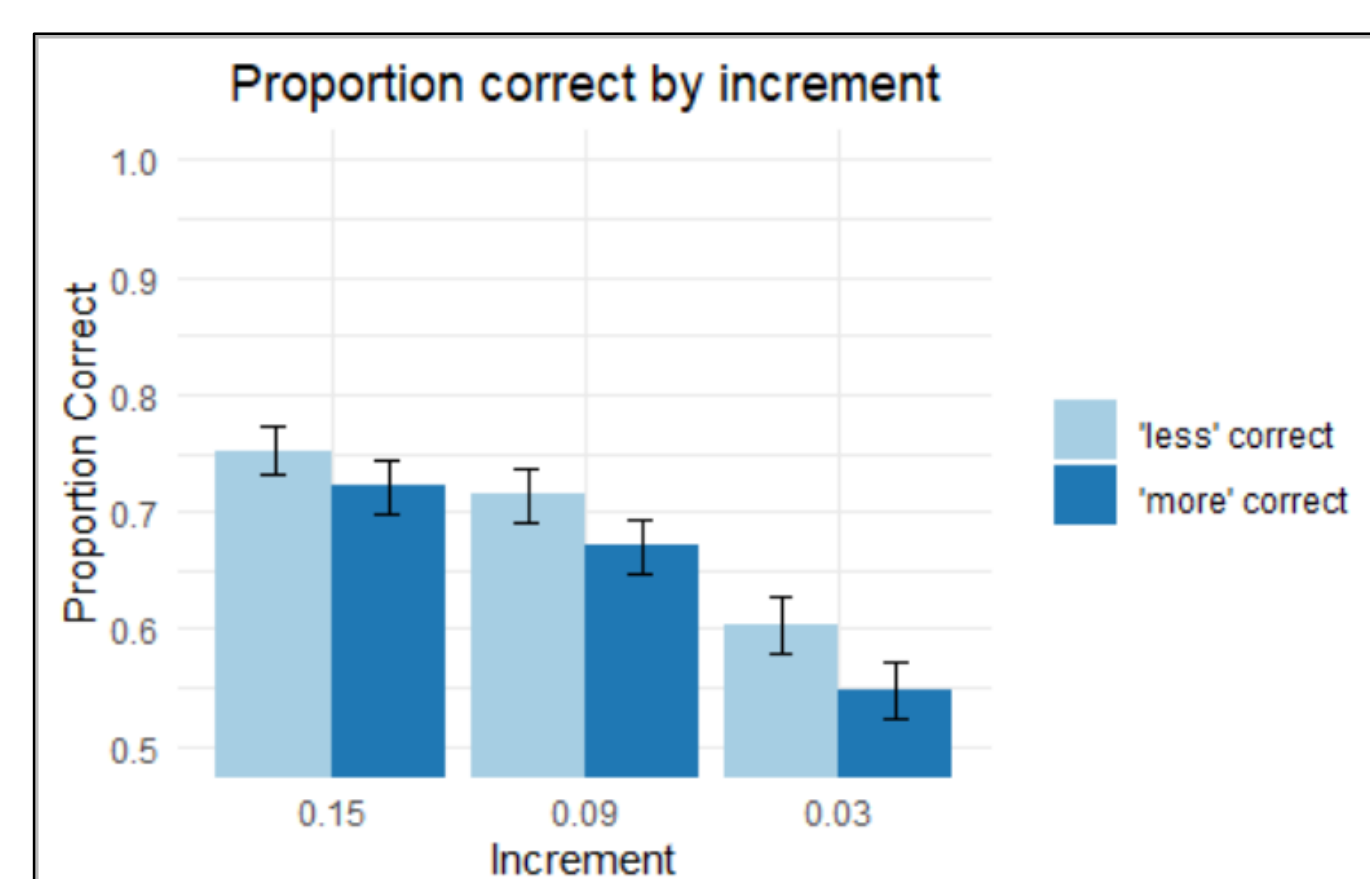


Figure 4 below shows the different pattern of response accuracy by trial type by frame. H2a was not supported. In the bill frame, accuracy varied only slightly, being highest for bi-monthly sequences where 'More' was correct. The abstract frame recorded significant UFIS in comparison, with performance significantly better when UFIS led to the correct answer i.e. 'Less' (binomial test of proportions, $p < 0.01$).

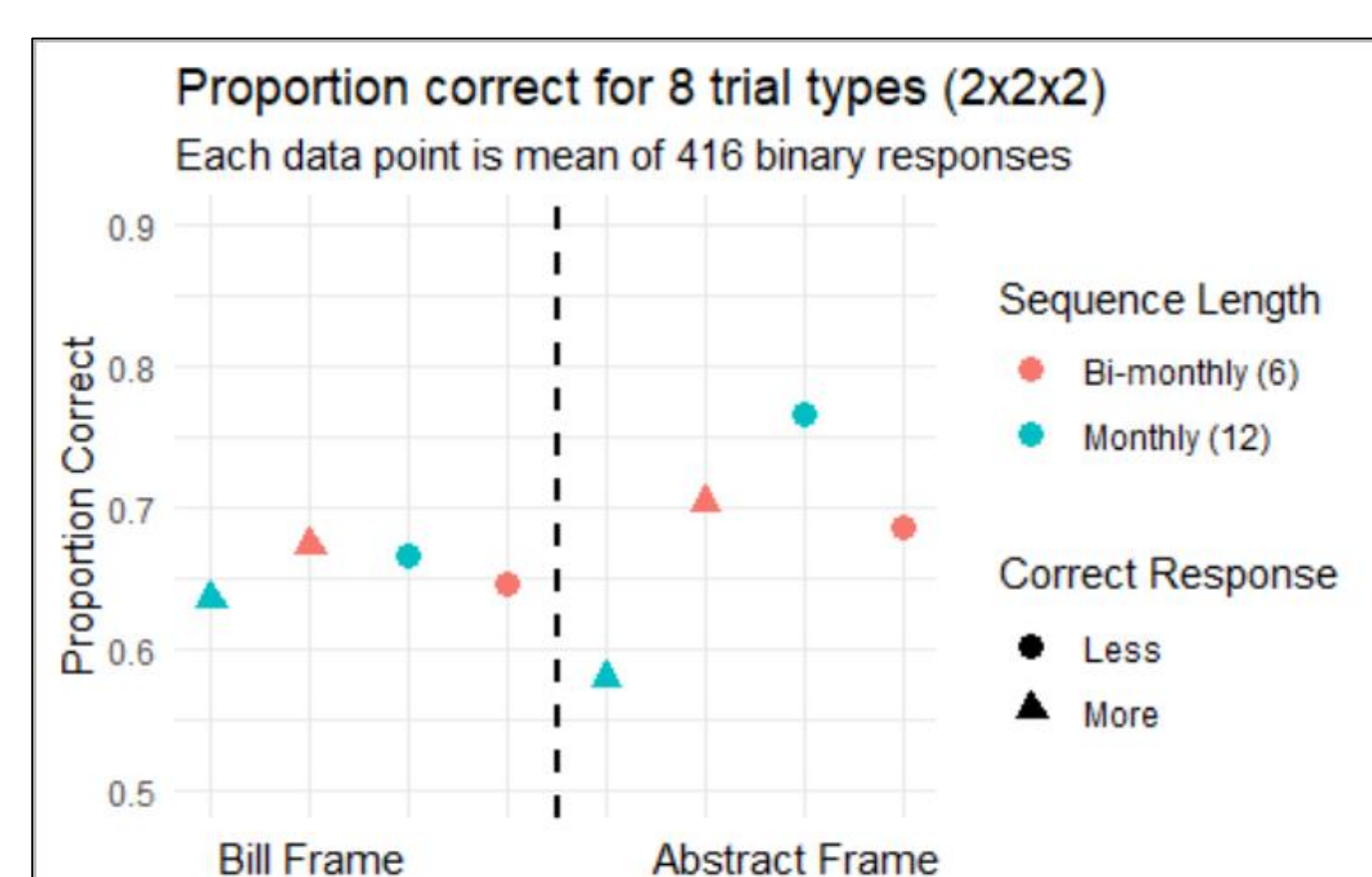
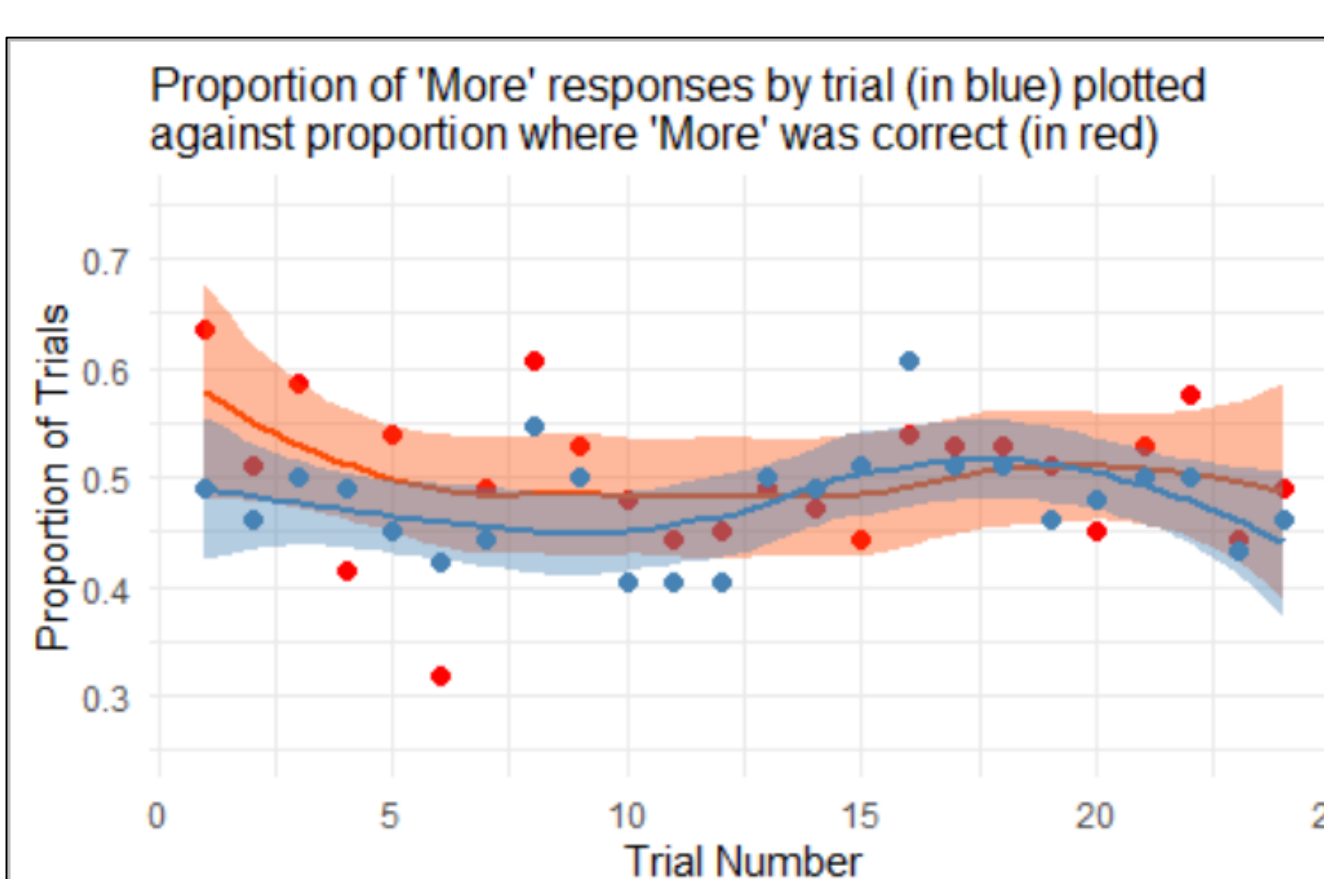


Figure 5 below shows how the proportion of 'More' responses changed in aggregate across the 24 trials. After the half way point the proportion of More responses increased. This may have been due to participants correcting for a tendency to choose less given a prior that the correct response was evenly split between the two options. This would mean the recorded UFIS was potentially a lower bound.



Discussion

Support for H1a constitutes a replication of the laboratory experiment in Scheibehenne (2019), except using a non-student sample and conducted online. The magnitude of underestimation was marginally lower in the bill frame, and the difference remains marginal in unreported regression analyses, so on balance we fail to reject H1b. In the 2AFC task, significant UFIS was observed only in the abstract frame, meaning H2a was not supported. Bill frame participants were less biased but also marginally less precise. This 2AFC task was included to test whether UFIS was an artefact of participants typing in their estimate – after all, in many situations a subjective perception need not be articulated before making a decision. Finally, H3 was strongly supported: in both tasks UFIS was greater for the longer monthly sequences, which had the same sum on average as shorter bi-monthly ones. UFIS was also stronger for larger sums, but space constraints precluded exploring that in this poster. That UFIS might be stronger for longer sequences is potentially pertinent to policy debates on whether pricing structures are benign or if they influence decision-making.

References

- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigenous cultures. *Science*, 320(5880), 1217-1220.
- Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain-specific representations and inferences in the Wason selection task. *Cognition*, 77(1), 1-79.
- Gounville, J. T. (1998). Pennies-a-day: The effect of temporal reframing on transaction evaluation. *Journal of Consumer Research*, 24(4), 395-408.
- Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *The review of economic studies*, 62(4), 515-539.
- Scheibehenne, B. (2019). The psychophysics of number integration: Evidence from the lab and from the field. *Decision*, 6(1), 61.
- Wichman, C. J. (2017). Information provision and consumer behavior: A natural experiment in billing frequency. *Journal of Public Economics*, 152, 13-33.

It also casts doubt on the assumption that more frequent billing reduces price misperceptions (Wichman, 2017). Future research could explore whether UFIS is a better explanator than mental accounting for the successful pennies-a-day marketing technique (Gourville, 1998). Within the domain of household finance, UFIS may also be a contributing factor to low switching rates. Consumer inertia is generally attributed to high search costs and hassle costs of switching (Klemperer, 1995) but underestimating the total current outlay could also play a role.

Conclusion

This experiment found evidence to suggest that in a controlled setting with incentives for accurate responses, people tend to underestimate sequence sums. Future research should explore why UFIS was greater for longer sequences, and why the pattern of responses differed across frames in the 2AFC task. One possibility is that people consulted their priors for bills, which had a moderating effect on accuracy.