It's Tough to Make Predictions, Especially About the Future:

On Interventions to Enhance the Crowd Within for Prospective Forecasts

David A. Illingworth PhD[1], Michael R. Dougherty PhD[1], Serra A. Erbas[1], & Rick P. Thomas PhD[2]

[1]University of Maryland, College Park, [2]Georgia Institute of Technology

Abstract

Three experiments examined methods for improving forecast accuracy and the accuracy of retrospective estimation. In Experiment 1, participants twice forecasted season statistics for the University of Maryland men's basketball team. Between estimates, participants consulted with two other judges, considered the perspective of a rival supporter, or performed no additional task. We found a crowd within effect where aggregate forecasts were more accurate than individual forecasts. Experiments 2 and 3 evaluated interventions for enhancing the crowd within. Experiment 2 found dialectical bootstrapping improved prospective and retrospective accuracy, while Experiment 3 found that inducing forgetting resulted in a similar improvement.

Background

- Judgment accuracy is often improved by pooling estimates across multiple judges—a phenomenon referred to as the "wisdom of the crowd" (Galton, 1907)
- Recent work suggests aggregating multiple estimates from the same judge results in improved accuracy (Vul & Pashler, 2008), including instances where judges explicitly consider alternative estimates (Herzog & Hertwig, 2009)
- Collaborative sharing of information has been promoted to improve judgment (Mellers et al., 2014); however, such collaboration may undermine the independence of judges necessary for the wisdom of the crowd (Lorenz et al., 2011)
- Others have shown that a benefit for mentally simulating the knowledge of a second judge via perspective-taking (Regan & Totten, 1975)

Method

Experiment 1

Participants were undergraduate students at the University of Maryland (N = 185)

- Task was to forecast the distribution of game statistics for the University of Maryland men's basketball team (2019)
  - Offensive points, points allowed, free throws, blocks, rebounds
- Each statistic distribution was estimated twice: once prior to an intervention, once after an intervention (3 between-subject conditions)
  - None—work alone to complete forecasts twice
  - Collaboration—discussion and information sharing with 2 additional judges
  - Perspective taking—simulating the knowledge of a second judge

Experiments 2 & 3

Participants were recruited from Mechanical Turk (Exp 2; N = 299) and Qualtrics (Exp 3; N = 2,458)

- Task was to estimate the answer for both prospective forecast and retrospective queries (Table 1)—R5 omitted from Exp 3 and prospective dates moved to January 1, 2021
- Each focal query was answered 5 times in Exp 2, and 3 times in Exp 3.
- Participants completed an intervening task between each set of responses to focal queries (4 conditions)
  - None—complete sets of focal estimates with no intervening task
  - Dialectical bootstrapping—explicitly consider reasons for error and view previous estimate
  - Modular arithmetic—complete 20 problems between sets of focal estimates
    - Changed to Elicitation—"largest plausible" and "smallest plausible"—in Exp 3
  - Distractor questions—complete 20 distractor geopolitical queries between sets of focal est.
    - Changed to Interference—entering values similar to focal estimates—in Exp 3

Table 1. Focal prospective and retrospective queries

P1	How many fatalities will be caused by novel coronavirus 19 in the US by November 1, 2020?
P2	What will the value of the Dow Jones Industrial Average be by November 1, 2020?
P3	How many acres in the Western United States will burn during the 2020 fire season before November 1, 2020?
P4	What will the price of oil be by November 1, 2020?
R1	How many members of Britain's Parliament voted in favor of Brexit in December 2019?
R2	How many migrants were arrested at the US-Mexico border in fiscal year 2019?
R3	How many members of the US House of Representatives voted to impeach Donald Trump?
R4	How many people marched in protest in Hong Kong in the largest demonstration in 2019?
R5	How many acres of the Amazon Rainforest burned in 2019?

Discussion

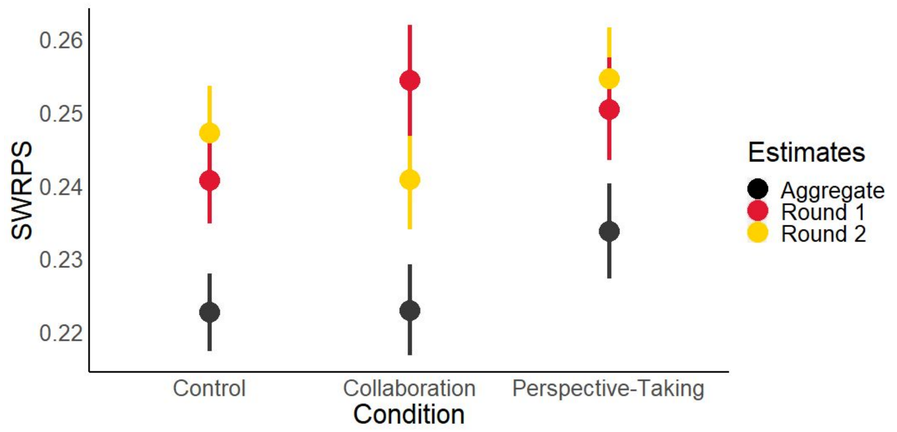
- Practices such as collaboration and perspective taking did not benefit prospective forecast
- Aggregating multiple estimates from an individual resulted in an improvement in accuracy
- Results were mixed for interventions intended to enhance the crowd within
- Evidence supports both dialectical bootstrapping and introducing interference by eliciting responses similar to estimates for focal queries
- Though significant, sizes for the reported effects are small

Results

Experiment 1

- **Neither collaboration nor perspective-taking improved on the control**
- **Aggregate estimates (mean of 2 est) reduced error regardless of intervention (Fig. 1)**

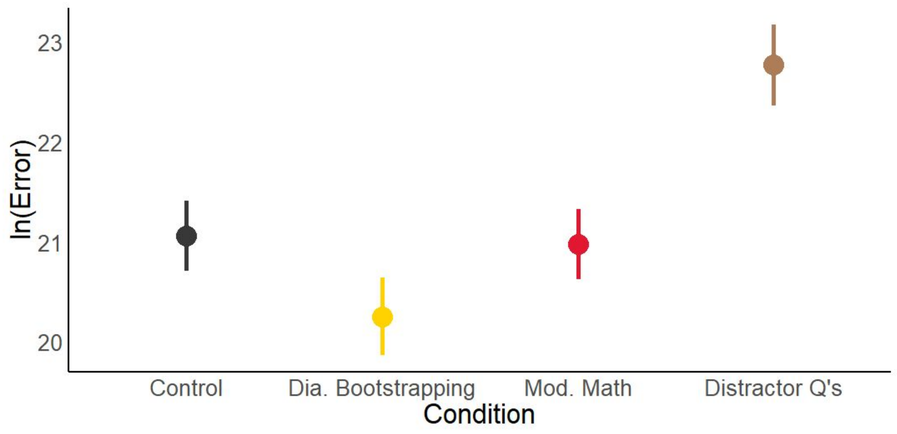
Figure 1. Mixed Model Estimates of Mean Error by Estimate and Condition



Experiment 2

- **Dialectical Bootstrapping was the most effective intervention for enhancing the crowd within (Fig 2)**

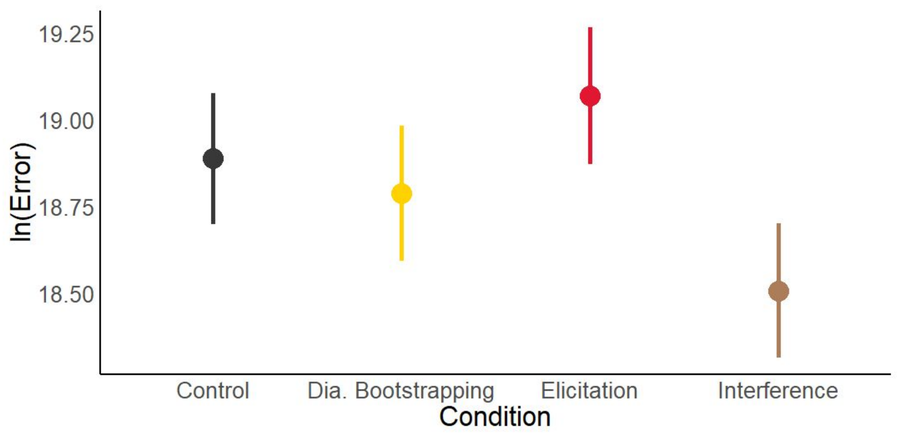
Figure 2. Mixed Model Estimates of Mean Aggregate Error by Condition



Experiment 3

- **Interference (induced forgetting) was the most effective intervention for enhancing the crowd within (Fig 3)**

Figure 3. Mixed Model Estimates of Mean Aggregate Error by Condition



SJDM Poster session link

- https://umd.zoom.us/j/99894488611

Contact

- **davidai@umd.edu**

References

Galton, F. (1907). Vox populi. *Nature, 75*(1949), 450-451. • Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20*(2), 231-237. • Lorenz, J., Rauhut, H., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *PNAS, 108*(22), 9020-9025. • Mellers, B., et al. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological science, 25*(5), 1106-1115. • Regan, D. T., & Totten, J. (1975). Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology, 32*(5), 850. • Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science, 19*(7), 645-647.