



A Crowd of 'Crowds Within': Improving Aggregated Crowd Accuracy in a Small Team Counterfactual Forecasting Task

Colin Widmer^{*1}, Amy Summerville¹, Alice Leung², Noelle Creagh¹, Andrea Humez², Ion Juvina³, Fred Bernardin², & Brandon Minnery¹

¹ Kairos Research

² Raytheon BBN Technologies

³ Wright State University

* Presenting Author



Abstract

Wisdom of crowd effects are well documented to produce more accurate estimates than those of any one estimator; they can be further improved by aggregation strategies. Additionally, multiple estimates from a single individual can serve as a 'crowd within' that is more accurate than a single estimate. In a recent counterfactual forecasting competition, small crowds completed a structured reasoning process. The judgments of each team member were aggregated to determine a final team answer. Aggregations incorporating multiple judgments from each team member (a 'crowd within') were found to be superior to other aggregation strategies, including performance- and behavior-based weightings.

Task

Participants performed counterfactual (i.e., "what if") analysis of simulated games following the IMPACT process, an analysis methodology developed in the context of the larger research program

Participants completed a 1-hour training on one of five possible simulations (CriticWorld, Stratego, Hanabi, Pathwayz, or ABStreets), as well as a 3-hour training and practice session on the IMPACT process

Participants were assigned into teams of 3 and completed 12 hours of analysis divided across six days, which included 2 hours of team synchronous work time every other day

Each team answered 18 questions in total, divided into sets of 6 questions for each of 3 counterfactual scenarios

IMPACT Process

Step 1: Assess the Question

- Participants engage in an "outside view" analysis of background materials
- Identify outcomes of interest and values in factual timelines
- Take stock of available information on possible outcome ranges and relevant historical cases

Step 2: Causal Analysis & Initial Forecast

- Lay out a causal model for the outcome of interest
- Indicate the direction and magnitude of at least five factors
- Forecast the single most likely outcome, as well as the distribution of likely outcomes across provided answer bins

Step 3: Review & Revise Forecast

- Discuss results of previous steps as a team, primarily focusing on any disagreements
- Consider new evidence brought up by teammates
- Make a final revised update to forecast

Aggregation Methods & Results

Aggregation Description	Improvement Over Unweighted (- is better)	Aggregation Description	Improvement Over Unweighted (- is better)
Crowd within	-0.0066	Self-confidence rating	-0.0009
Process compliance	-0.0052	Training performance	-0.0007
Verbal engagement	-0.0047	Self & peer confidence	0.0007
Dialectical thinking words	-0.0041	Factual outcome accuracy	0.0019
Numeracy & CRT Scores	-0.0023	Peer confidence rating	0.0034

Conclusions

Although it is non-intuitive that including the earlier forecasts from participants should improve team forecasts, as both individuals and teams should generally become more accurate in their final forecasts after having thought about the problem more deeply and considered teammate viewpoints, our empirical results support the validity of this aggregation method. In the case of the IMPACT process, participants spent the bulk of their analysis time prior to making their initial forecasts, and have already performed most of their causal reasoning and analysis, so these "initial forecasts" are not simply wild guesses.

Acknowledgments