

Diversity X Size group forecasting task (March 2018). (#9192)

Created: 03/19/2018 09:23 AM (PT)

Public: 11/18/2020 05:25 AM (PT)

Author(s)

Niccolo Pescetelli (MIT) - pescetelli@mpib-berlin.mpg.de

Iyad Rahwan (MIT) - irahwan@mit.edu

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

The study investigates the effects of online information and social deliberation on individual and group first-order (accuracy) and second-order (calibration) performance in series of real life forecasting problems. Two dimensions of interest are orthogonally manipulated across groups in a standard 2x2 design. The first dimension is group diversity. Diversity has been suggested to improve group performance by increasing the variability of the solutions, information and reasoning mindsets used by the group. The second dimension is group size. Both theoretical and empirical work (Navajas et al. 2017, NHB) suggests that assigning N participants to one single group might lead to worse results than assigning them to M groups of N/M members each. Fragmenting the initial population into smaller groups has the benefit of (a) reducing the impact of potential groups failures (herding, social loafing, risky shift etc.) on the whole population and (b) increasing the independence and variability of the solutions proposed across groups (although potentially reducing the variability within groups) and the reasoning used to tackle individual problems. Following Navajas et al. methodology, we expect that aggregating consensus forecasts and final individual forecasts (see below) across small groups will lead to better forecasting accuracy than aggregating within groups or over members of large groups. We expect diverse groups to perform better than homogeneous groups and small group-based aggregation to perform better than large group-based aggregation. We do not have specific expectations on how the two dimensions will interact with each other. We expect that private revised forecasts (see below) will show reduced variability compared to private initial forecasts. Overall we expect a reduction in variability as people interact with each other (see Lorenz et al 2011) but increase in accuracy due to private evidence gathering and social processes.

3) Describe the key dependent variable(s) specifying how they will be measured.

Our independent variables of interest are (a) first-order accuracy and (b) second-order accuracy. As all forecasting problems we use refer to binary possible outcomes (e.g., Will or will not X happen?), first-order accuracy will be defined as a correct/incorrect binary variable. Second-order accuracy (calibration) will be defined as the forecast Brier score, where a Brier score of 0 represents perfect performance and a score of 2 represents worst performance. For each forecasting problem, four forecasts (stages) will be elicited from each participant: 1. initial private forecast; 2. revised private forecast (after browsing task-relevant online information); 3. final private forecast (after social deliberation); 4. group consensus forecast (forecast agreed by the whole group). Participants will be rewarded based on time spent and performance achieved (Brier). Finally, verbal group discussions (e.g., chat-logs) will be recorded to understand whether group successes and failures can be better understood by analyzing the language and arguments used by group members.

4) How many and which conditions will participants be assigned to?

We will assign participants into 4 conditions, derived from our 2x2 design. Questionnaire responses (mainly demographics, reasoning tests and personality tests) will be used to divide participants into the four conditions according to the following procedure. (A) Individual question responses will be treated as independent feature dimensions unless belonging to the same test (e.g., CRT test, AOMT test and NFC test), in which case, responses will be aggregated and scored first and the resulting score be treated as one single dimension. (B) Features will be normalized across participants. (C) Group size condition will be fully randomized. (D) Randomization in the assignment to group diversity condition - due to the nature of this variable - will proceed as follow. a) Setting ϵ and k parameters in a DBSCAN clustering algorithm so that: (1) An inner core cluster, an outer core cluster and a periphery can be defined; (2) the inner core cluster is approximately double the size of the outer core and periphery clusters (outer core and periphery being about the same size). (3) the inner core cluster includes all participants who have at least k neighbors within ϵ distance in feature space; (4) the outer core cluster includes participants whose neighbors are in the inner core cluster but do not reach themselves the minimum number of neighbors k ; (5) the periphery group includes every other participant that does not satisfy the previous two conditions. (b) participants in the inner core cluster are randomly assigned to the outer core cluster or to the periphery. This creates two groups of approximately equal size that are expected to differ in diversity (due to the outer core and periphery participants). Randomization of the inner core participants ensures that the effects found can be attributed to our manipulation.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We intend to carry out the following analyses both for individual and group forecasts:

1. Descriptive analyses
2. ANOVAs to compare means in different conditions
3. t-test for direct comparisons
4. hierarchical mixed-effect models to account for within-group correlations
5. analysis of the forecasts distribution (changes in mean and variance after each stage)
6. Comparison (using methods 1-5) between inner core participants' individual forecasts across conditions and stages to test causal effects of our manipulation.

We have not agreed on how to analyze verbal communication among participants.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We aim at keeping all participants in our analysis. Although we do not anticipate any participant to troll their groups or actively sabotaging the experiment (e.g. DOS attacks on the chat service), if this happens we will consider excluding those groups from further analysis.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

The study recruits participants from Amazon Mechanical Turk. 150 participants will be asked submitted their answers to the survey. The final sample size will be defined by how many of these respondents will submit their answers and show up on the day of the main experiment and will not drop out before the study is over. We expect to form groups of 10-20 people, depending on condition.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)