7 Supplementary Material

7.1 List of Questions

Id	*Question*	*Dynamic*	*Hard*	*Outcome*
1	Will D.C. United win the D.C. United vs. FC Dallas soccer game (Major League Soccer) on Sat Oct. 13th?	N	N	Y
2	Will the Chicago Bears win the Chicago Bears vs Miami Dolphins game on Sun Oct. 14th?	N	Y	N
3	Will the Seattle Seahawks score more than 60 points in total in the games Los Angeles Rams vs Seattle Seahawks on Sun. Oct 7th and the game Seattle Seahawks vs Oakland Raiders on Sun. Oct 14th?	Y	N	N
4	Will Italy win against Poland in the Uefa Nations league match on Sun. Oct 14th?	N	N	Y
5	Will the Houston Texans score more points in their games against Dallas Cowboys on Sun. Oct 7th than their game against Buffalo Bills on Sun. Oct 14th?	Y	N	N
6	Will Orlando City score more goals in their game against the FC Dallas on Sat Oct 6th than their game against the New England on Sat Oct 13th?	Y	Y	N
7	Will the price of Bitcoin in USD on Sat Oct 13th (EST time zone) be, at any point of the day, above 6500? (resource: https://xe.com/currencycharts/?from=XBT&to=USD)?	Y	Y	N
8	Will the end-of-day (EST time zone) closing value for the British pound against the US dollar drop below $1.32 on Mon Oct. 15th?	Y	Y	Y
9	Will Facebook's stock price quote on NASDAQ on Sat Oct. 13th go above $160? (resource: https://www.nasdaq.com/symbol/fb)	Y	N	N
10	Will one of Alphabet Inc., Facebook, Amazon, or Microsoft make an announcement on Sat Oct. 13th (EST time zone) about a security breach?	N	N	N
11	Will Theresa May propose a new Brexit plan on Mon Oct. 15th (BST time zone)?	N	N	N
12	Will Donald Trump fire a member of the White House staff on Mon Oct. 15th (EST time zone)?	N	N	N
13	Will the official YouTube video of the Taylor Swift song 'Delicate' reach 300,000,000 views before Sat Oct. 13th midnight (EST time zone)?	Y	Y	N
14	Will 'Crazy Rich Asians' have a 'Rotten Tomatoes' score (in tomatometer) above 90% on Sat Oct 13th at noon (EST time zone)?	Y	N	Y
15	Will Elon Musk tweet more than 5 times on Sat Oct. 13th (EST time zone)?	N	Y	N
16	Will Trevor Noah (a comedian) be more popular than John Oliver (a comedian) on Google Trends (https://trends.google.com/trends/) for settings (United States, All Categories, Web Search) on Sat Oct. 13th at noon (EST time zone)?	Y	Y	Y
17	Will Ryan Gosling have more than 2.4 Million followers on Twitter by Sat Oct. 13th midnight (EST time zone)?	N	N	N
18	Will more than five patents containing 'Blockchain' in the title be published on Sat Oct. 13th (EST time zone) in the online search repository of the US patent system? (resource: http://appft.uspto.gov/netahtml/PTO/search-adv.html)?	N	N	N

7.2 Hard versus Easy Split

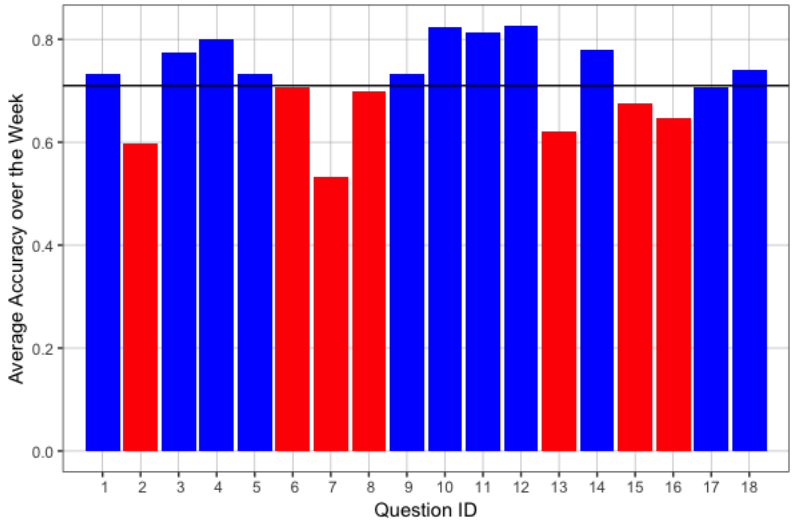
7.3 Scoring Matrix

CA uses the following scoring matrix :

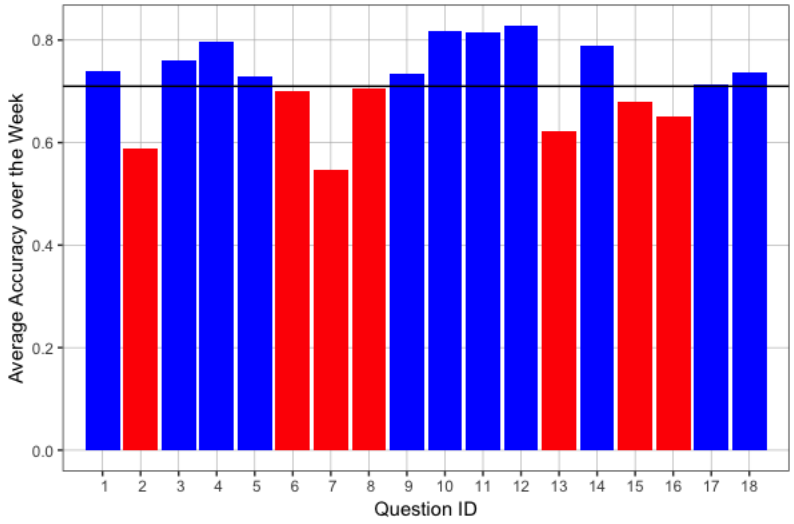
$$S(i,j) = \text{Sign}(P(i,j) - P(i)P(j)) = \text{Sign}(\Delta(i,j))$$

Here $P(i,j)$ is the joint probability of observing the signal pairs i and j and $\Delta(i,j)$ measures the correlation between the signals i and j. Suppose the signals are positively correlated i.e. $\Delta(i,i) > 0$ and $\Delta(i,j) \leqslant 0$ for $i \neq j$. In that case, CA rewards for an agreement on the same question and punishes for an agreement on two separate questions.

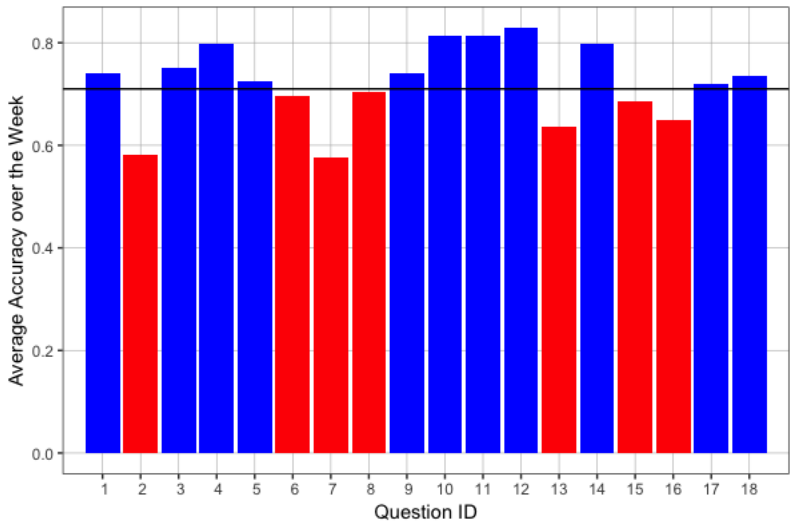
We make two adjustments to the CA mechanism described above.



(a) Average Accuracy on October 8



(b) Average Accuracy over the Whole Week



(c) Average Accuracy on Oct 13

Figure 3: The average accuracy of the predictions for the 18 questions. We computed the average accuracy for the predictions made on the first day, last day and over the whole week. Questions 2,7,13,15, and 16 have average accuracy consistenly below 0.71 and we label these questions as "hard" and the remaining questions as "easy".

1. On a given day, CA awards a given response to a question by pairing it with a response from another user. This might produce a reward with high variance, so we compute the reward by averaging the scores obtained by pairing with 100 users.
2. We defined the scoring matrix so that it takes as input two continuous forecasts instead of two discrete signals. In order to achieve this, we discretize the interval $[0, 1]$ into 10 bins and compute the delta matrix using reports from the good judgement platform Ungar et al. (2012).Figure 4 displays the scoring matrix for two reports.

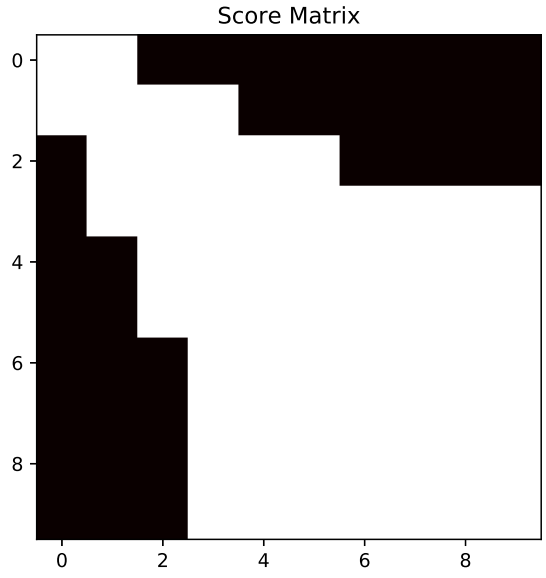
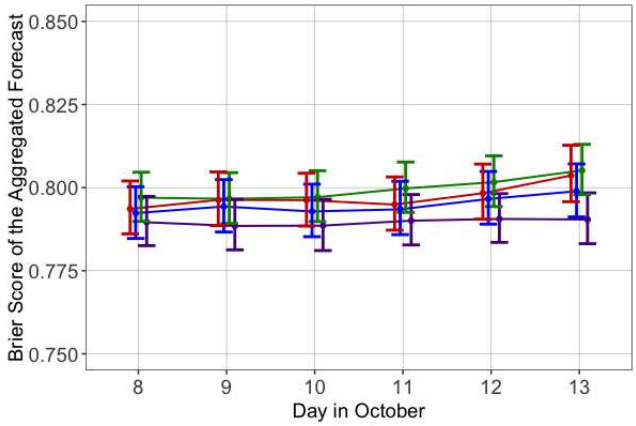
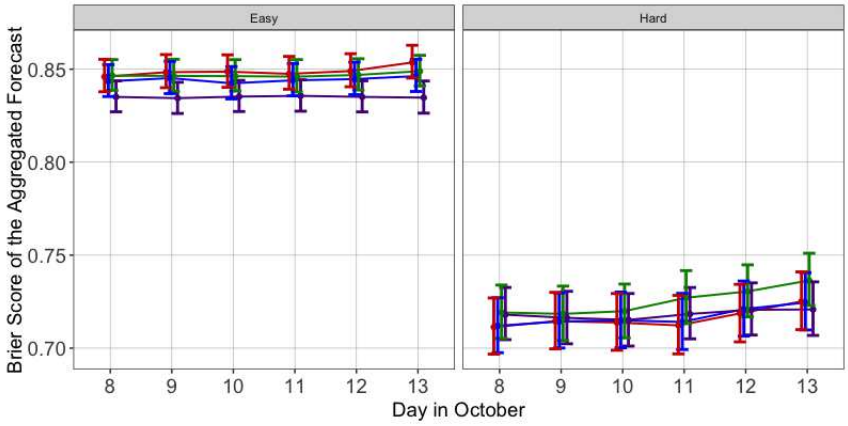


Figure 4: Score Matrix

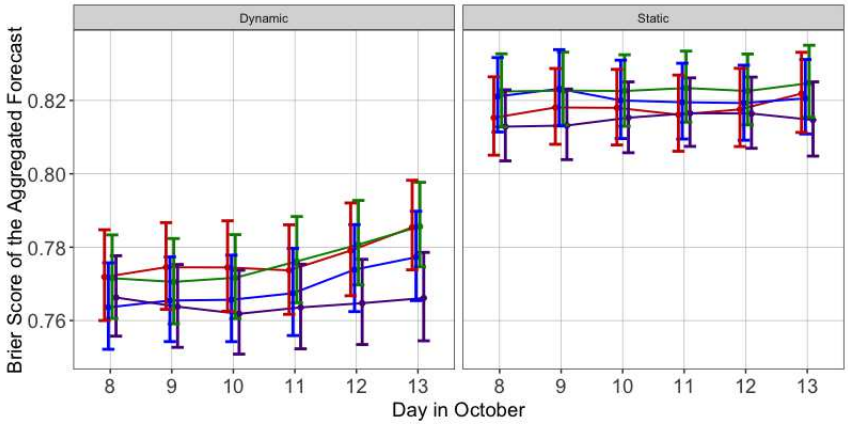
7.4 Performance of Different Aggregators



(a) All questions combined

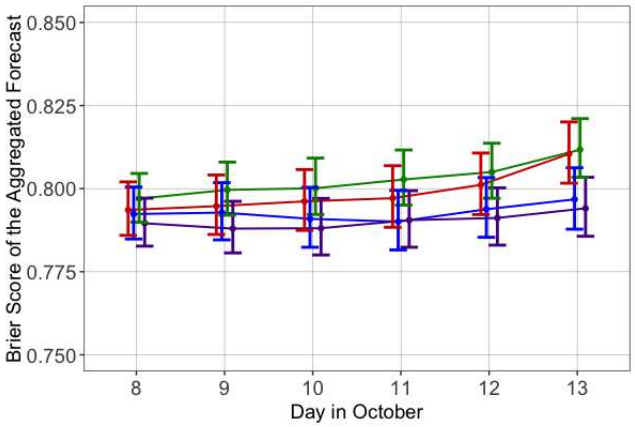


(b) Easy versus Hard Questions

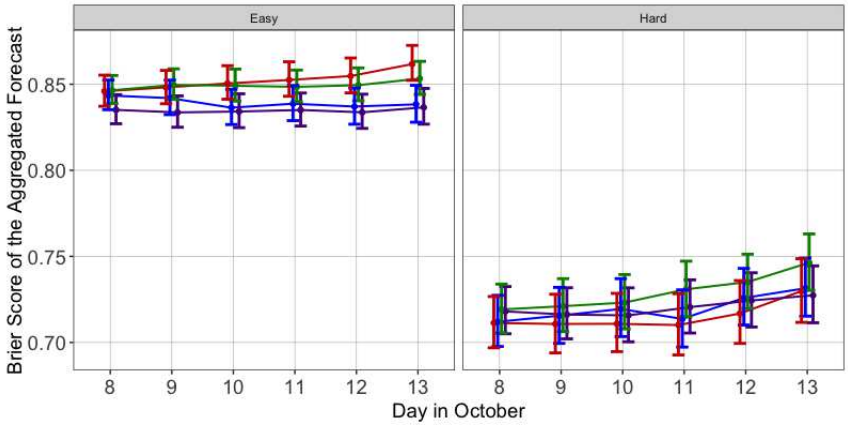


(c) Dynamic versus Static Questions

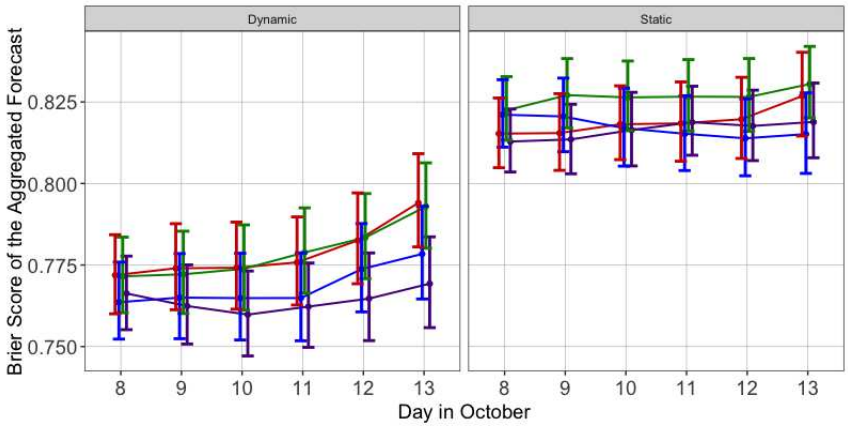
Figure 5: Aggregator 1.a: Mean



(a) All questions combined

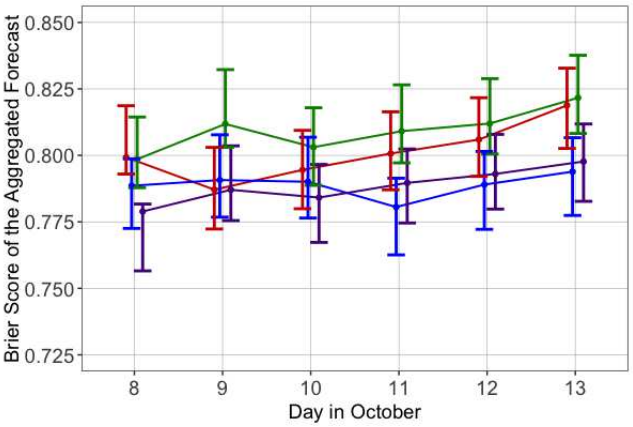


(b) Easy versus Hard Questions

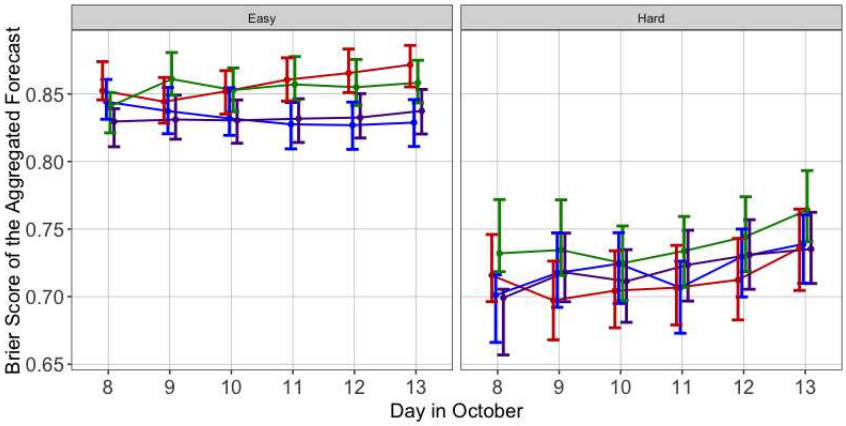


(c) Dynamic versus Static Questions

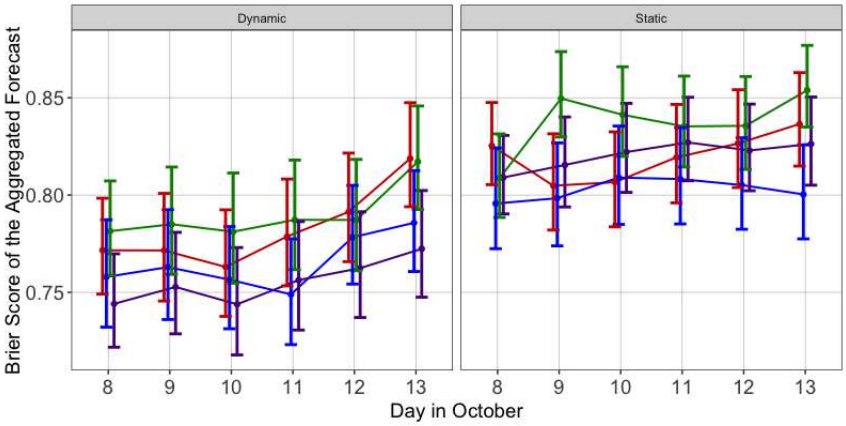
Figure 6: Aggregator 1.b: Weighted Mean



(a) All questions combined

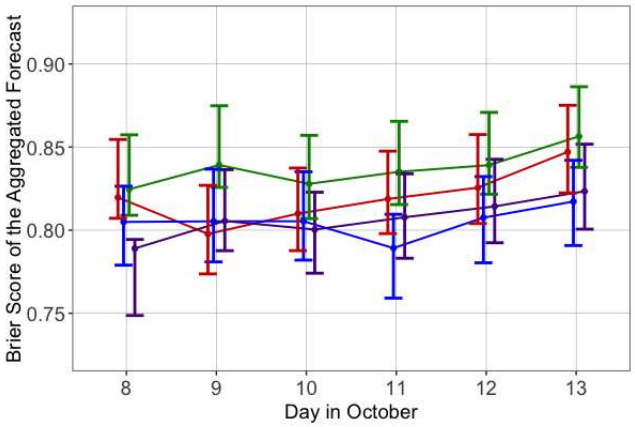


(b) Easy versus Hard Questions

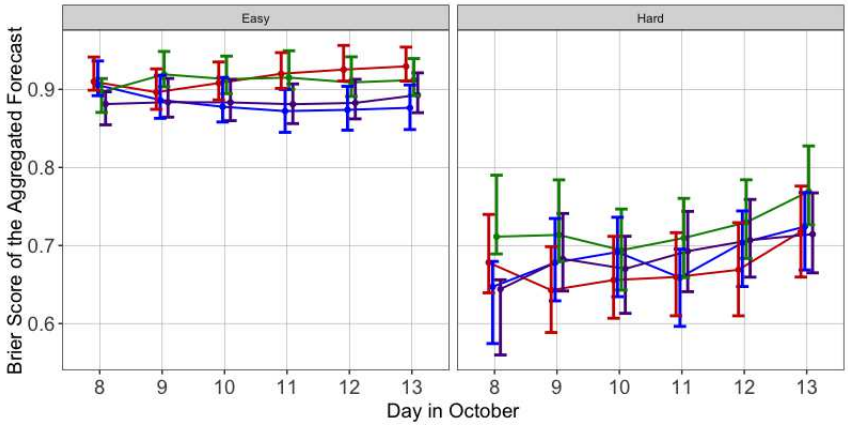


(c) Dynamic versus Static Questions

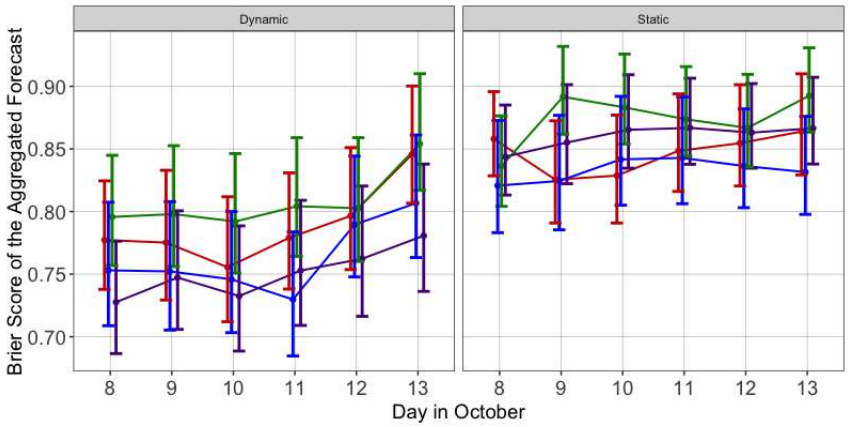
Figure 7: Aggregator 1.c: Top-k + Weighted Mean



(a) All questions combined

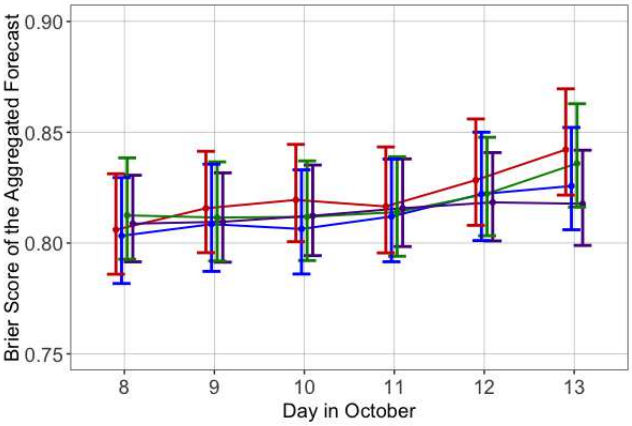


(b) Easy versus Hard Questions

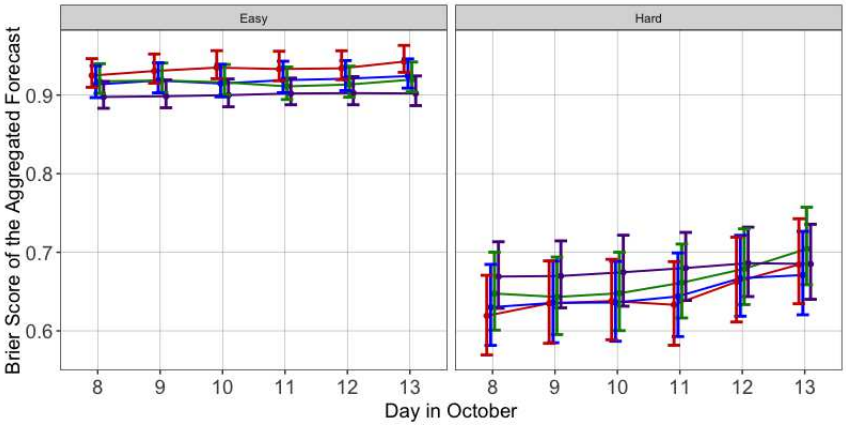


(c) Dynamic versus Static Questions

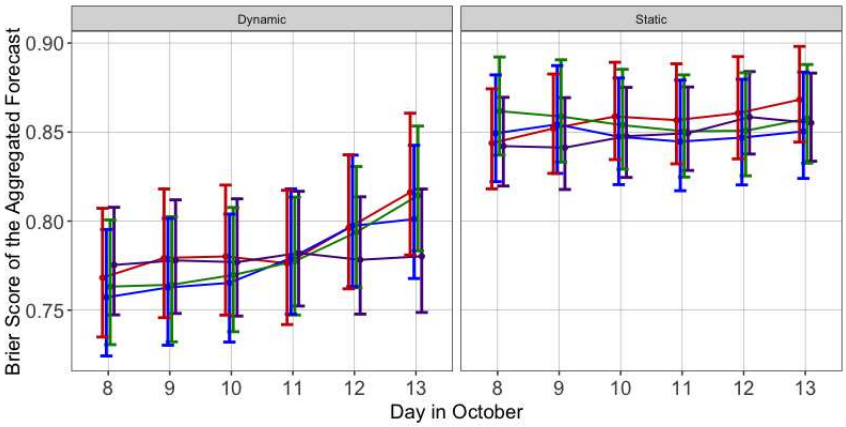
Figure 8: Aggregator 1.d: Top-k + Weighted Mean + Extremize



(a) All questions combined

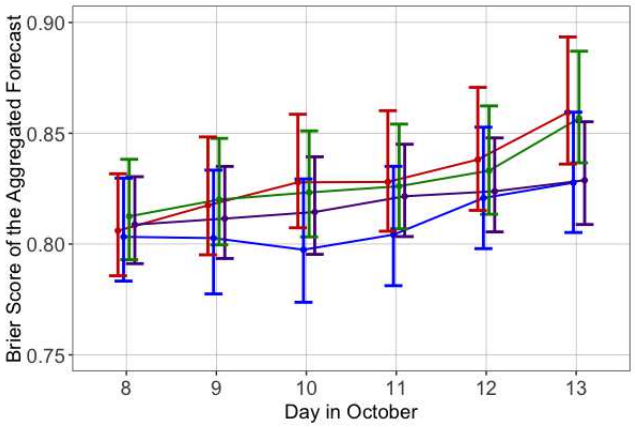


(b) Easy versus Hard Questions

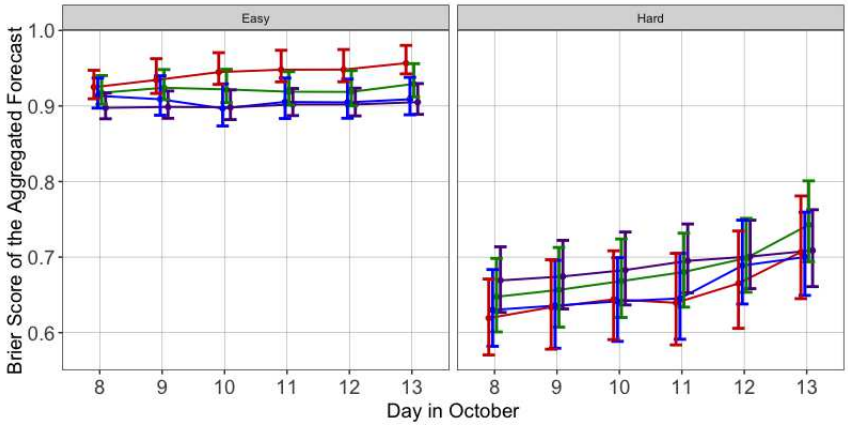


(c) Dynamic versus Static Questions

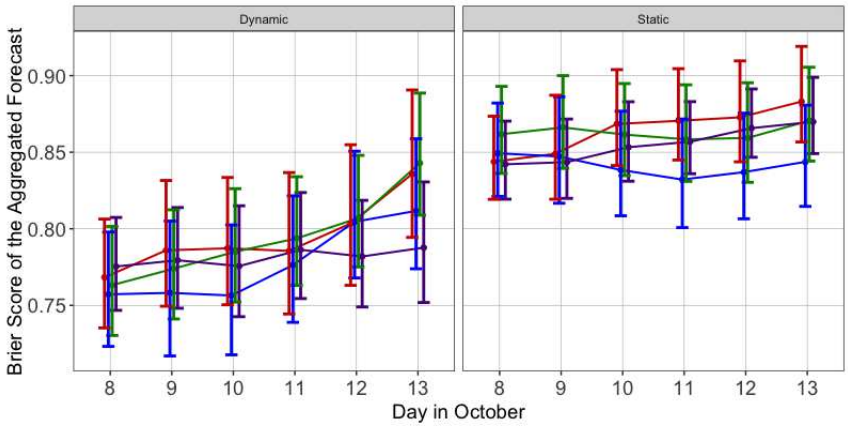
Figure 9: Aggregator 2.a: Logit



(a) All questions combined

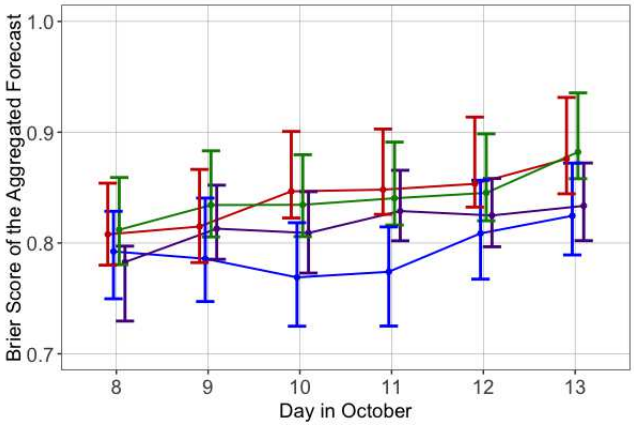


(b) Easy versus Hard Questions

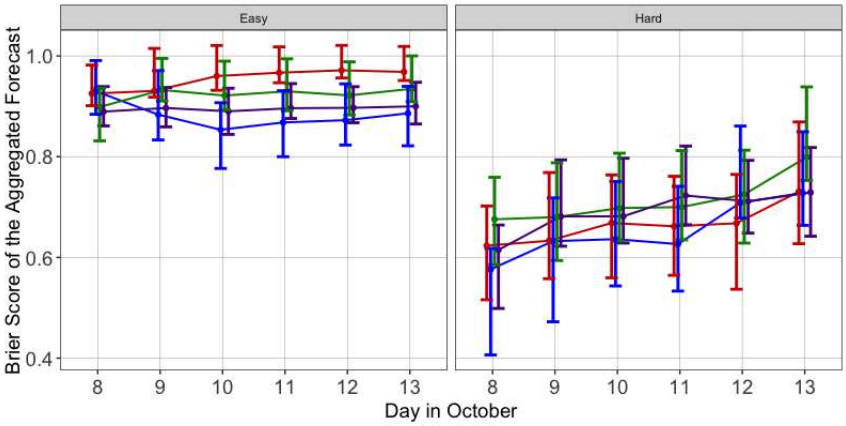


(c) Dynamic versus Static Questions

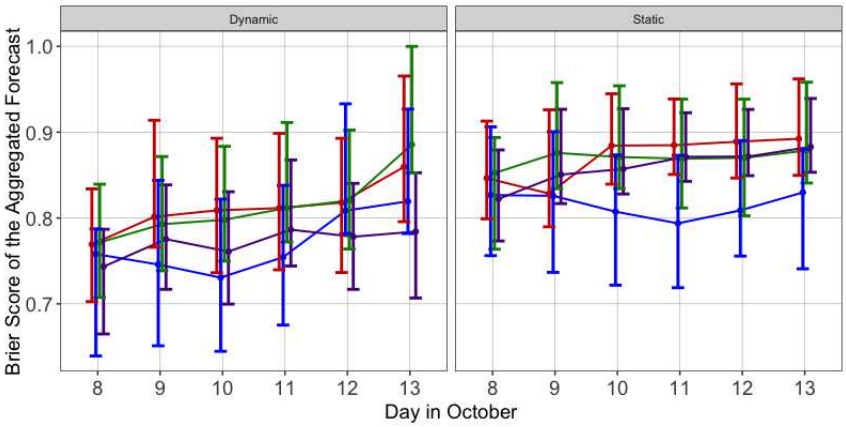
Figure 10: Aggregator 2.b: Weighted Logit



(a) All questions combined

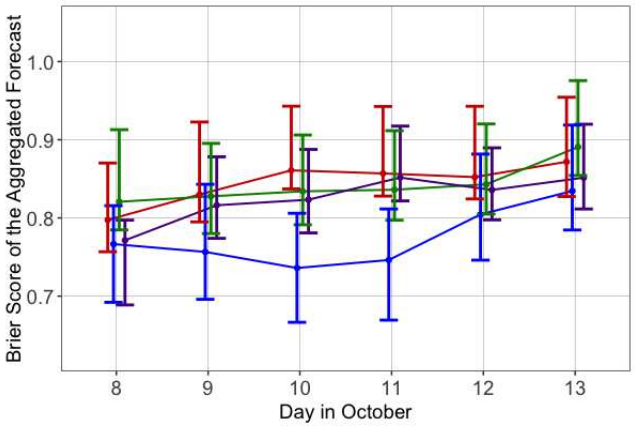


(b) Easy versus Hard Questions

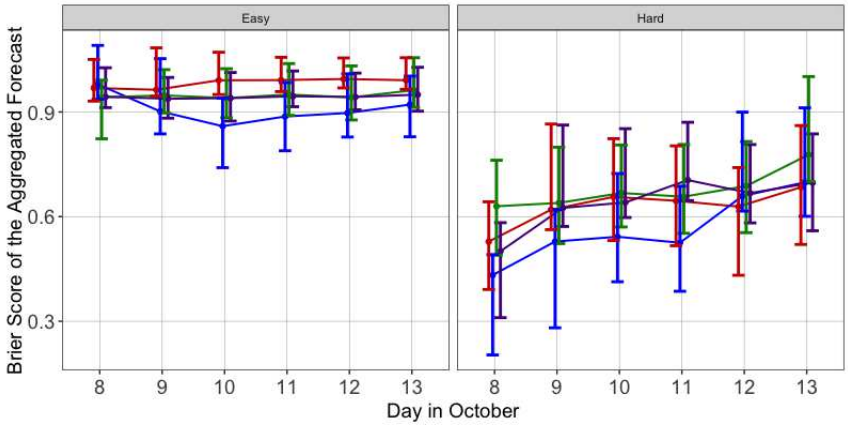


(c) Dynamic versus Static Questions

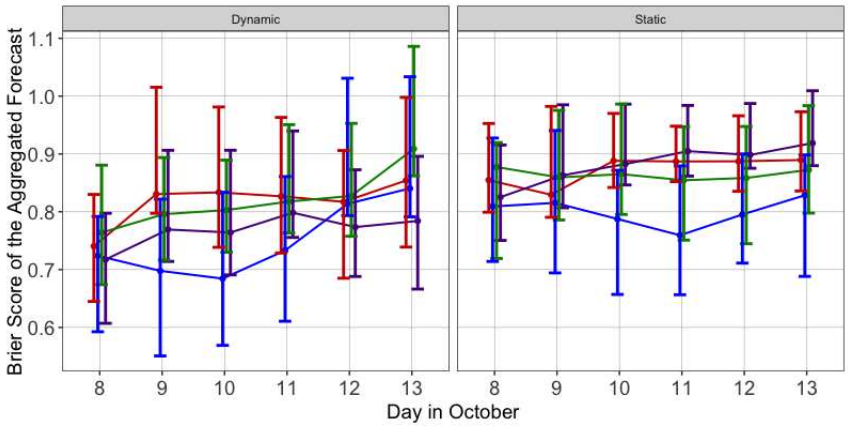
Figure 11: Aggregator 2.c: Top-k + Weighted Logit



(a) All questions combined



(b) Easy versus Hard Questions



(c) Dynamic versus Static Questions

Figure 12: Aggregator 2.d: Top-k + Weighted Logit + Extremize