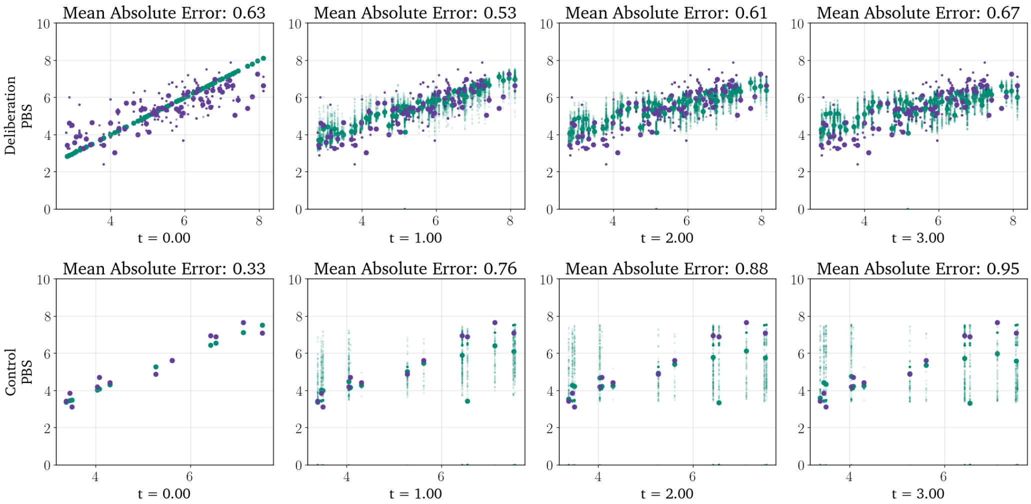

FIGURE 1: PBS, purple indicating the PBS score after deliberation in the original data, green indicates the results of the simulation in that time step.

0.1 DeGroot Model

0.1.1 PBS Scores

We first proceed with analyzing the performance of the DeGroot model on substantive agreement. Figure 1 shows the PBS of both the deliberation and control group, and the simulation results for both instances. As mentioned in **??**, the PBS (PBS) is the average of the 26 most polarizing questions, where a low PBS corresponds to more liberal answers, and high PBS indicates more conservative answers. As expected the model performs poorly at predicting the control group, as there was no significant change for control group members. Similar to the control group, the starting PBS of the participants is a strong indicator for their PBS. Therefore the model at t=0 is already reasonably aligned with the final PBS. We see that after the first time step the PBS scores get predicted more accurately, after which the model starts making larger prediction errors. This is because the model keeps averaging all opinions until a steady state is reaches in which most voters hold non-extreme positions. Whether this is positive depends on reality, as **?**] remarked, if deliberation is able to reach full consensus, the model might give a glimpse into how this works. If this is not the case however, then the model is overly naive suggesting that people come to hold a weighted average of all original opinions. The latter seems more likely.

Figure 2 shows the change in PBS for the deliberation group, the original data shows most change happens in people with high PBS, getting lower PBS and thus becoming less extreme. The model does not capture this effect, showing the most change for people with low initial PBS in later time steps. This might be because there is a correlation between PBS score and knowledge. As shown by **?**], most extreme voters, in terms of

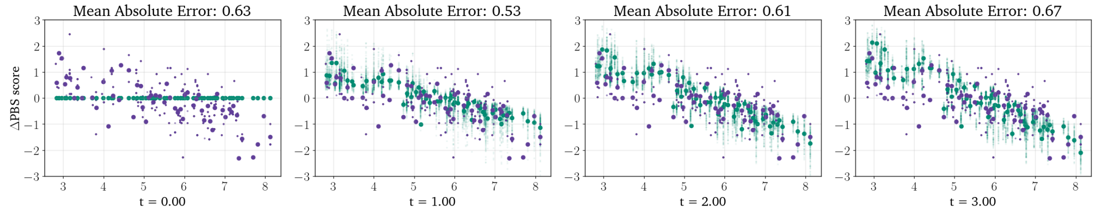

FIGURE 2: Change in PBS, relative to the original, pre deliberation, measurement. The control is omitted as there was no significant change.

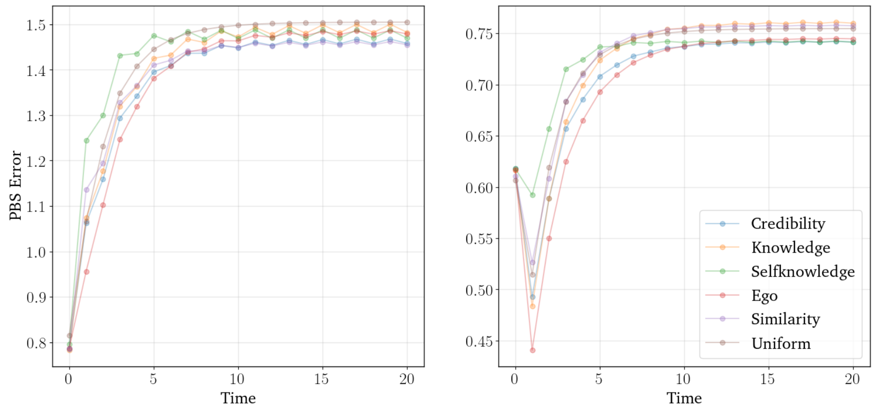

FIGURE 3: Prediction error of the model as a function of time, binned relative to the original PBS.

PBS, seemed to also be the most knowledgeable, if this was skewed towards voters with high PBS, then these voters would have more effect on people's opinions in this model when using knowledge-based trust. Looking at the binned errors in Figure 3, we see that the model performs better when we do not include knowledge, further indicating that knowledge is a poor predictor of trust, or persuasiveness. Of course this claim might be weakened by noting that knowledge in this case is measured by questions regarding the current state of the America government, such as know which party currently has a majority in the senate. Thus, this specific knowledge might be insufficient to predict someone's persuasiveness on the topic of immigration for example.

We note that this slight positive results appear only when the voters are grouped by their original PBS, thereby giving the model reasonable predictive power over a population of voters. We note that this hold even for different sizes of bins. Figure 3 shows the progression of errors over time when the error is calculated on a per-individual basis, and we find the model consistently does worse than predicting someone to not change their opinion, which is what t=0 indicates.

Figure 4 shows the relation between the bias factor and the PB score, showing that the bias does not improve the models predictive power. As one might expect a bias is

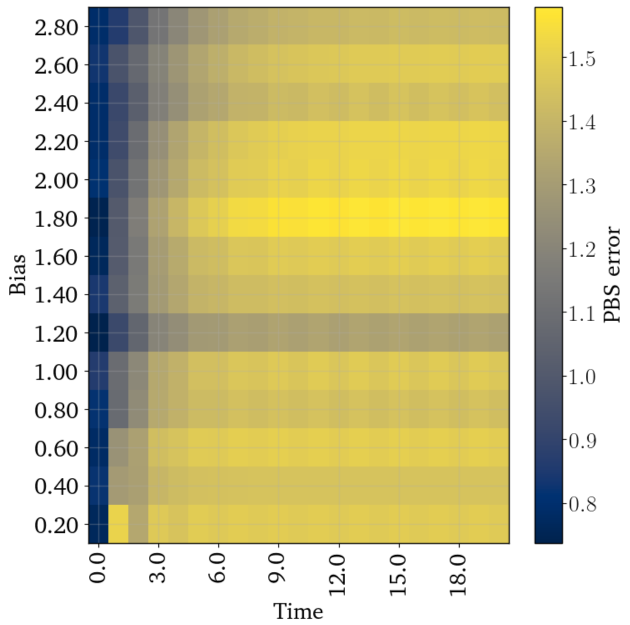

FIGURE 4: PBS Errors as a function of bias and time. bias acts as a damper, when bias is higher the model take longer to over estimate the change in opinion.

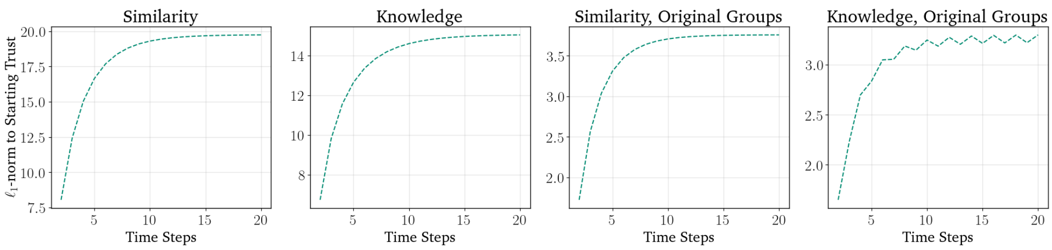

FIGURE 5: Convergence of trust matrices, as measured by the ℓ_1-norm between the trust matrix at the start and trust matrix at the current time step t

"slowing down" the model. Because of this the model is slower to diverge away from the true opinions.

0.1.2 Convergence

From **??**, we have seen that in the limit some matrices are convergent, while some are not, in particular if the matrix is aperiodic, this it is convergent. As we model the deliberation group as having fully connected matrices, the matrices are aperiodic, and thus convergent. We look at the distance between the estimated support matrix, and the true support matrix, to get a sense of the rate of convergence. The distance in the element wise ℓ_1 norm.

0.2 Elections

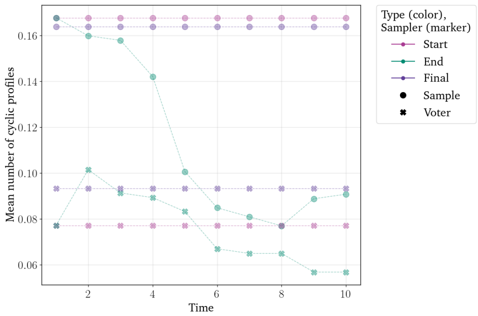

FIGURE 6: The proportion of cyclic profiles remaining, 0 indicating that no cyclic profiles were present after deliberation.

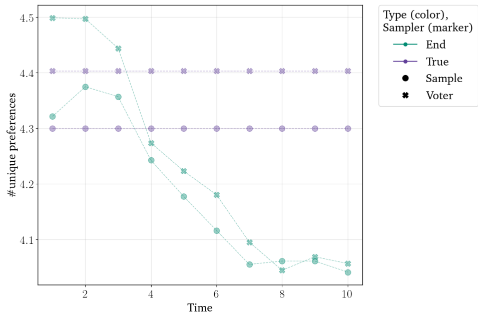

FIGURE 7: Number of unique preferences at the final step of deliberation.

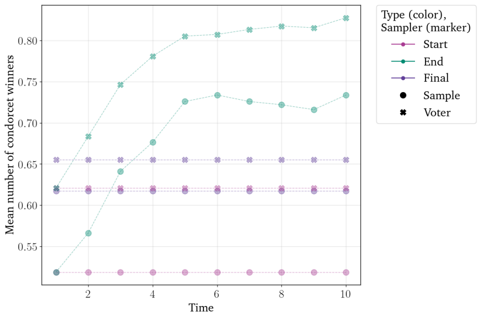

FIGURE 8: The proportion of Condorcet winners left after deliberation, value above one indicate Condorcet winners emerging during deliberation

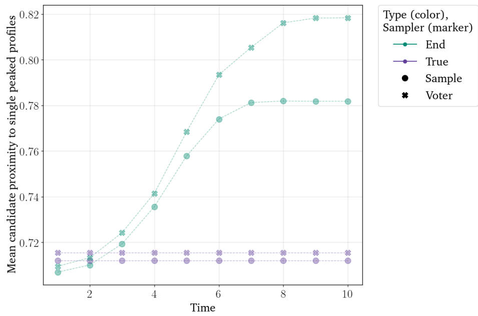

FIGURE 9: Proximity to single-peakedness after deliberation. Proximity to single-peakedness as defined in **??**.

Firstly, looking at the different voter generation mechanisms, we find that in general they do not affect the metrics much. The metric for which this is not true is that of the fraction of elections with a Condorcet winner. Though slightly unintuitive, we suspect the reason why a single voter's opinion is more likely to result in a Condorcet winner than the average of 10 voters is that the true opinions before deliberation were more polarized. As a result, having multiple "average" candidates results in little difference between them, while an individual voter is more likely to fall close to a large camp of voters, and thus become a Condorcet winner through being closer to a majority of voters.

Looking at the metrics evaluated on the model, we see similar results as with the analysis for substantive agreement. At first the simulation starts far from the metrics from the true score, then it moves towards it and overshoots it until it starts to converge. Interestingly, for these metrics, the model does require more steps to converge to the same values as the true data.

This last analysis is very short and needs work