Confirmation Bias in Opinion Dynamics of Closed Populations: A Preliminary Model

Oliver Lee

[a]*New York University, MATH-UA 251: Introduction to Math Modeling Fall 2023*

Abstract

Decision making has been a heavily researched topic in the last decade, specifically the phenomenon of confirmation bias. However, many studies focus on qualitative methods in order to understand the phenomenon in addition to neglecting social implications. In this study we aim to explore the effects of confirmation bias in decision making by employing and altering an established mathematical model for bias assimilation in opinion dynamics of a closed population. Through numerical simulations of sample data fitted from real survey data we discovered the polarizing nature of opinions and how over time populations seem to converge to opposing opinions. Limitations in the one dimensional and random nature of our simulation identify opportunities for future research that take these into account.

Keywords: Decision making, Opinion Dynamics, Confirmation Bias

Over the past decade, decision making has been a heavily researched topic in both psychology (Rollwage and Fleming, 2021; Talluri et al., 2018) and mathematics (Wierzbicki, 1982; Rainer and Krause, 2002). From game theory to mathematical programming, the applications of decision making are vast. By deconstructing mental processes and their functions, capabilities, as well as their stressors, may we better understand cognition. One specific phenomena of cognition is the idea of confirmation bias. Talluri et al. (2018) define confirmation bias as "committing to a categorical proposition, or a course of action, biases subsequent judgment and decision-making." Many scholars have attempted to explore this idea and it's implications in addition to modeling it's implications to decision making (Powell et al., 2012; Talluri et al., 2018; Rollwage and Fleming, 2021). Others have also attempted to model confirmation bias and it's effect on dependent on time (Xia et al., 2020; Lange et al., 2021). However, the literature mainly focuses on decision making while lacking exploration of social implications. A specific application of this is the field of opinion dynamics and social networks. In this exploratory paper we aim to build upon pre-existing models of opinion formation within closed populations in order to investigate how confirmation bias effects the opinions of individuals over time.

1. Theoretical Model

In order to sufficiently model we chose a model based upon bias assimilation. The model used in this paper is based upon Xia et al. (2020)'s use of Dandekar et al. (2013)'s model for "biased assimilation by a single agent in a fixed environment." First, a fixed population of n agents is defined through a directed graph $G(V, E)$ such that V is the set of agents labeled $1, 2, ...n$ and $E = \{(i, j) : i, j \in V, i \neq j\}$ is the set of edges, or in our case, neighbors. We also define a set $N_i = j : (i, j) \in E$ as the set of all neighbors of agent i. Next, a $n \times n$ matrix W contains weights w_{ij} in which self weight $w_{ii} \geq 0$. We also do not allow self-looping in G. However, w_{ii} can be positive. We set $x_i(t) \in [0, 1]$ to represent agent i's opinion at discrete time t.

$$x_i(t+1) = \frac{w_{ii}x_i(t) + x_i(t)^{b_i}s_i(t)}{w_{ii} + x_i(t)^{b_i}s_i(t) + (1 - x_i(t))^{b_i}(d_i - s_i(t))} \tag{1}$$

$x_i(t + 1)$ represents agent i's update. $b_i > 0$ represents agent i's bias. $s_i(t) = \sum_{j \in N_i} w_{ij}x_j(t)$ is defined as the "weighted average support" for the opinion denoted by $x_i(t) = 1$. $d_i = \sum_{j \in N_i} w_{ij}$ is also defined such that $d_i - s_i(t)$ represents the weighted average support for the opinion denoted by 0. Furthermore, 1 and 0 represent the polarizing opinions of a particular subject.

In order to better represent confirmation bias, we choose to alter this model by considering Rainer and

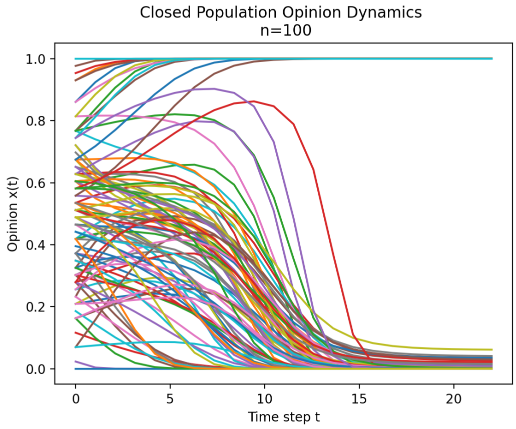

Figure 1: Dandekar et al. (2013)'s model for opinion dynamics with bias assimilation. Initial opinions $x_i(0)$ are randomly sampled from a poisson distribution. Weights $w_{ij} = 1$. Bias b_i is randomly sampled from [0.5, 1.5]

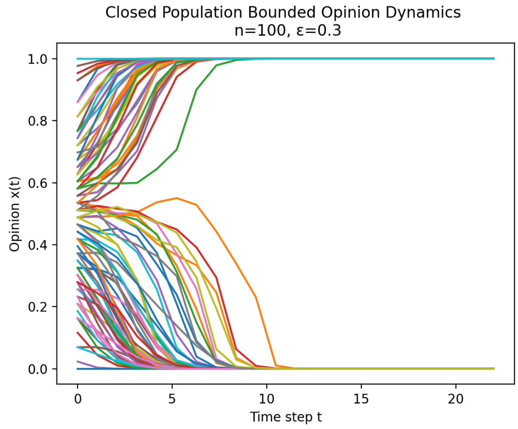

Figure 2: Dandekar et al. (2013)'s model for opinion dynamics with bias assimilation with bounded confidence. Initial opinions $x_i(0)$ are randomly sampled from a poisson distribution. Weights $w_{ij} = 1$. Bias b_i is randomly sampled from [0.5, 1.5]. Confidence interval $\epsilon = 0.3$

Krause (2002)'s model on opinion dynamics with bounded confidence. Therefore, we define:

$$I(i, x) = \{1 \leq i \leq n : |x_i - x_j| < \epsilon\} \quad \text{s.t. } I_i \subseteq N_i$$

$I(i, x)$ represents the set of neighbors of agent i where the difference between opinions x_i and x_j is less than some ϵ. This then allows us to better model confirmation bias by only allowing opinions that are relatively similar to each other to effect one another.

2. Method

In this paper, we plan to analyze both the original model and our altered model in addition to their application to real data. Specifically, we use data from In order to analyze our model, we choose varying initial conditions for $x_i(t)$, W, and b_i. For our first analysis we chose a random Poisson distribution within $[0, 1]$ in hopes of somewhat accurately representing a sample population's opinions. We also randomize agent connections. In order to avoid a near complete graph, or where each agent is neighbors with every other agent, through numerical exploration, we only randomize connections for $\frac{n}{3}$ agents. This ensures some agents are sufficiently connected while others only have a few connections. Thus we make a simplifying assumption that these connections are sufficient to represent populations.

Next, for the sake of simplification, we assume that each agent's opinion has the same effect on one another. Thus we set $w_{ij} = 1$ in order to give equal influence to each agents' opinions. Furthermore, we also randomly pick b_i from a uniform distribution. Specifically, we employ the `np.random.uniform()` python function for b_i. Next, we assume an acceptable $\epsilon = 0.3$ in which individuals can influence each other. We rationalize this by understanding that $x(t) = 0.5$ represents a neutral stance. Thus, opinions above and below 0.5 are somewhat categorical, however, varying by severity. Then, we chose an interval of 0.3 where opinions are within 0.1 factor of each category.

For our second analysis, following Xia et al. (2020) we also pulled a random sample for w_{ij} from [0.5, 1.5] to represent varying influence of opinions instead of uniform influence.

For our third analysis, we pulled survey data from NORC GSS (2020). Specifically, we used data from the question "On a scale from 0 to 10, how bad or good do you think the impacts of climate change will be for the world as a whole? 0 means extremely bad, 10 means extremely good" (NORC GSS, 2020). For this specific question, NORC GSS (2020) surveyed 2016 out of 4032 total participants. However, only 1284 provided an answer. From this interval data, we attempt to fit the data using `scipy.norm.fit()` to get a probability density function. Then, from this pdf we use the `scipy.norm.rvs()` to get an artificial sample of 1000 participants based upon NORC GSS (2020) survey data. Furthermore, the sample data was also normalized to fit our model. However, some unwanted outliers greater than 1 and less than 0 were also generated in this sample. Thus, for the sake of simplicity, we assume they are not factors of the sample. This leaves us with a net

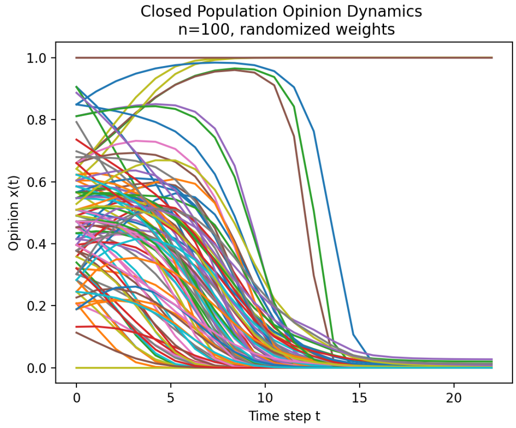

Figure 3: Dandekar et al. (2013)'s model for opinion dynamics with bias assimilation. Initial opinions $x_i(0)$ are randomly sampled from a poisson distribution. Weights w_{ij} and bias b_i is randomly sampled from [0.5, 1.5].

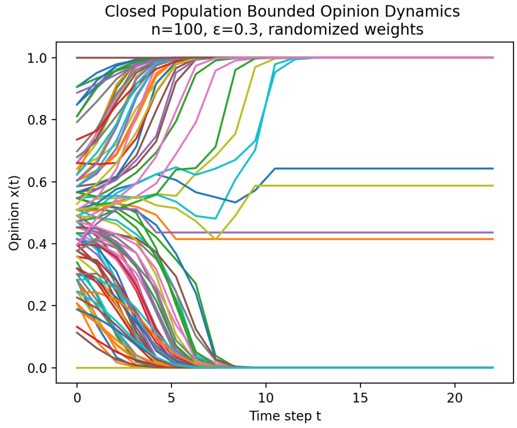

Figure 4: Dandekar et al. (2013)'s model for opinion dynamics with bias assimilation and bounded confidence. Initial opinions $x_i(0)$ are randomly sampled from a poisson distribution. Weights w_{ij} and bias b_i are randomly sampled from [0.5, 1.5]. Confidence interval $\epsilon = 0.3$

sample of 937 participants. We implemented our altered model and original model in python. Opinion data preprocessing was also done in python. A full implementation can be found in the submission for this paper.

3. Results

3.1. Model Comparison

When running our numerical implementation of the original model, observe there are two stable equilibrium at 0 and 1 in figures 1 and 3. We also take note that equilibria exist both for when weights are randomized and when they are not. Next, notice how some opinions in figure 3 are growing towards 1 but end up dipping towards 0 around $t = 10$. This further indicates the importance of weights.

In Figures 2 and 4 we notice the different pattern in opinion progression than without a bounded confidence interval. However, equilibria at 0 and 1 still exist. Opinions seem to converge to these equilibrium values dependent on their initial opinion. Furthermore, weight, or an individual's opinion influence doesn't seem to change this pattern. However, in figure 4 we acknowledge the existence of a few opinions staying stable in time. Possible explanations for this phenomenon could be that when agents have a lack of neighbors with similar opinions. In other words, numerically, d_i and $s_i(t)$ both drop to zero which leads to $x_i(t + 1) = x_i(t)$.

3.2. Integrating Real Data

When using survey data NORC GSS (2020) to initialize our model, notice how most of the opinions in figure 6 seem to rapidly decay to 0. This is notably faster than in both figures 1 and 3. However, this may be due to the impact of randomized weights. Also note when applying our confidence interval, our model exhibits the same consistent polarizing behavior shown in figure 5.

4. Discussion

Modeling confirmation bias through adding a confidence bound to an existing model proved to further polarize the data. This is consistent with previous research that discovered individuals prefer to form polarized groups of similar opinion (Cinelli et al., 2021). Figure 7 pulled from Cinelli et al. (2021) demonstrates this through different Facebook communities and their opinion severity with respect to vaccines.

We also discovered that this introduction of a confidence bound altered the evolution of opinions dependent on their initial opinion. This finding suggests implications towards a difficulty in changing opinions. Specifically, in terms of climate change, those who are slightly concerned are predicted to only intensify their concern with time. Additionally, their opinions will not tend towards the opposite. This also seems to be an additional implication of confirmation bias as well. Individuals well within a certain belief will always seemingly tend towards the polar extreme of such belief. However, those around the neutral opinion, or $x_i(t) \approx 0.5$, will tend towards opinions of those near them that hold the most weight or influence.

Furthermore, we note the importance of opinion weight and bias. When running numerous numerical

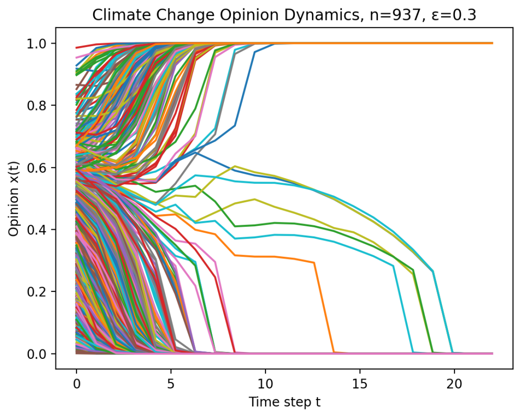

Figure 5: Initial opinions $x_i(0)$ sampled from real data distribution are used on Dandekar et al. (2013)'s model with bounded confidence. Weights w_{ij} and bias b_i are randomly sampled from [0.5, 1.5]. Confidence interval $\epsilon = 0.3$

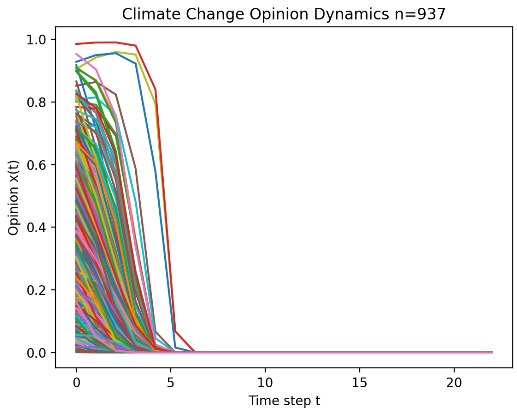

Figure 6: Initial opinions $x_i(0)$ sampled from real data distribution are used on Dandekar et al. (2013)'s model. Weights w_{ij} and bias b_i are randomly sampled from [0.5, 1.5].

simulations, we noticed that often, the original model simulation would look different each time. For instance, they would represent either a slow (figure 1) or fast (figure 6) convergence to either polarized opinion. However, in our altered model, it would remain relatively similar.

5. Conclusion & Limitations

In this paper, we explored the effects of confirmation bias on populations through fitting sample data to a modified opinion dynamics model. Through numerical simulation, we discovered how initial opinions, opinion weight, and bias all influence and ultimately lead to opposing polarized opinions within a closed population.

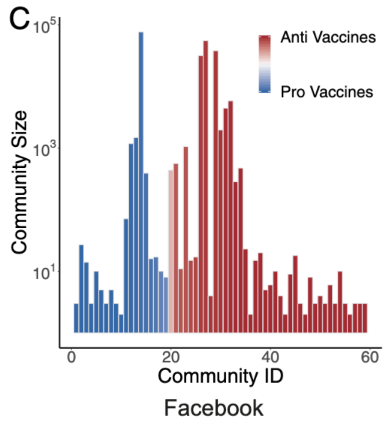

Figure 7: Cinelli et al. (2021)'s data on average opinion of different communities of Facebook

However, many limitations to this study exist. First, measuring opinions on a scale from 0 to 1 is arbitrary and one-dimensional. Intricate opinions that beg for elaboration are not considered. Thus, our model may only be valid for yes/no or agree/disagree type topics. Next, randomizing connections between neighbors, opinion weight, and bias all pose major threats to our model's validity. For instance, an individual with a very strong opinion towards the threat of climate change may naturally seek out in addition to having great influence on the opinions of others. Without considering the implications each measure has upon one another, we cannot accurately represent a human population.

Furthermore, without considering cultural and social context of our population we hope to model, we neglect external pressures towards such population it's implications that may effect initial as well as the severity of confirmation bias on opinion development.

In addition, by only considering a closed population, we neglect the possibility for additional agents and perspectives to be accounted for. Furthermore, we only consider connections that have already been established. Specifically, new connections being made within a closed population are not accounted for. All of these shortcomings limit our study's range of validity.

As a result, future research should account for multivariate measures of opinion in addition to using valid measures and thus data for opinion weight as well as bias. By also focusing on specific populations as well as accounting for additional individuals and connections may we better model and observe the effects of confirmation bias and opinion development over time.

References

Cinelli, M., Morales, G.D.F., Galeazzi, A., Quattrociocchi, W., Starnini, M., 2021. The echo chamber effect on social media. Proceedings of the National Academy of Sciences 118, e2023301118. doi:10.1073/pnas.2023301118.

Dandekar, P., Goel, A., Lee, D.T., 2013. Biased assimilation, homophily, and the dynamics of polarization. Proceedings of the National Academy of Sciences 110, 5791–5796.

Lange, R.D., Chattoraj, A., Beck, J.M., Yates, J.L., Haefner, R.M., 2021. A confirmation bias in perceptual decision-making due to hierarchical approximate inference. PLOS Computational Biology 17, 1–30. URL: `https://doi.org/10.1371/journal.pcbi.1009517`, doi:10.1371/journal.pcbi.1009517.

NORC GSS, w.f.f.N., 2020. National Opinion Research Center General Social Survey 1972-2021, 2020 [Dataset]. Roper #31118778, Version 2. NORC at the University of Chicago [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research [distributor]. Access Date: Dec-21-2023.

Powell, M.B., Hughes-Scholes, C.H., Sharman, S.J., 2012. Skill in interviewing reduces confirmation bias. Journal of Investigative Psychology and Offender Profiling 9, 126–134. doi:https://doi.org/10.1002/jip.1357.

Rainer, H., Krause, U., 2002. Opinion dynamics and bounded confidence: Models, analysis and simulation. Journal of Artificial Societies and Social Simulation 5.

Rollwage, M., Fleming, S.M., 2021. Confirmation bias is adaptive when coupled with efficient metacognition. Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20200131. URL: `https://doi.org/10.1098/rstb.2020.0131`, doi:10.1098/rstb.2020.0131.

Talluri, B.C., Urai, A.E., Tsetsos, K., Usher, M., Donner, T.H., 2018. Confirmation bias through selective overweighting of choice-consistent evidence. Current Biology 28, 3128–3135. e8. Pmid:30220502.

Wierzbicki, A.P., 1982. A mathematical basis for satisficing decision making. Mathematical Modelling 3, 391–405. doi:10.1016/0270-0255(82)90038-0. iD: 272294.

Xia, W., Ye, M., Liu, J., Cao, M., Sun, X.M., 2020. Analysis of a nonlinear opinion dynamics model with biased assimilation. Automatica 120, 109113. doi:10.1016/j.automatica.2020.109113.