2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) | 979-8-3503-2745-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACIIW59127.2023.10388085

Understanding the Rigidity of Beliefs in Temporal Social Networks

Adiba Mahbub Proma
Department of Computer Science
University of Rochester
Rochester, USA
aproma@cs.rochester.edu

***Abstract*—Current research in psychological, cognitive, and social sciences suggests that belief rigidity is a significant cause of opinion polarization. People cling to their beliefs despite contradicting evidence, making them vulnerable to misinformation. This is exacerbated by the algorithmic structure of current social media, where it is easy for people to create their own echo chambers and only interact with like-minded people. For my PhD dissertation, I propose to explore how various attributes and actions of social networks affect the rigidity of one's beliefs in temporal social networks. I plan to study these effects within a virtual lab setting, specifically focusing on climate change denial and the politics around climate change. Through our experiments, I investigate whether it is possible to replicate social media signals in a laboratory setting, and thus introduce a new way of empirically studying social signals without having to use social media data. Finally, I plan to use the resulting insights to design intervention strategies for combating extreme polarization of opinions in social media, and thus aid in reducing misinformation.**

***Index Terms*—belief rigidity, social networks, nudging**

I. Motivation

Research in psychological, cognitive, and social sciences suggests that belief rigidity is a significant cause of opinion polarization. This rigidity means that people stubbornly cling to their beliefs even when faced with evidence that contradicts them [1]. At an individual level, people tend to reject counterevidence to avoid cognitive dissonance. At the social level, mutual reinforcement among like-minded peers can result in increasingly stubborn beliefs over time. This can lead to increased susceptibility to misinformation, where misinformation is defined as information that is false, misleading or inaccurate [2].

Moreover, the algorithmic design of social media platforms makes it easier for misinformation to penetrate the public discourse since these platforms promote echo chambers through their homophily-based recommendation systems. Social media users tend to create their own echo chambers, or "filter bubbles", by selectively exposing themselves to news and information that aligns with their pre-existing beliefs and values [3]. Research suggests that encouraging more diverse and balanced media consumption habits, and improving the design of social media platforms to better facilitate exposure to different perspectives can help reduce this phenomenon [3].

The work was supported by grants from the US Army Research Office (W911NF-22-1-0182)

Intuitively, the peer and content recommendation engines of social media may have an important—yet underexplored—role to play in mitigating polarization. Moreover, the literature largely overlooks how people's belief dynamics are affected by the interplay between node attributes and evolving network structures in social media. Therefore, it is crucial to investigate the underlying factors that contribute to increased belief rigidity. This will help us understand what changes need to be implemented in the algorithmic design of social media platforms to combat the issue of polarization.

In this thesis, I propose to explore how various attributes and actions of social networks affect the rigidity of one's beliefs in temporal social networks. I plan to study these effects within a virtual lab setting, specifically focusing on the climate change denial and the politics around climate change. The resulting insights can be used to design intervention strategies for combating extreme polarization of opinions in social media, and thus aid in reducing misinformation.

II. Scope of the Research

My research objectives are given below.

- **Introducing a platform that we can use to study various attributes of social networks in a lab setting:** Researchers often face challenges in accessing social media data due to data privacy concerns, and even when they do, they may not have access to the relevant data at the right time. Additionally, testing intervention strategies on social media is challenging. To address these issues, the aim is to investigate whether it is possible to replicate social media signals in a laboratory setting, allowing for the study of specific social cues and the development of effective intervention strategies.
- **Designing and conducting experiments to study belief rigidity on social media:** I will design a series of experiments to study various social signals including the effect of tweaking network structure within the lab setting.
- **Designing effective interventions to reduce belief rigidity and analyse its role in the spread of misinformation:** Our understanding of the networks would allow us to design better interventions and aid in tackling the spread of misinformation.

Keeping the research objectives in mind, we formulate the following research questions:

- RQ1: How does exposure to diverse content affect belief rigidity in social networks?
- RQ2: Do demographic identity factors contribute to belief rigidity for social networks?
- RQ3: What effect does the structure of social networks have on belief rigidity?
- RQ4: How can we design interventions to reduce belief rigidity, and how effective are they in tackling misinformation?

III. Background and Related Works

Misinformation is defined as information that is misleading, irrespective of its intent to deceive [2]. Strategies to combat misinformation either tackle algorithmic design or how humans perceive information [2]. For social networks, tackling algorithmic design includes early detection of malicious accounts and using ranking algorithms, such as text-based fake news detection pipelines and Graph Neural Networks (GNNs) [4] to model news diffusion patterns. There has also been research on fact-checking algorithms such as Twittertrails, FactWatcher and TruthTeller. Twittertrails uses deep learning to model the spread and trustworthiness of news [5]. FactWatcher complements previous approaches by considering different types of facts, and also provides fact ranking, fact-to-statement translation and keyword-based fact search [6]. TruthTeller transcribes political videos and checks them against a database on PolitiFact and FactCheck.org [7]. Recently in July 2022, Meta announced developing Sphere, which is an AI-driven tool for fake news detection by fact-checking against millions of web articles. Once they spot fake or questionable information, the AI would flag the post, warn the user and might also suggest more reliable sources [8].

While the computer science literature is rich in fact-checking and reliability labelling methods, tackling the spread of misinformation goes beyond this [9]. For tackling the human factors, strategies include inoculation and nudging [10], [11]. Inoculations provide people with information to counteract misinformation, but it is difficult to apply them to every case as they need to happen before the misinformation reaches the person. [10]. Similarly, subtle changes in the "choice architecture" such as facilitation, framing and personalization have been known to influence behaviors in predictable ways [12]. Social influence, especially social comparison, has also proven to be an effective nudging technique since people are open to overriding their own beliefs to replicate others' actions due to the herd instinct bias [13]. Nudging has also been used to encourage critical thinking and fact-checking behavior among digital media users [11].

There have been multiple modeling efforts from the social and psychological domains to capture the social dynamics of belief rigidity. For example, the Hierarchical Ising Opinion Model captures both the individual and social processes in a unified complex adaptive system (i.e., in a 'network of networks') [14]. Moreover, there has also been some work on understanding how people react to fake news [15]. For example, Gabriel et al. developed a framework that includes four "reaction frames" to map how readers might react to a given headline: suspicion, confirmation, disbelief, and misunderstanding. The authors also developed a machine learning model that can predict the reaction frame that a given reader is likely to have based on their demographic information and past reading history [15].

IV. Prior Work

A. Nudging through user interface attributes in e-commerce platforms

Our earlier work studies the effect of *nudging* by facilitation through user interface attributes. We select the domain of sustainable consumption in e-commerce platforms as a use case. The prevalence of the "attitude-behavior gap", which is when consumers fail to be eco-friendly despite wanting to opt for greener consumption [16]–[18], makes this domain appropriate to study the effects of nudging. We develop a prototype named SEER that targets addressing three major factors responsible for the attitude-behavior gap - inconvenience, lack of knowledge, and lack of trust [19]–[21]. The key component of SEER is an environmental rating scale (1-5 scale) to rate products on its environmental impact. To increase trust and knowledge - we add statement summaries that explain the rating and also highlight environment-related keywords.

We then conduct a quasi-randomized case-control study with 98 participants. Participants are given a prompt to help a local school with limited budget by purchasing products. To elicit real-life behavior, they receive 10% of the money they save for the school. All participants watch a 3 minute video on the effects of climate change and how individual actions can help. Therefore, the participants face conflicting objectives of saving money for an immediate reward or investing in eco-friendly products for long-term environmental benefits. Only the case group has access to the additional features. The one-tailed t-test reveals that the participants from the case group are significantly more eco-friendly than the control group ($p < 0.005$), as they select a higher number of eco-friendly products.

Our experiment shows that nudging through even simple interface changes can encourage specific behavior in people, but only to a certain extent since other factors also have an impact. For example, price significantly affects consumer behavior [22]–[24], and we found a significant negative correlation between the extra price a consumer has to pay for an eco-friendly product and the number of consumers still willing to buy it (Pearson's correlation co-efficient $r = -0.73$ and $p = 0.007$). While this experiment especially focused on nudging through interface re-design, our current proposed work focuses on social signals that often have a similar impact as nudging.

V. Methods

To study the effect of various social signals on belief rigidity, the experiment is designed to be an incremental case-control study, where the control is designed to imitate the signals found in current social media. Each of the cases would be designed to study a specific social signal. The study will have multiple rounds (1 round per day) and each round consists of three stages - response, revision, and rewiring. The subjects will carry out the required tasks each day through the custom-built website. Our experiment is designed to validate control and study the effects of exposure to more diverse opinions on belief rigidity.

A. Current Work

The participants are split in a quasi-randomized process into three groups - control, case 1 and case 2 - such that the groups have similar demographic characteristics, sentiment, and knowledge about climate change. Each network has 12 participants, and the experiment is repeated thrice for each group. The pilot study has 5 rounds (spanning over 5 days).

At the start of each round (i.e., the response stage), all participants will be shown prompts based around the politics of climate change, and these statements will be the same for all groups. The prompts will be generated using existing fact-checking datasets (e.g., ClimateFever [25]), actual posts on social media (eg. the Climate Change Twitter Dataset [26]) and climate-change-focused knowledge graphs (e.g., KnowUREnvironment [27]). Participants are required to rate the statement on a 7 point Likert scale on how strongly they agree or disagree with the statement, along with their reason for the rating. Participants are assigned 3 base connections. In the next stage (i.e., response stage), they can see the responses (rating and reason) of these connections, and have the option to update their answer if they want to. Finally, in the rewiring stage, they can see responses of more participants that are ”recommended” to them. The participants must like or dislike the responses they see. This will ensure that they are reading all the responses, and we can also analyse their responses later. Moreover, they select 3 people they want as connections whose answers they will see for the next round. They can stick to the connections assigned or pick new ones (or do a mixture of both). The whole process repeats for the following rounds.

For the control, the connections assigned and recommendations are such that participants only see responses of other participants with similar answers. For case 1, participants will be “recommended” a mixture of accounts with similar and opposing views during the rewiring stage, and they get to see the responses of whoever they follow in the next round’s revision stage. For case 2, we intentionally inject opposing views during the revision stage so that participants have access to opposing views before they update their answers.

Some of the hypotheses are:

- Control imitates current social media signals such as homophily and polarization
- Belief rigidity should be same or increase in control
- Polarization should occur naturally after a few iterations in control
- For case 1, rigidity should slightly decrease as they are exposed to opposing views
- Polarization should still occur naturally in case 1
- For case 2, the decrease in rigidity should be greater than the other two groups

For the experiment, temporal belief rigidity is measured by the difference between the initial Likert scale value and the updated Likert scale value in each round. I will then conduct a temporal analysis to see whether, over time, participants are less rigid in their beliefs. Across case and control group, the correlation of the independent variables such as race, gender, political affiliation, and location with the target variable (temporal difference in belief rigidity) will be analyzed. I will also invite participants to provide feedback for an in-depth qualitative analysis. The analysis will help us gain useful insight into what network structures contribute to belief rigidity.

VI. Tentative Plan for Future Work

A. Studying Framing and Node Injection

Currently, the statements are worded to sound neutral, but there is scope to experiment with complex statements that are closer to the real-world social media discourse. Moreover, bots designed to be highly provocative or highly logical in their responses can be injected into the networks. This would help quantify the effect of framing and node injection in social networks. Moreover, I plan to incorporate the option of interacting with other participants’ answers. Users would be able to provide responses to other participants’ responses, which would then be visible to those participants.

B. Quantifying Nudges on Social Media

I would also like to experiment with social factors such as trust, reputation and demographic cues. To do this, I plan to add and experiment with various UI components meant to increase trust on the user (such as the blue tick mark on Twitter), increase reputation (such as showing number of upvotes and downvotes on Reddit) or provide demographic cues about the user. These components can act as “nudges”, and the experiments can help us quantify the effect of such signals on belief rigidity.

C. Adding Multimodal Content

All statements and responses in the pilot experiment are text-based, and we can perform semantic analysis. In the future, I plan to integrate multimodal content to our experiments, including using multimodal content as statements, or allowing users the option to add image data and videos to the responses. This integration would bring more technical challenges such as using appropriate data integration techniques, building ML classification models that can handle both text-based and vision data, and so on.

D. Generalizing Concepts

Right now, I am working with climate change statements. However, the same set of experiments can be repeated with various topics such as health and politics. This would help us

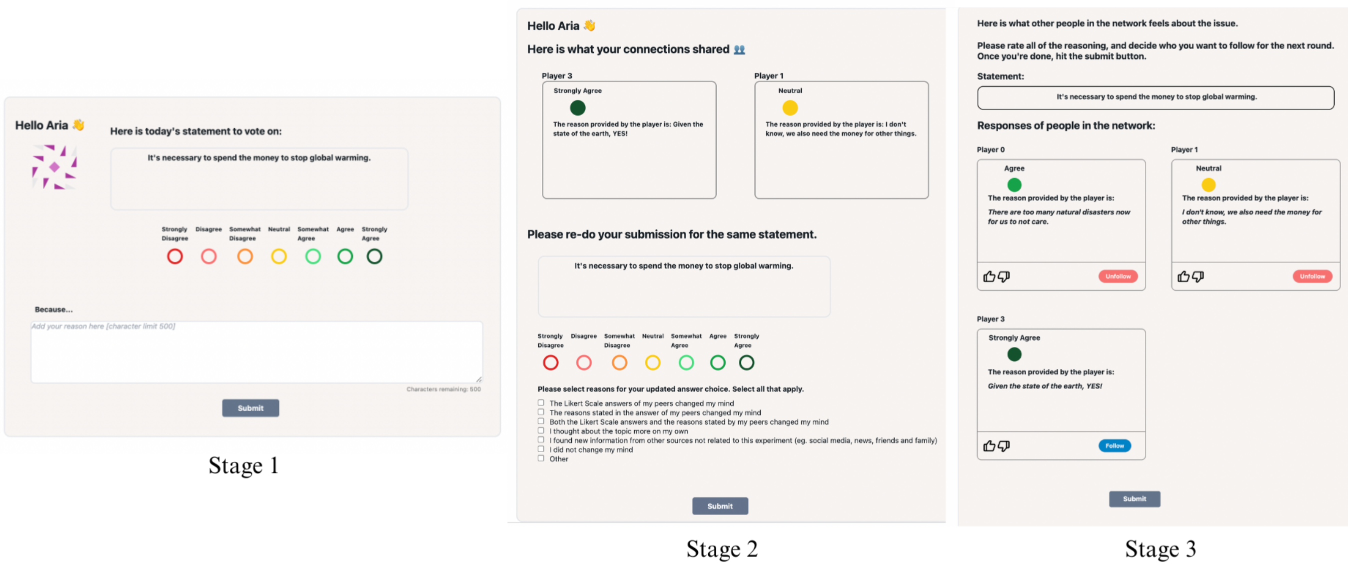

Fig. 1. The interface for the three stages of our current experimental design. In stage 1, users can rate and provide reason. In stage 2, they can see responses of their connections, and in stage 3, they can see responses of other people in the network, and have the option to follow them for the next round.

generalize our concepts to tackle other contextual domains of misinformation.

Moreover, it would also be interesting to see how gender, age, race and culture affect belief rigidity. Currently, the target is the US population, but it is possible to conduct the experiment on a more global scale. Our designed framework enables us to easily extend the scope of our work to multinational, multilingual and multicultural levels, thus making it possible to identify which characteristics are population-specific and which are generalizable. Extending to such scales would also bring additional technical challenges such as tackling code-switching among languages and taking into account cultural differences in interaction.

E. Designing Interventions

In the future, I would like to design effective intervention strategies to reduce belief rigidity in social media. The goal of the intervention would be to reduce the effect of polarization, homophily and confirmation bias to mitigate the propagation of misinformation within these networks. The specific features of the intervention strategy would strongly depend on prior findings from our experiments. Intuitively, it should be a combination of introducing tweaks in the social network structure, and designing better user-adaptive interfaces.

VII. Expected Contributions to Affective Computing Field

My dissertation takes a social network-based approach to empirically understand belief rigidity and its role in tackling misinformation. As beliefs can influence people's emotions towards a certain topic, the impact of belief on people's emotions, cognition and thought processes are often permanent. Therefore, my work aims to extend the sub-field of "Issues in Psychology and Cognition of Affect in Affective Computing Systems", and would allow us to empirically quantify people's reaction to different social media signals. For example, we aim to show that it is possible to computationally capture temporal opinion polarization.

Using these findings, my dissertation also aims to design an effective intervention system to reduce belief rigidity. While it is difficult to know at this point what that intervention would look like, we can intuitively say that it would be a combination of designing more user-adaptive interfaces and inherent changes in the structure of social networks. Therefore, our work also contributes to the sub-field of "Human-Centred Human-Behaviour-Adaptive Interfaces". Thus the proposed thesis promises to extend the traditional techniques of affective computing and inspire new directions for the community.

Ethical Impact Statement

The designed experiments have been approved by the IRB of our university. Research on reducing belief rigidity can help design interventions to aid individuals become more open-minded and flexible in their thinking, improving decision-making and reducing bias. It can also play a crucial role in reducing online hate and cyber-bullying, making the social media space less toxic.

However, there are also potential negative ethical implications that need to be considered. For example, people might misuse this research to reinforce harmful beliefs by framing them as merely "rigid" rather than fundamentally flawed or unjust. To mitigate these issues, education in critical thinking is crucial, and care must be taken during such research to ensure that appropriate safeguarding mechanisms are in place.

References

[1] M. Scheffer, D. Borsboom, S. Nieuwenhuis, and F. Westley, "Belief Traps: Tackling the Inertia of Harmful Beliefs," *Proceedings of the National Academy of Sciences*, vol. 119, no. 32, p. e2203149119, 2022.

[2] K. M. d. Treen, H. T. Williams, and S. J. O'Neill, "Online Misinformation about Climate Change," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 11, no. 5, p. e665, 2020.

[3] D. Spohr, "Fake News and Ideological Polarization: Filter Bubbles and Selective Exposure on Social Media," *Business Information Review*, vol. 34, no. 3, pp. 150–160, 2017.

[4] G. Barnabò, F. Siciliano, C. Castillo, S. Leonardi, P. Nakov, G. Da San Martino, and F. Silvestri, "Fbmultilingmisinfo: Challenging Large-Scale Multilingual Benchmark for Misinformation Detection," in *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2022.

[5] P. T. Metaxas, S. Finn, and E. Mustafaraj, "Using Twittertrails.com to Investigate Rumor Propagation," in *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, pp. 69–72, 2015.

[6] N. Hassan, A. Sultana, Y. Wu, G. Zhang, C. Li, J. Yang, and C. Yu, "Data In, Fact Out: Automated Monitoring of Facts by Factwatcher," *Proceedings of the VLDB Endowment*, vol. 7, no. 13, pp. 1557–1560, 2014.

[7] A. Lotan, A. Stern, and I. Dagan, "Truthteller: Annotating Predicate Truth," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 752–757, 2013.

[8] "Introducing Sphere: Meta AI's Web-scale Corpus for Better Knowledge-Intensive NLP." Accessed: 11-30-2022.

[9] T. Schuster, R. Schuster, D. J. Shah, and R. Barzilay, "The Limitations of Stylometry for Detecting Machine-generated Fake News," *Computational Linguistics*, vol. 46, no. 2, pp. 499–510, 2020.

[10] J. Cook, P. Ellerton, and D. Kinkead, "Deconstructing Climate Misinformation to Identify Reasoning Errors," *Environmental Research Letters*, vol. 13, no. 2, p. 024018, 2018.

[11] L. Konstantinou, A. Caraban, and E. Karapanos, "Combating Misinformation through Nudging," in *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part IV 17*, pp. 630–634, Springer, 2019.

[12] A. Caraban, E. Karapanos, D. Gonçalves, and P. Campos, "23 Ways to Nudge: A Review of Technology-mediated Nudging in Human-computer Interaction," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–15, 2019.

[13] R. B. Cialdini and N. Garde, "Influence (vol. 3)," *Port Harcourt: A. Michel*, 1987.

[14] H. L. van der Maas, J. Dalege, and L. Waldorp, "The Polarization within and across Individuals: The Hierarchical Ising Opinion Model," *Journal of Complex Networks*, vol. 8, no. 2, p. cnaa010, 2020.

[15] S. Gabriel, S. Hallinan, M. Sap, P. Nguyen, F. Roesner, E. Choi, and Y. Choi, "Misinfo Reaction Frames: Reasoning about Readers' Reactions to News Headlines," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, (Dublin, Ireland), pp. 3108–3127, Association for Computational Linguistics, May 2022.

[16] W. Young, K. Hwang, S. McDonald, and C. J. Oates, "Sustainable Consumption: Green Consumer Behaviour when Purchasing Products," *Sustainable Development*, vol. 18, no. 1, pp. 20–31, 2010.

[17] I. Vermeir and W. Verbeke, "Sustainable Food Consumption: Exploring the Consumer "Attitude–Behavioral Intention" Gap," *Journal of Agricultural and Environmental Ethics*, vol. 19, no. 2, pp. 169–194, 2006.

[18] P. Wheale and D. Hinton, "Ethical Consumers in Search of Markets," *Business Strategy and the Environment*, vol. 16, no. 4, pp. 302–315, 2007.

[19] M. R. Gleim, J. S. Smith, D. Andrews, and J. J. Cronin Jr, "Against the Green: A Multi-method Examination of the Barriers to Green Consumption," *Journal of Retailing*, vol. 89, no. 1, pp. 44–61, 2013.

[20] Y.-S. Chen and C.-H. Chang, "Enhance Green Purchase Intentions: The Roles of Green Perceived Value, Green Perceived Risk, and Green Trust," *Management Decision*, 2012.

[21] Y. Joshi and Z. Rahman, "Factors Affecting Green Purchase Behaviour and Future Research Directions," *International Strategic Management Review*, vol. 3, no. 1-2, pp. 128–143, 2015.

[22] A. S. Ehrenberg, K. Hammond, and G. Goodhart, "The After-effects of Price-related Consumer Promotions," *Journal of Advertising Research*, vol. 34, no. 4, pp. 11–22, 1994.

[23] D. R. Lichtenstein, N. M. Ridgway, and R. G. Netemeyer, "Price Perceptions and Consumer Shopping Behavior: a Field Study," *Journal of Marketing Research*, vol. 30, no. 2, pp. 234–245, 1993.

[24] N. Ramya and S. M. Ali, "Factors Affecting Consumer Buying Behavior," *International Journal of Applied Research*, vol. 2, no. 10, pp. 76–80, 2016.

[25] T. Diggelmann, J. Boyd-Graber, J. Bulian, M. Ciaramita, and M. Leippold, "Climate-Fever: A Dataset for Verification of Real-world Climate Claims," *arXiv preprint arXiv:2012.00614*, 2020. (Accepted for the Tackling Climate Change with Machine Learning Workshop at NeurIPS 2020).

[26] D. Effrosynidis, A. I. Karasakalidis, G. Sylaios, and A. Arampatzis, "The Climate Change Twitter Dataset," *Expert Systems with Applications*, vol. 204, p. 117541, 2022.

[27] M. S. Islam, A. Proma, Y. Zhou, S. N. Akter, C. Wohn, and E. Hoque, "KnowUREnvironment: An Automated Knowledge Graph for Climate Change and Environmental Issues," in *AAAI 2022 Fall Symposium: the Role of AI in Responding to Climate Challenges*, 2022.