Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

THE EFFECT OF INFLUENCERS ON SOCIETAL POLARIZATION

John M. Betts

Department of Data Science and Artificial Intelligence
Faculty of Information Technology
Monash University
Wellington Road
Clayton, Victoria 3800, AUSTRALIA

Ana-Maria Bliuc

Department of Psychology
School of Social Sciences
The University of Dundee
Nethergate
Dundee, DD1 4HN, SCOTLAND

ABSTRACT

Societies can polarize when there is disagreement on important issues. The rise of social media in recent years has led to the phenomenon of influencers, who are now prominent in public debate, especially online. However, their effect on polarization is not well understood. To address this gap in knowledge, we create an Agent-Based Model of a society into which an influencer is introduced and their effect on polarization observed. Results show that an influencer holding extreme views will always increase the rate of polarization, with this effect increasing in line with their reach and activity level. The effect of a neutral influencer varies with the tolerance for opposing beliefs in the society: slowing the rate of polarization for relatively tolerant societies, but increasing the rate when societies are more conservative, or the influencer has narrow reach. Consequently, these results have implications for the design of influencer campaigns for social good.

1 INTRODUCTION

Despite the increased availability and use of modern communication technologies, contemporary society seems more polarized on key social issues than ever before, McCoy et al. (2018); McCoy and Somer (2019). The rise of social media as an important channel for entertainment, shopping, news and communications over recent years has led to the phenomenon of the influencer. These are groups or individuals with a strong persuasive power due to expertise, personality, fame or notoriety. Charismatic opinion leaders have the ability to influence the behavior or opinions of others, Tur et al. (2021). Influencers are now a common feature of social media platforms such as Facebook, Instagram and Twitter, as well as mainstream media, where they promote products, causes and opinions. Although the image of an influencer promoting an extreme position on an issue would be most familiar in the socio-political sphere, the promotion of fact-based information is often part of public discourse on important issues. In this case the influencers tend to represent a government or public institution rather than providing an individual point of view, and the messaging tends to be framed in a more neutral tone. Examples of this include the Centre for Disease Control and Prevention, CDC, on public health (including vaccination), or the US Environmental Protection Agency, EPA, on climate change. Despite their ubiquity, the role of influencers in affecting social attitudes and political opinions is not well understood, Riedl et al. (2021). Although it is likely that exposure to extremist messaging increases polarization, very little is known about the complex relationship between influencers, the community (to be influenced), the broader societal context (the type of society) in which the interactions between influencers and their followers occur or type of message (neutral or extremist) being transmitted.

To examine the way influencers affect societal polarization, we create an Agent-Based Model of a society, into which we introduce an influencer. Varying the confidence threshold at which an agent would

have a positive interaction with another agent (that is, one in which each agents belief moves slightly closer to the others, and a social connection is strengthened) enables relatively closed, moderately open and very tolerant societies to be simulated. Similarly, the confidence threshold for influencer-agent interactions is varied to simulate an influencer with narrow, moderate and wide reach. By increasing the probability that an influencer will interact with an agent, it is possible to amplify the influencer's effect. Since the message to be broadcast doubtlessly affects the interaction, we consider two types of influencers: one having a neutral message (that would be intended to promote consensus in the real world), and the other taking an extremist position on an issue. Under these varied parameter settings it is then possible to observe the effect of the influencer on the rate of polarization, interpreted as both the ideological extremitization of belief (as each agents belief becomes increasingly polarized), and psychological distancing (through the formation of polarized clusters).

Results show that an extremist influencer will always increase the rate of polarization in a society by increasing the rate of both ideological separation and psychological distancing. Neutral influencers having a narrow reach will also increase the rate of polarization due to the repulsive effect of their interactions. However, when neutral influencers have a wider reach, they slow the rate of ideological extremitization *and* ideological separation in more open communities. These experimental results offer insights into the design of communication strategies to slow the rate of societal polarization on key issues.

In the following section we present a brief review of research into the social processes leading to polarization, the use of Agent-Based Modeling as a means to understand the social networks that result from human interaction, and the psycho-social models of influence and bonding that underpin these processes. We then introduce the Agent-Based Model of a society that includes influencers. Section 4 presents the design of experiments and results. We conclude the paper with a summary of results and discussion of the implications of our findings.

2 LITERATURE REVIEW

2.1 Societal Polarization

In political psychology, polarization tends to be regarded as either issue-driven by support for opposing policy positions, Duffy et al. (2019), or manifest as distrust and intense dislike of the outgroup (the opposing ideological camp), Abramowitz and McCoy (2019); Druckman and Levendusky (2019). Recent studies propose that polarization results from both the ingroup (one's own group) position becoming more uniform and extreme as members of the ingroup interact, amplifying the consensus (ideological extremitization), and increased distancing from the outgroup as interactions between the ingroup and outgroup lead to dissent and hostility (psychological distancing), Bliuc et al. (2021). Thus, polarization results from these two processes operating concurrently: ideological extremitization of belief and psychological distancing as the difference between ingroup and outgroup belief increases. Classic studies in social psychology support this view by showing that communication in closed groups (or echo-chambers in the modern context), can lead to group members positions becoming more extreme, Myers (1975); Myers and Lamm (1976); Cinelli et al. (2021). These findings help explain why in groups that polarize because of opposing views on policy issues, for example (climate change supporters vs climate deniers), interaction within the group can lead to more extreme group positions and therefore a greater ideological distance between ingroup and outgroup.

Previous research on polarization has shown that selective exposure to communication from ingroup on social media is likely to produce greater polarization (by increasing confirmation bias) Ramírez-Dueñas and Vinuesa-Tejero (2021); Modgil et al. (2021). Moreover, we know that influencers on social media platforms such as Twitter help spread extreme content faster Asatani et al. (2021). However, less is known about the effect of different types of influencers on polarization, and in particular about the potential 'buffering' effects of neutral influencers compared to their extremist counterparts in various types of societies. Although empirical studies have analyzed the diffusion of ideas through social networks, for example, Bakshy et al.

(2011), and the structural position of influencers within polarized clusters, for example, Soares et al. (2018), the role of influencers in the polarization of society at large remains unclear.

2.2 Agent-Based Modeling as a Means to Understand Society

While models of social influence or opinion dynamics can be used to observe that polarization of belief is an outcome of the interaction between individuals, these models do not clearly describe the effect of this on social structure (see for example, Flache et al. (2017), and Ishii and Kawahata (2018)). Agent-Based Modeling, Macal (2016), is one means by which the social networks formed by the repeated interaction of individuals can be observed. These models enable the interaction of autonomous entities (agents), each having certain characteristics, to be simulated according to specific rules. By simulating successive interactions between agents over the longer term, the complex net effect of these individual interactions on a population at large can be observed. This makes them an effective tool to connect the micro-level assumptions about individual agent behaviour to macro-level patterns in artificial societies, Smith and Conrey (2007). They have been used, for example, to study opinion dynamics and social processes that may lead to the clustering of extreme beliefs and polarization in societies, for example, Nowak et al. (1990); Deffuant et al. (2000); Flache and Macy (2011); Turner and Smaldino (2018); Rychwalska and Roszczyńska-Kurasińska (2018). We now extend this research to study the effect of influencers on these processes.

2.3 Psycho-Social Models of Influence and Bonding

Social influence can occur when there are interactions between members of different social groups or members of the same social group holding differing opinions, and these interactions result in attitude or opinion change. The rules governing the interaction between agents in the Agent-Based Model that follows are derived from three types of social influence models:

- **Assimilative**, Friedkin and Johnsen (2011); Groeber et al. (2014), based on classic social psychological theory that individuals who are connected by a shared social identity always influence each other towards reducing the differentiation between them (through a process of averaging);
- **Similarity biased**, Sherif and Hovland (1961); Festinger (1962); Axelrod (1997), whereby only individuals who are sufficiently similar can influence each other; and
- **Repulsive**, Festinger (1962); Flache and Macy (2011); Dykstra et al. (2015), where it is assumed individuals who are highly different still influence each other, but towards increasing those differences. Thus, clusters may form as maximally oppositional views through a process of bi-polarization Flache et al. (2017).

Pairs of agents form a social bond with each other when they hold a similar level of belief, McPherson et al. (2001). These bonds strengthen after repeated interactions between pairs holding a similar level of belief, but also weaken over time if the agents either fail to interact, or interact but now hold sufficiently different levels of belief Skyrms and Pemantle (2009). Following these processes, a social network of agents is formed, which evolves over time reflecting the change in belief held by agents, and the history of their interactions.

Influencers are treated as agents having special properties in this modeling framework. They broadcast a belief to others in the society through their interaction. However, their belief on an issue does not change as a result of interactions with others, and they do not form social bonds with other agents. In this regard, we treat them as zealots or ideologues holding either a neutral or extreme stance on an issue, interacting with others while not being a member of the groups with whom they interact. Of interest to our study is the rate at which polarization occurs in the agent population, reflected in its ultimate form as the formation of distinct groups and extremitization of belief, and the effect of influencers on this process.

3 AN AGENT-BASED MODEL OF A SOCIETY WITH INFLUENCERS

3.1 Modeling Assumptions and Notation

In the model that follows, agents are represented by X_i, where $i \in \{1,...,n\}$. Each agent has a level of belief on a single issue (for example, vaccination or climate change etc., which may be positive or negative), $B(X_i)$, where $\{-1 \leq B(X_i) \leq 1\}$. The single influencer is indicated by Y. The influencer also holds (broadcasts) a belief on the same issue, $B(Y)$, where $\{-1 \leq B(Y) \leq 1\}$. A key difference between the agent and the influencer is that the belief of each agent varies over successive iterations as the result of interactions between agents, and between agents and the influencer. The influencers level of belief, however, remains constant throughout each trial. For our experiments we tested two scenarios, the case of a neutral influencer, where $B(Y) = 0$ and an extremist influencer, where $B(Y) = 1$. The strength of the social connection between each pair of agents, X_i and X_j, is is indicated by $N(X_i, X_j) \in \{0, 1, ..., \infty\}$. This is used to construct the social network of agents, and varies over sucessive iterations as a result of interactions between agents.

The confidence threshold between agents, Flache et al. (2017), CT_X, is the maximum difference in belief between a pair of agents, in which each has confidence in the other as a source of information, and influences the other positively. When two agents are within this range during an interaction, their beliefs converge and the social connection between the pair is increased. Conversely, outside this range, agents beliefs diverge and the network connection decreases. In the experiments that follow we assume that all agents have the same confidence threshold for an agent-agent interaction, which remains constant throughout each simulation trial. The confidence threshold for an interaction between an agent and the influencer, CT_Y, may be different to that for agent-agent interactions, reflecting a different level of trust in the influencer held by agents. It is also the same for all agents, and remains constant throughout each trial. The behavior when an agent interacts with the influencer is similar to the interaction between two agents with the exception that the influencers belief is unchanged and there is no social bond with the influencer to make or break. The probability of an agent interacting with an influencer during each iteration is given by p_Y. The degree by which belief changes during an interaction between agents is proportional to the difference in agent belief, scaled by a constant rate, r, and truncated at either terminal during these interactions to maintain belief on the interval $[-1, 1]$.

3.2 Design of the Simulation

Parameters that remain constant during each simulation trial are: the confidence threshold between agents, CT_X, between an agent and the influencer, CT_Y, the proportional change in the belief of an agent at each interaction, r, and the probability that an influencer will interact with an agent in any iteration p_Y. The model is initialized by randomly assigning a level of belief to each agent chosen from the uniform distribution on the interval $[-1, 1]$. Thus $B(X_i) \sim U(-1, 1)$, for $i \in \{1, 2, 3, ..., n\}$. Agents have no social connections at this stage, thus $N(X_i, X_j) = 0$ for all i, j. At each iteration a pair of agents is randomly chosen for interaction. To simplify notation, it is assumed that that $B(X_i) \geq B(X_j)$ and $\delta_a = B(X_i) - B(X_j)$, so that $\delta_a > 0$. When $\delta_a \leq CT_X$ agents have an assimilative interaction, where each agent's belief moves incrementally toward the other as

$$B(X_i) \leftarrow B(X_i) - r\delta_a \quad \text{and} \quad B(X_j) \leftarrow B(X_j) + r\delta_a,$$

and the network connection between agents is strengthened as $N(X_i, X_j) \leftarrow N(X_i, X_j) + 1$. When $\delta_a > CT_X$ agents engage in a repulsive interaction whereby

$$B(X_i) \leftarrow min(B(X_i) + r\delta_a, +1) \quad \text{and} \quad B(X_j) \leftarrow max(B(X_j) - r\delta_a, -1).$$

The network connection between agents is weakened as $N(X_i, X_j) \leftarrow max(N(X_i, X_j) - 1, 0)$. A second interaction, between a randomly chosen agent and the influencer, then occurs, with probability p_Y. The

difference in belief between the influencer and agent is calculated as $\delta_i = B(X_i) - B(Y)$. When $|\delta_i| \leq CT_Y$ an assimilative interaction occurs, and agent belief moves closer to that of the influencer as

$$B(X_i) \leftarrow B(X_i) - r\delta_i.$$

Otherwise a repulsive interaction occurs in which

$$B(X_i) \leftarrow \begin{cases} min(B(X_i) + r\delta_i, +1) & \text{if} \quad \delta_i \geq 0 \\ max(B(X_i) + r\delta_i, -1) & \text{if} \quad \delta_i < 0. \end{cases}$$

A summary of the social network of agents, and individual agent belief, is then reported periodically. The operation of the simulation is shown in Algorithm 1. The algorithm was implemented in the R programming language, using the igraph package for social network analysis and graph plotting.

3.3 The Simulation Algorithm

Algorithm 1 Agent-Based Model with Influencer

$B(X_i) \sim U(-1,1)$ for all $i \in \{1,2,3,...,n\}$
$N(X_i,X_j) = 0$ for all i, j
set CT_X, CT_Y $B(Y)$, p_Y, randomize
for $I = 1$ to total iterations **do**
    # agent-agent interaction
    randomly choose X_i and X_j where $i \neq j$ and $B(X_i) \geq B(X_j)$
    $\delta_a = B(X_i) - B(X_j)$
    if $\delta_a \leq CT_X$ **then**
        $B(X_i) \leftarrow max(B(X_i) - r\delta_a, -1)$; $B(X_j) \leftarrow min(B(X_j) + r\delta_a, +1)$
        $N(X_i,X_j) \leftarrow N(X_i,X_j) + 1$
    else
        $B(X_i) \leftarrow min(B(X_i) + r\delta_a, +1)$; $B(X_j) \leftarrow max(B(X_j) - r\delta_a, -1)$
        $N(X_i,X_j) \leftarrow max(N(X_i,X_j) - 1, 0)$
    end if
    # influencer-agent interaction
    if $x \sim U(0,1) < p_Y$ **then**
        randomly choose X_i
        $\delta_i = B(X_i) - B(Y)$
        if $|\delta_i| \leq CT_Y$ **then**
            $B(X_i) \leftarrow B(X_i) - r\delta_i$
        else
            if $\delta_i \geq 0$ **then** $B(X_i) \leftarrow min(B(X_i) + r\delta_i, +1)$
            if $\delta_i < 0$ **then** $B(X_i) \leftarrow max(B(X_i) + r\delta_i, -1)$
        end if
    end if
    # periodic reporting
    if I mod reporting interval $= 0$ **then** calculate and report statistics
end for

3.4 Output from the Agent-Based Model

The output from the model at each iteration is each agent's level of belief, $B(X_i)$ and the matrix containing the level of connection between each pair of agents, $N(X_i,X_j)$. From these data, the social network formed

by the agents was constructed and the number of clusters counted. In the simulation trials that follow, this information was recorded at initialization and then after every 1000 periods. Figure 1 shows the evolution of the social network formed during one simulation trial. Agent belief is indicated as a gray-scale with -1 shown by white, 1 by black and mid-gray being neutral. Social connections between agents are represented by edges, and the social network plot shows the formation of clusters of agents. Beginning from a random, disconnected state at initialization ($I = 1$) the progressive polarization of agent belief and the gradual clustering of agents having a similar level of belief can be seen over successive iterations. By $I = 40000$ the social network is completely polarized.

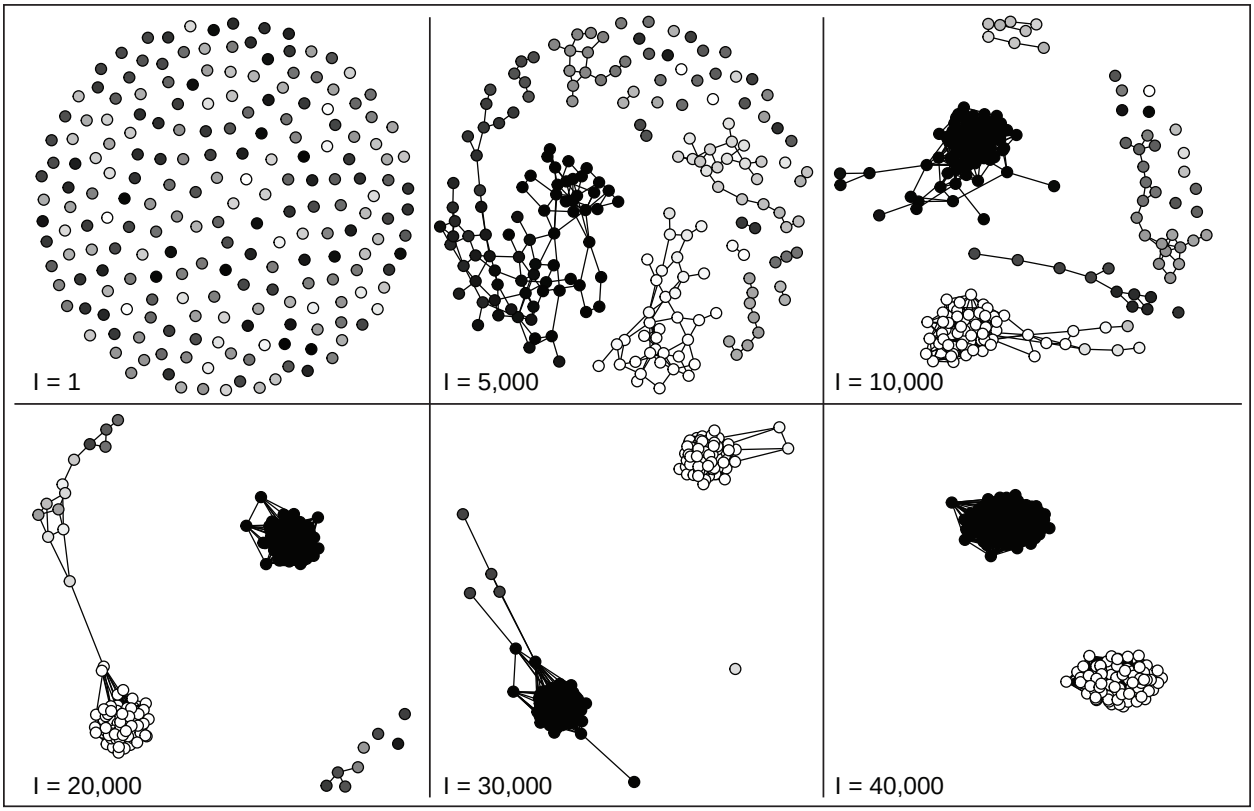

Figure 1: Evolution of the social network from initialization $(I = 1)$ and then at selected intervals, showing the formation of increasingly polarized clusters, and complete polarization at $I = 40,000$ iterations.

Figure 2 presents summary output from the same simulation over successive iterations. The left panel shows the trajectory of each agents belief over the duration of the simulation, where the complete polarization of agent belief can be observed at approximately 35,000 iterations. The center panel shows the number of clusters over the simulation as a natural logarithm, from which it can be observed that the rate of decrease in the number of clusters is approximately exponential. The right panel shows the standard deviation of agent belief over successive iterations, commencing from an initial value of approximately 0.58, being σ_x when $x \sim U(-1,1)$, to approximately 1 when $x = -1$ or 1 with equal probability. In the analysis that follows, the rate at which polarization occurs is determined by two measures obtained from these summary calculations: the number of iterations at which two distinct (polarized) clusters have formed, representing the psychological distance between groups; and the number of iterations by which at least approximately 80% of agents have polarized (that is, hold a belief of either $+1$ or -1), reflecting the ideological separation between groups Bliuc et al. (2021).

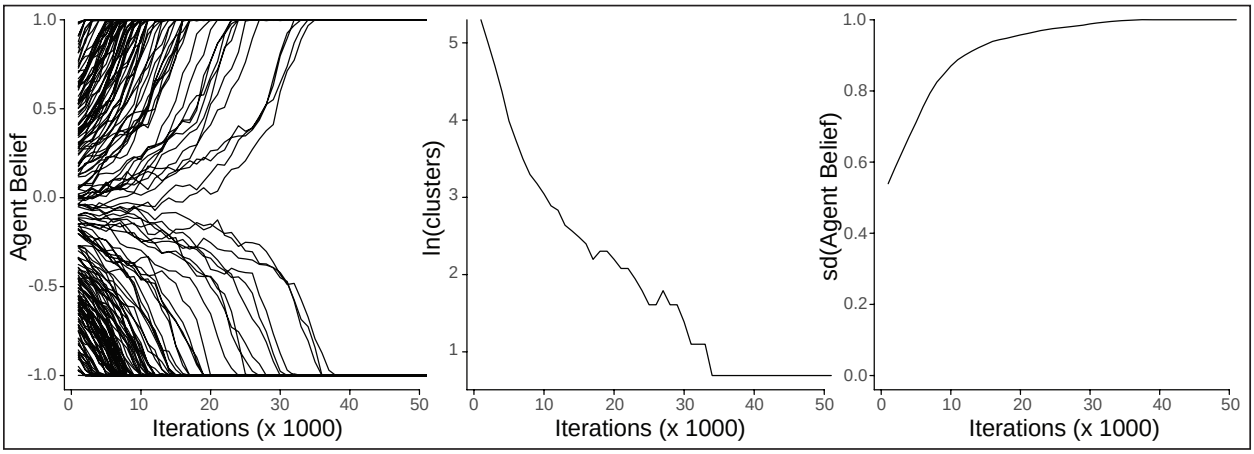

Figure 2: Summary output from a single simulation trial showing, at each iteration: the trajectory of each agents belief (left panel), the natural log number of clusters formed (center panel), and the standard deviation of agent belief (right panel).

4 DESIGN OF EXPERIMENTS AND RESULTS

4.1 Inputs to the Model

Parameter settings for the simulation trials were set as follows: The confidence threshold for agent-agent interactions were chosen from $CT_X \in \{0.1, 0.5, 0.9\}$, reflecting relatively narrow, medium, and wide differences in belief over which agents would have an assimilative interaction and form a social connection. At these values of CT_X, polarization is the natural outcome of the interactions described by Algorithm 1, with $-1 \leq B(X_i) \leq +1$, whereas $CT_X > 1$ would typically result in consensus. The confidence threshold for influencer-agent interactions were also chosen from the same range. Thus, $CT_Y \in \{0.1, 0.5, 0.9\}$. Influencer belief was set as $B(Y) \in \{0, 1\}$, representing the case of a neutral influencer, having belief 0 and an extremist influencer having belief 1. The probability of an influencer-agent interaction in any iteration was $p_Y \in \{0, 0.1, 0.2, 0.5, 1.0\}$, giving the base case, where there was no influencer interaction ($p_Y = 0$) as well as increased levels of influencer activity to the point where there were as many influencer-agent interactions as there were agent-agent interactions. The multiplier for the proportional change in agent belief during each agent-agent or influencer-agent interaction, r, was 0.01 throughout. A population of 200 agents was created. Trials were fully factorial, with 100 repetitions at each parameter combination. Each trial ran for 50,000 iterations, with summary statistics reported every 1,000 periods.

4.2 Measures of Polarization

Two measures are used to evaluate the relative rates of polarization across these varied parameter settings. These are: The first reporting period during which two completely polarized clusters emerged; and the first reporting period at which at least approximately 80% of agents had polarized. The censoring of trials meant that for a small proportion of cases (1.2%) the trial terminated before two clusters were reached (that is, held a belief level of either +1 or -1). In these cases the terminal value of 50,000 iterations was reported. The polarization of 80% of agents was estimated as the period at which the standard deviation of agent belief was 0.93, this being the standard deviation of a population of 200 agents in total, comprising 80 agents having belief $+1$, 80 having belief -1 and the remainder having distribution $B(X_i) \sim U(-1, 1)$. Again, the censoring of trials meant that for a small proportion of cases (1.3%) the trial terminated before a standard deviation of 0.93 was reached. Again, the terminal value was reported in these cases.

4.3 Results

Figure 3 shows the number of iterations by which polarization, in terms of psychological distancing ($n = 2$ clusters), occurs as a function of agent-agent confidence threshold, CT_X, influencer-agent confidence threshold, CT_Y and influencer belief, $B(Y)$. This figure and the next can be viewed as two main panels showing results for the neutral influencer (LHS) and the extremist influencer (RHS). Each panel in the plot shows the effect of increasing the probability of an influencer-agent interaction during each iteration. Ignoring the effect of influencers temporarily by considering the only cases where $p_Y = 0$, the results overall show the rate of clustering increases (that is, the time to form two clusters decreases) as the confidence threshold for agent-agent interactions increases. There is a small, and insignificant increase from $CT_X = 0.1$ ($\bar{x} = 35,293$, $s = 7,986$) to $CT_X = 0.5$ ($\bar{x} = 34,196$, $s = 8,167$). However, there is a significant increase in the rate of polarization when $CT_X = 0.9$ ($\bar{x} = 23,230$, $s = 4,945$), confirmed by a two-sided t-Test (here and following tests), $p \ll 0.001$. This results from the increased likelihood of assimilative iterations (and making network connections) due to the increased confidence threshold, reflecting a more tolerant society.

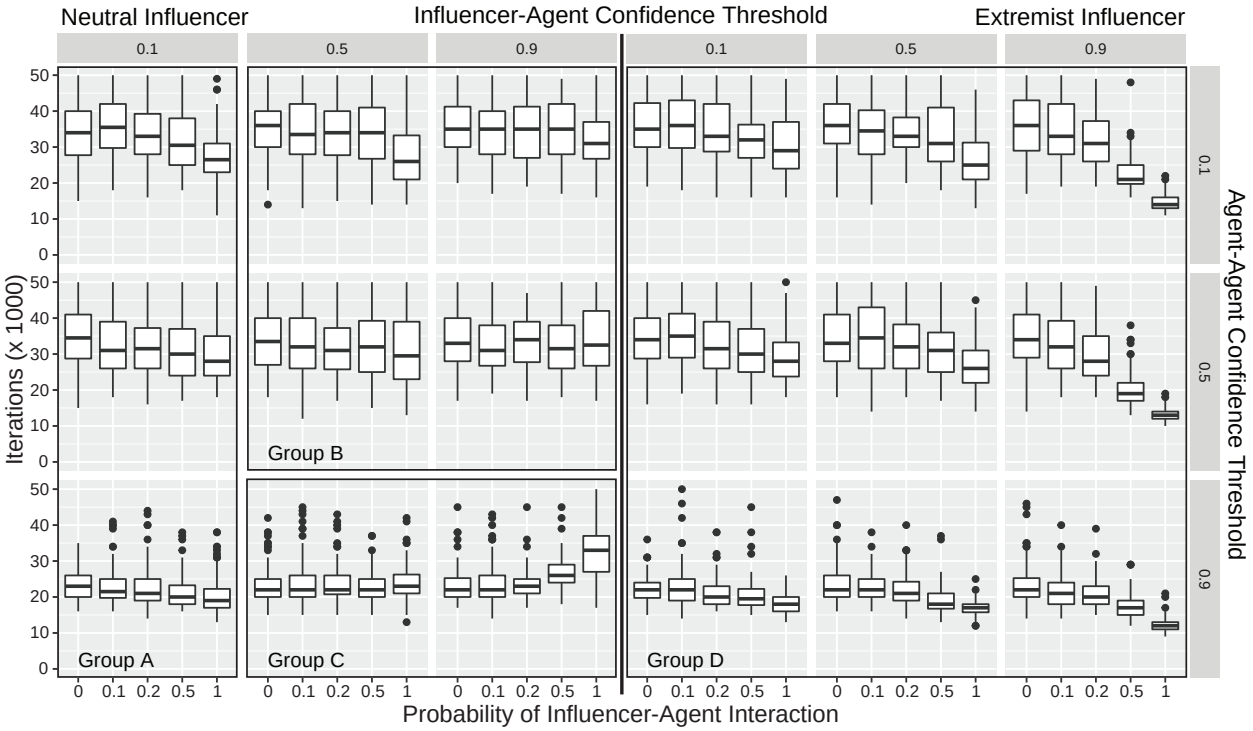

Figure 3: The number of iterations by which 2 clusters have formed, by agent-agent confidence threshold, influencer-agent confidence threshold, probability of influencer-agent interaction and whether the influencer is neutral (LHS) or extremist (RHS).

Subgroups of parameter combinations that exhibit similar properties have been identified for the following analysis and discussion. Groups A, B and C show the period at which polarized clusters have formed for a neutral influencer at each parameter combination. These groups show the effect on cluster formation is mixed. Group A shows that increased interaction with a neutral influencer having a narrow confidence threshold increases the rate of clustering regardless of the confidence threshold for agent-agent interactions over the range tested. This is because the majority of influencer-agent interactions in this situation are repulsive, increasing the rate at which agent's belief polarize and hence increasing the likelihood of assimilative agent-agent interactions. Groups B and C show results when the confidence

threshold for influencer-agent interactions is relatively wider. The results here are mixed. Group B shows that the neutral influencer only increases the rate of clustering by a small amount for the most part, and the increase is not statistically significant. It is only when $CT_X = 0.1$ and $CT_Y = 0.5$ that the increase in $p_Y = 0.5\,(\bar{x} = 33,850,\, s = 9,240)$ to $p_Y = 1.0\,(\bar{x} = 27,910,\, s = 8,289)$ results in a significant increase in the mean time to polarization $p < 0.001$. Group C shows that when the confidence threshold is high for agent-agent interactions $(CT_X = 0.9)$, increasing influencer activity slows cluster formation when the confidence threshold for influencer-agent interactions is high. For example, when $CT_Y = 0.9$, increasing from $p_Y = 0.5\,(\bar{x} = 26,470,\, s = 4,644)$ to $p_Y = 1.0\,(\bar{x} = 33,140,\, s = 7,489)$ results in a significant increase in the average time to polarization $p < 0.001$. Group D shows the periods at which of clusters $(n = 2)$ form for an extremist influencer where it is evident that increasing the rate of influencer interactions always increases the rate of clustering regardless of the confidence threshold for agent-agent or influencer-agent interactions. Due to the consistency of these results they are presented without tests of statistical significance.

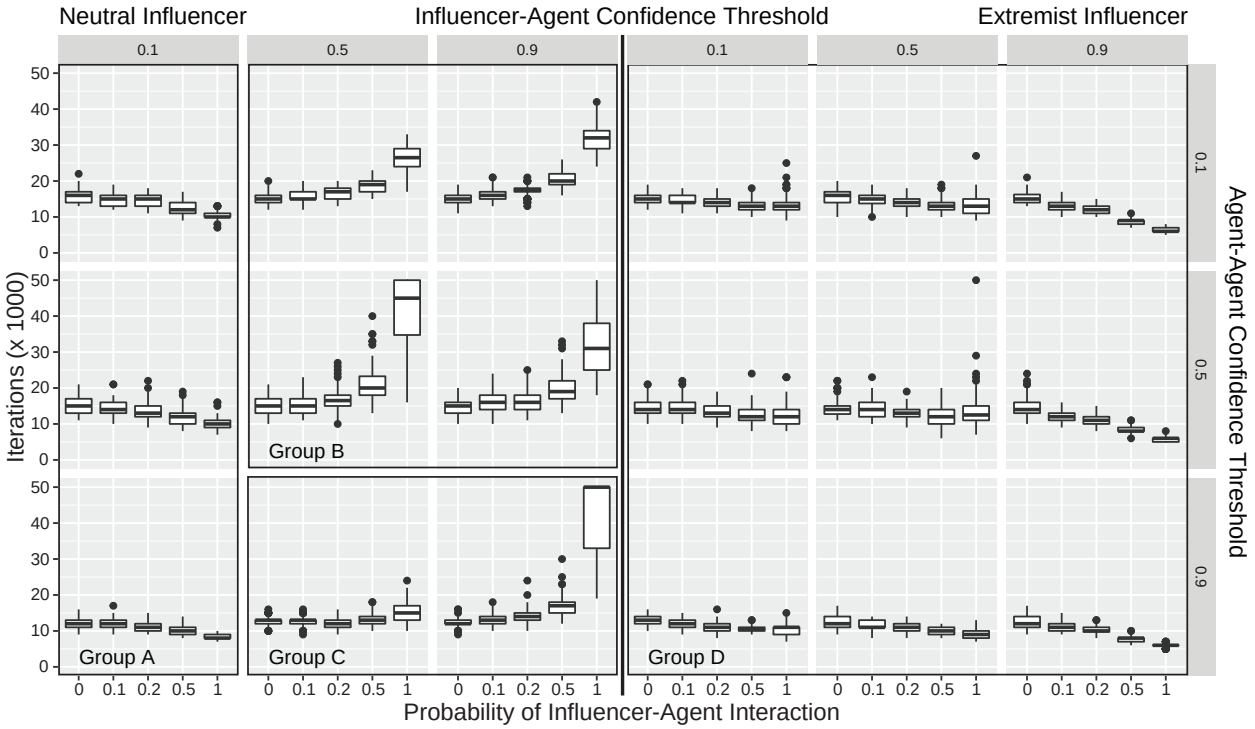

Figure 4: The number of iterations by which agent belief has polarized, (that is, $B(X_i) \approx +1$ or -1) for approximately 80% of agents, by agent-agent confidence threshold, influencer-agent confidence threshold, probability of influencer-agent interaction and whether the influencer is neutral (LHS) or extremist (RHS).

Figure 4 shows the number of iterations by which approximately 80% of agents are polarized in terms ideological extremitization of belief (that is, the standard deviation of agent belief ≥ 0.93), as a function of agent-agent confidence threshold, CT_X, influencer-agent confidence threshold, CT_Y and influencer belief, $B(Y)$. Ignoring the effect of influencers, the results overall show the rate of ideological extremitization increases as the confidence threshold for agent-agent interactions increases. There is a significant decrease in the average number of iterations at which 80% polarization occurs from $CT_X = 0.1\,(\bar{x} = 15,371,\, s = 1,642)$ to $CT_X = 0.5\,(\bar{x} = 14,653,\, s = 2,392)$, $p \ll 0.001$ and again from $CT_X = 0.5$ to $CT_X = 0.9\,(\bar{x} = 12,450,\, s = 1,498)$ to , $p \ll 0.001$. Looking now at the effect of the influencer, Group A shows that increased activity by a neutral influencer $(B(Y) = 0)$ having a narrow confidence threshold for influencer-agent interactions

($CT_Y = 0.1$), always increases the rate of agent polarization due to the influencer-agent interactions being predominantly repulsive. Groups B and C show that when there is a wider confidence threshold for influencer-agent interactions, the increased activity by the neutral influencer always reduces the rate of polarization. The censoring of trials at $50,000$ iterations is evident for these two groups, indicating the extreme retardation of belief extremitization when influencer activity is high, that is, when $p_Y = 1.0$. Group D shows that in the case of an extremist influencer the rate of polarization always increases in line with an increased rate of influencer-agent interactions. Due to the consistency of results within each group, they are presented without tests of statistical significance.

5 SUMMARY AND CONCLUSION

Taking the results in Figures 3 and 4 together, Group D shows that increased activity by extremist influencers always increases the rate of polarization by increasing both the rate of ideological separation (polarization of belief), and the rate of psychological distancing (the formation of polarized clusters) regardless of the confidence threshold for agent-agent or influencer-agent interactions. This result confirms the commonly held concern that messaging by extremists of any political persuasion or ideology will increase the likelihood of polarization within communities.

The effect of a neutral influencer is less straightforward. Group A shows that a neutral influencer having a narrow confidence threshold for influencer-agent interactions will increase the rate of polarization, due to the largely repulsive influence of the interaction. This accords to real world experience when, in situations of ideological division (for example, on the issue of climate change) even when confronted by ostensibly neutral opinion leaders (such as scientists), people in opposing ideological camps may still maintain their belief regardless of the evidence in the influencers message. We think there are two possible reasons for this. Firstly, even a neutral message (that is, equidistant from either pole) may still be polarizing. For example, a balanced message by a scientist on climate change may still lead to confirmation bias in the pro-environmental camp (resulting in psychological distancing), and dis-confirmation bias in the deniers camp, with information more likely to be endorsed by the outgroup as unreliable (increasing ideological extremitization). Secondly, ineffective, non-charismatic influencers tend to be counter-productive as they increase the rate of polarization. Information from an unpopular (unconvincing) communicator is likely to be seen as an outsider, and hence unreliable, by both camps, leading to increased polarization in both.

Groups B and C show that increased activity by a neutral influencer with a moderate to wide reach will always reduce the rate of ideological extremitization, that is the polarizing of agent belief. Although the effect of the neutral influencer on psychological distancing (clustering) is neutral when agents have a fairly narrow confidence threshold for agent-agent interactions, Group C shows that when the confidence threshold is wide, cluster formation is slowed. In either case, an important observation is that these clusters contain agents that are less polarized (that is, more moderate) in belief than would be the case if interaction with the influencer was reduced. In this regard, Group B can be thought of as being similar to a conservative community, where individuals form more moderate clusters, even though the rate of clustering is similar to society as a whole. Group C represents a very tolerant community, where a neutral influencer reduces the rate of clustering to a small degree, and has the potential to reduce the rate of ideological separation significantly. Again, even though clusters form in this community, they contain more moderate members than would be the case without the neutral influencer.

Our findings suggest that rather than trying to stop or prevent polarization (which most research indicates is ineffective), evidence-based communication strategies could be used to slow the rate of ideological extremitization. For example, education campaigns may be employed to increase the publics tolerance to diverse opinions in combination with moderate messaging from charismatic opinion leaders.

A limitation of the current work is that in the real world, individuals are confronted by multiple issues of varying significance simultaneously. In this situation, the tendency to polarize on a single issue may be overwhelmed by competing issues. Notwithstanding, this research offers an insight into the mechanism by which dissent may lead to polarization, and the effect of influencers in accelerating or mitigating that

process. Although different societal issues in varying social contexts will each require a unique strategy to address polarization, a greater understanding of the processes by which polarization occurs, and the role played by influencers, can only help improve public strategies to address this societal concern. This project has only employed a small range of parameter values to describe the artificial society. Future research includes testing the robustness of our current conclusions across a broader parameter space.

6 ACKNOWLEDGEMENTS

The authors would like to thank Professor David Green and Mr Robert Milligan for productive discussions about the development of related Agent-Based Models.

REFERENCES

Abramowitz, A., and J. McCoy. 2019. "United States: Racial resentment, negative partisanship, and polarization in Trumps America". *The Annals of the American Academy of Political and Social Science* 681(1):137–156.

Asatani, K., H. Yamano, T. Sakaki, and I. Sakata. 2021. "Dense and influential core promotion of daily viral information spread in political echo chambers". *Scientific reports* 11(1):1–10.

Axelrod, R. 1997. "The dissemination of culture: A model with local convergence and global polarization". *Journal of conflict resolution* 41(2):203–226.

Bakshy, E., J. Hofman, M. Winter, and D. Watts. 2011. "Everyone's an influencer: quantifying influence on twitter". In *Proceedings of the fourth ACM international conference on Web search and data mining*. February 9th-12th, Hong Kong, China, 65–74.

Bliuc, A.-M., A. Bouguettaya, and K. D. Felise. 2021. "Online Intergroup Polarization Across Political Fault Lines: An Integrative Review". *Frontiers in Psychology* 12:1–15.

Cinelli, M., G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini. 2021. "The echo chamber effect on social media". *Proceedings of the National Academy of Sciences* 118(9):1–8.

Deffuant, G., D. Neau, F. Amblard, and G. Weisbuch. 2000. "Mixing beliefs among interacting agents". *Advances in Complex Systems* 3(01n04):87–98.

Druckman, J. N., and M. S. Levendusky. 2019. "What do we measure when we measure affective polarization?". *Public Opinion Quarterly* 83(1):114–122.

Duffy, B., K. A. Hewlett, J. McCrae, and J. Hall. 2019. "Divided Britain? Polarisation and fragmentation trends in the UK". *The Policy Institute, Kings College London*:1–108. https://www.kcl.ac.uk/policy-institute/assets/divided-britain.pdf, accessed 25th March 2022.

Dykstra, P., W. Jager, C. Elsenbroich, R. Verbrugge, and G. R. De Lavalette. 2015. "An agent-based dialogical model with fuzzy attitudes". *Journal of Artificial Societies and Social Simulation* 18(3):3.

Festinger, L. 1962. *A Theory of Cognitive Dissonance, Vol. 2. Redwood*. USA: Stanford University Press: Stanford, CA.

Flache, A., and M. W. Macy. 2011. "Small worlds and cultural polarization". *The Journal of Mathematical Sociology* 35(1-3):146–176.

Flache, A., M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, and J. Lorenz. 2017. "Models of social influence: Towards the next frontiers". *Journal of Artificial Societies and Social Simulation* 20(4):1–32.

Friedkin, N. E., and E. C. Johnsen. 2011. *Social influence network theory: A sociological examination of small group dynamics*, Volume 33. Cambridge University Press.

Groeber, P., J. Lorenz, and F. Schweitzer. 2014. "Dissonance minimization as a microfoundation of social influence in models of opinion formation". *The Journal of Mathematical Sociology* 38(3):147–174.

Ishii, A., and Y. Kawahata. 2018. "Opinion dynamics theory for analysis of consensus formation and division of opinion on the internet". 1–11. https://arxiv.org/abs/1812.11845, accessed 22nd June 2022.

Macal, C. M. 2016. "Everything you need to know about agent-based modelling and simulation". *Journal of Simulation* 10(2):144–156.

McCoy, J., T. Rahman, and M. Somer. 2018. "Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities". *American Behavioral Scientist* 62(1):16–42.

McCoy, J., and M. Somer. 2019. "Toward a theory of pernicious polarization and how it harms democracies: Comparative evidence and possible remedies". *The Annals of the American Academy of Political and Social Science* 681(1):234–271.

McPherson, M., L. Smith-Lovin, and J. M. Cook. 2001. "Birds of a feather: Homophily in social networks". *Annual review of sociology* 27(1):415–444.

Modgil, S., R. K. Singh, S. Gupta, and D. Dennehy. 2021. "A confirmation bias view on social media induced polarisation during Covid-19". *Information Systems Frontiers*:1–25.

Myers, D. G. 1975. "Discussion-induced attitude polarization". *Human Relations* 28(8):699–714.
Myers, D. G., and H. Lamm. 1976. "The group polarization phenomenon.". *Psychological bulletin* 83(4):602.
Nowak, A., J. Szamrej, and B. Latané. 1990. "From private attitude to public opinion: A dynamic theory of social impact.". *Psychological review* 97(3):362.
Ramírez-Dueñas, J. M., and M. L. Vinuesa-Tejero. 2021. "How does selective exposure affect partisan polarisation? Media consumption on electoral campaigns". *The Journal of International Communication* 27(2):258–282.
Riedl, M., C. Schwemmer, S. Ziewiecki, and L. M. Ross. 2021. "The Rise of Political InfluencersPerspectives on a Trend Towards Meaningful Content". *Frontiers in Communication* 6:1–7.
Rychwalska, A., and M. Roszczyńska-Kurasińska. 2018. "Polarization on social media: when group dynamics leads to societal divides". In *Proceedings of the 51st Hawaii International Conference on System Sciences*. January 3rd-6th, Hilton Waikoloa Village, Hawaii, 2088–2097.
Sherif, M., and C. I. Hovland. 1961. *Social judgment: Assimilation and contrast effects in communication and attitude change.* Yale University Press.
Skyrms, B., and R. Pemantle. 2009. "A dynamic model of social network formation". In *Adaptive networks*, 231–251. Springer.
Smith, E. R., and F. R. Conrey. 2007. "Agent-based modeling: A new approach for theory building in social psychology". *Personality and social psychology review* 11(1):87–104.
Soares, F., R. Recuero, and G. Zago. 2018. "Influencers in polarized political networks on Twitter". In *Proceedings of the 9th international conference on social media and society*. July 18th-20th, Copenhagen, Denmark, 168–177.
Tur, B., J. Harstad, and J. Antonakis. 2021. "Effect of charismatic signaling in social media settings: Evidence from TED and Twitter". *The Leadership Quarterly*:101476.
Turner, M. A., and P. E. Smaldino. 2018. "Paths to polarization: How extreme views, miscommunication, and random chance drive opinion dynamics". *Complexity* 2018:1–17.

AUTHOR BIOGRAPHIES

JOHN M. BETTS is a Senior Lecturer and Director of Education in the Department of Data Science and Artificial Intelligence, Faculty of Information Technology at Monash University, Australia. His research applies computational modeling, optimization and simulation across diverse fields. Ongoing research with social and political scientists investigates the language use and social dynamics of online white supremacist communities to better understand the spread of influence and social cohesion in these groups. His e-mail address is john.betts@monash.edu. His website is https://research.monash.edu/en/persons/john-betts.

ANA-MARIA BLIUC is a social and political psychologist at the University of Dundee. She has a PhD in Psychology from the Australian National University. Her research examines how peoples social identities influence their behavior in a range of contexts including health, environmental (climate change), and socio-political (collective action and social change). Research into online communities (including far-right and white supremacist, and recovery from addiction) has examined how collective identities and behaviors are shaped through online interactions. Her email address is ABliuc001@dundee.ac.uk. Her website is https://www.dundee.ac.uk/people/ana-maria-bliuc.