

How polarization extends to new topics: an agent-based model derived from experimental data - model explanation

[Anonymised review copy]

Abstract:

Polarization is an key phenomenon which has been linked to increasing disliking between people of opposite political groups. Furthermore, polarization can extend to new topics such as the debate on COVID-19 vaccines, making more complex to coordinate efforts for such a problem.

The social identity approach (SIA) offers a robust theoretical framework for understanding identity-based social processes. This approach suggests that people's perceptions and behaviour depend on their group identity (e.g. Democrat vs Republican).

In this article, we developed an opinion-dynamics model integrating SIA to explore how polarization can extend to new topics. Furthermore, we developed this model from empirical experiments. This allows us to use already validated micro-dynamic rules.

Experimental results show lack of repulsive effects, more attraction during in-group interactions and a new effect: increased stubbornness when people are exposed to opinions of an out-group member.

The model was built mimicking the interaction structure of the experiment. At each iteration, an agent observes the opinion of another agent. Depending on their respective groups the agent will experience a stronger or weaker attractive force, together with additional noise.

This model was able to produce polarization without the use of repulsive forces. Furthermore, the sensitivity analysis tells us that polarization in new topics can appear when all the following conditions are satisfied: (1) people should recognize their belonging to political groups, (2) they should have more in-group than out-group interactions and (3) there should be some initial asymmetry on the topic.

Keywords: experimental validation, micro-dynamic rule, opinion dynamics, update rule

● Agent-based model

Model properties

- 1.1 Following the previous results, we built an agent-based model to reproduce the observed behavior and see which kind of dynamics it can generate in case of repeated interactions. The main purpose of this model would be to study how polarization can appear in novel topics and its relationship with social groups. As previously mentioned, the biggest advantage of building a model from experimental data lies in the fact that the micro-dynamic rule (i.e. how agents update their own opinion) is already validated. Indeed, as we will discuss, most of the parameter values used in the model come from the values summarized in table 1.
- 1.2 We coded the model in python. The code is available at [LINK](#) in the form of a Jupyter notebook for ease of exploration. For the simulations, we used 1,000 agents divided into two groups of 500 each (called group A and group B). For each model iteration, two agents are selected randomly. They firstly check if they belong to the same group and, therefore, if they will have an in-group or an out-group interaction. Afterwards, they update their opinion using the values of shift and noise belonging to the type of interaction. As previously mentioned, shift is modeled as a constant opinion change. Therefore if the initial opinion is -0.3 and the agent is exposed

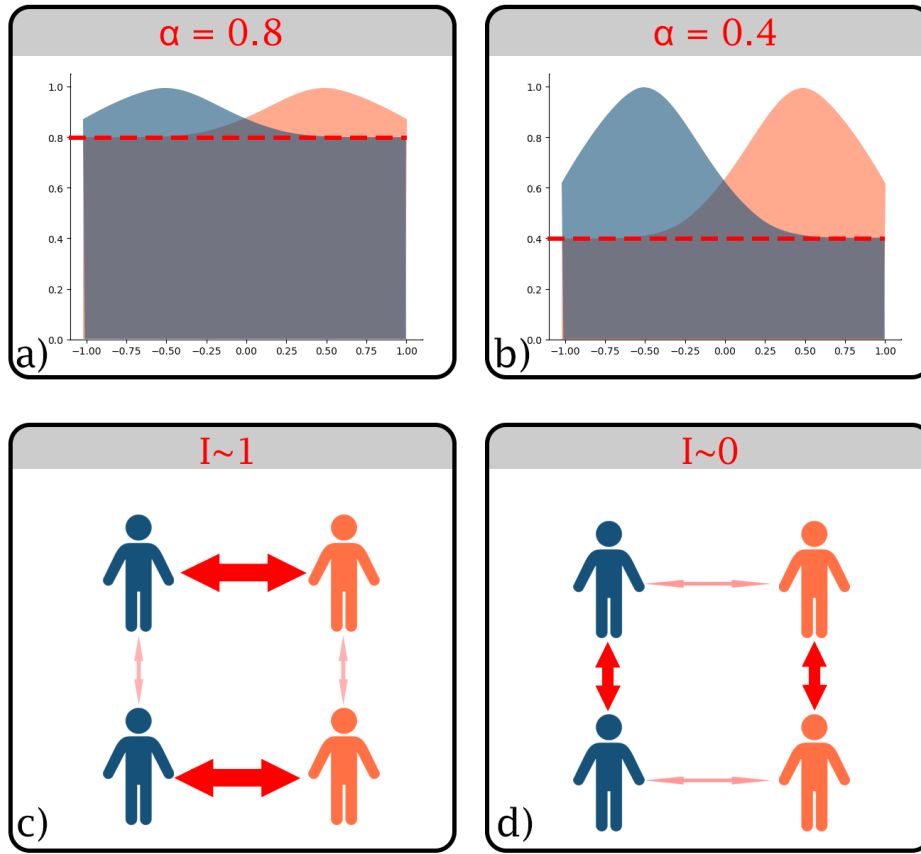


Figure 1: Visual representation of the parameters α and I . a) and b) how α affects the initial distribution. c) and d) scheme of how I affects the probability of in-group and out-group interaction.

with agreement (i.e. +1), for a shift of 0.068, the agent will update her opinion to $-0.3 + 0.068 = -0.232$. On top of the shift, each agent will also experience some noise in the interaction, modeled as a uniform distribution of amplitude specified by the type of interaction.

1.3 The values representing shift and noise are obtained from the experimental data in the following conditions:

1. "Minimal groups" uses the values of the table for the control setting (i.e. Klee and Kandinsky)
2. "Political groups" instead uses the values from the experimental setting (i.e. Republicans and Democrats)

Besides these conditions, we introduced two other parameters for exploring the model. The first one is the initial opinion distribution. Indeed, while in opinion dynamics it is common to assume uniform distributions for the initial opinion, in this case would be interesting to know what happens if there is an initial bias (or, similarly a correlation) between the two groups and the expressed opinion. This will allow us to see how this will impact the appearance of polarization in new topics.

To produce this asymmetric distribution, we used a combination of both a normal and a uniform distribution (see figure 1). In formula:

$$f(o) = (1 - \alpha) * N(o)_{\mu, \sigma} + \alpha * U(o) \quad (1)$$

where $f(o)$ is the overall opinion distribution, $N(x)$ is the normal distribution, and U is the uniform distribution. While for σ we selected 0.5 for both groups, we choose $\mu = 0.5$ for group A and $\mu = -0.5$ for group B to produce the initial asymmetry in the data. In this way, when $\alpha = 1$ the two distribution would be perfectly identical (i.e. two uniform distributions). Instead, for $\alpha = 0$ we would have two normal distributions centered in +0.5 and -0.5, thus maximizing initial asymmetry.

The final parameter that we would like to study is the amount of in-group interactions (as opposed to the amount of out-group interactions). We formalized this into the parameter I which represents the probability of having in-group interactions. Therefore, for $I = 1$ agents will only interact with in-group members, while

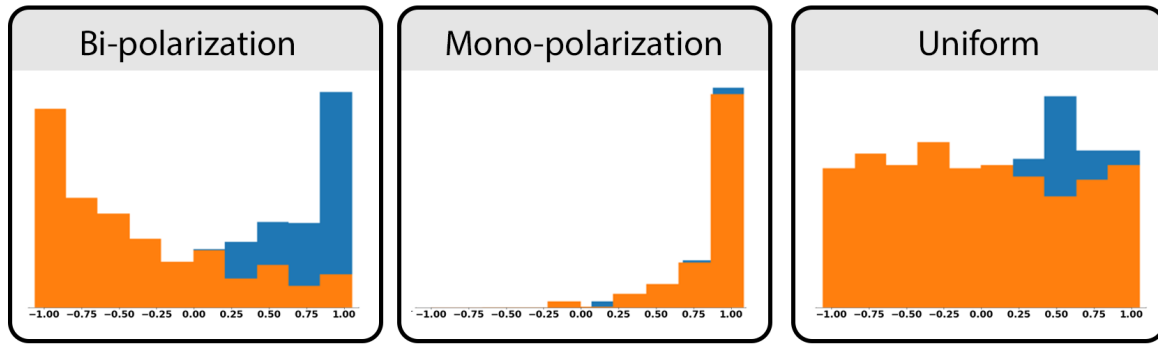


Figure 2: The three main opinion distribution after convergence for group A (blue) and group B (orange).

for $I = 0$ they will interact only with out-group members, while for $I = 0.5$ we will have equal probability of having in-group and out-group interactions.

For each run we let the model evolve for 100,000 iterations (i.e. every agent on average will interact 200 times) as this was sufficient for the model to reach convergence.

Model convergence and sensitivity analysis

- 1.4 Before moving to the sensitivity analysis, we preferred to run an intermediate phase of manual exploration of the model. This helped in the interpretation of the results from the sensitivity analysis.
- 1.5 In this phase, we found three main possible final distributions, which we will refer to as bi-polarization, mono-polarization and uniform (see fig 2).
- 1.6 Bi-polarization happens when the two distributions move to opposite extremes (fig 2). Qualitatively this configuration also coincides with the common interpretation of political polarization (Lelkes 2016). As we will discuss after the sensitivity analysis, this configuration appears often when people interact mostly with in-group members.
- 1.7 Mono-polarization, instead represents the case in which both populations move to the same extreme. This configuration appears mostly when agents interact with out-group members. Indeed, due to the lack of repulsive forces, out-group members are still able to influence each other and, with enough interactions, behave as a unique cohesive group.
- 1.8 Finally, the uniform case appears when the interaction between agents is too weak with respect to the effect of noise. In this case dominated by noise, both groups "converge" to a uniform distribution. As can be seen in figure 2 this will not produce a perfectly uniform distribution. Indeed, due to random fluctuations we will constantly have the formation of temporary peaks.
- 1.9 Finally, we run the sensitivity analysis to better understand how the selected parameters will influence the final outcome. Every point in figures 3 is the average of 10 runs, meaning that we repeated each simulation 10 times without changing the values of α and I , but re-initializing the simulation each time.
- 1.10 Since our main goal is to study the appearance of political polarization in novel topics, we introduced the polarization measurement as the difference of the mean of the two final distributions, as it is often calculated in the literature (Lelkes 2016; Pew Research 2014). In formula:

1.11

$$P = |Mean_A - Mean_B| \quad (2)$$

- 1.12 In the in figure 3 we reported P_{mean} calculated as the average value of P calculated across multiple runs having same values of α and I .
- 1.13 To provide more insight, we offered an alternative measure P_{th} which represented the probability that P would be above a threshold of 0.3 at the end of the simulation. We introduced this value to test if some parameter combinations can actually generate situations in which for some run we have strong polarization, while for other runs (still having the same parameters) we have little or no polarization.

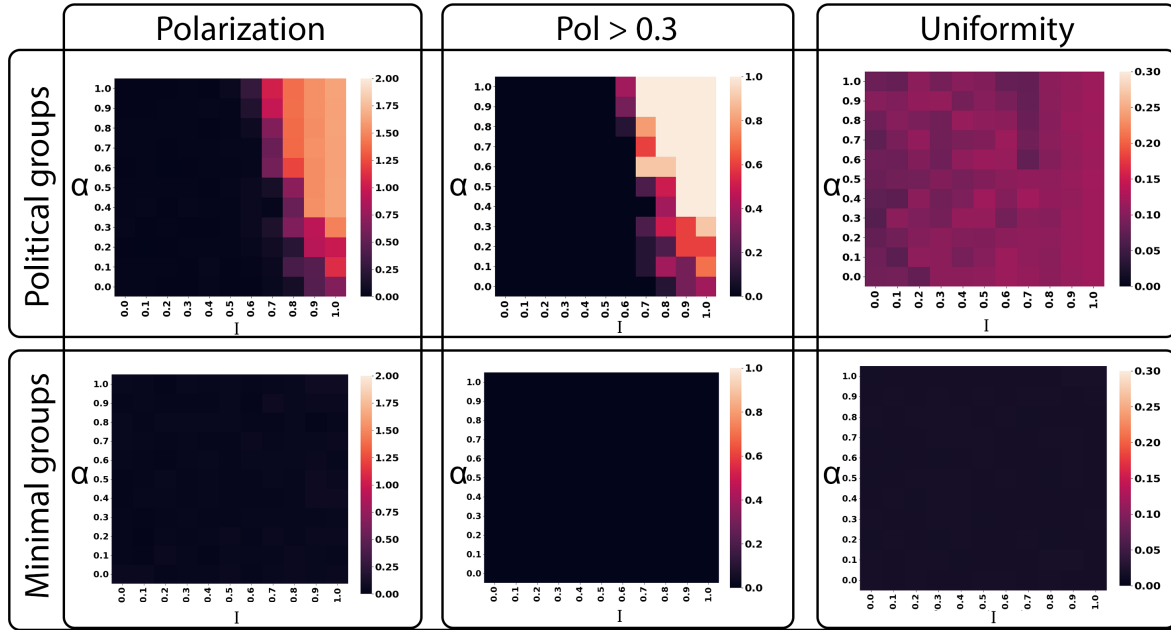


Figure 3: Heatmaps for P , P_{th} and u while varying α (i.e. initial asymmetry) and I (i.e. in-group interactions).

1.14 Finally, we introduced the uniformity parameter u . We introduced it for distinguishing cases of mono-polarization from cases in which both distributions are roughly uniform (as in both cases the polarization parameter would be close to 0). We calculated u for a distribution D as:

1.15

$$u(D) = \sum_{bins} |h(D) - \text{mean}(h(D))| \quad (3)$$

1.16 where h is the histogram of the distribution D . Therefore $u = 0$ when the histogram is flat (i.e. uniform distribution) and progressively bigger as the distribution deviates from the uniform one.

1.17 The top part of figure 3 shows the results for the case of political groups (i.e. using the values of shift and noise for the experimental setting). As seen, when $I < 0.5$ (i.e. people interact mostly with out-group members) polarization cannot appear. However, as people start interacting more and more with their in-group members, initial asymmetry can give rise to the appearance of polarization. In extreme cases in which people interact only with in-group members, even in the case of two initially uniform distributions, these can still give rise to polarization simply by random fluctuations. Furthermore, by observing the plot of the u parameter, we see that with political groups we always achieve either bi- or mono-polarization.

1.18 The situation is very different when, instead, we used the minimal groups configurations. In this case, the attraction between agents is so weak that the entire dynamic is dominated by the random movement. Because of that, no configuration is sufficient for producing polarization. This is also confirmed by the fact that the u parameter is way smaller than the one for political groups.

Acknowledgements

2.1 This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (grant agreement No 891347) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 802421).

References

Lelkes, Y. (2016). Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(S1), 392–410

Pew Research, C. (2014). Political polarization in the american public