9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Polarization without Echoes? An Agent-Based Framework for Modeling Affective Polarization under Exposure to Diverse Content

Keywords: Affective Polarization, Agent-Based, Sorting, Social Media, Echo Chambers

Extended Abstract

Affective polarization or inter-party hostility is increasingly perceived to pervade democracies worldwide and debilitate social cohesion (1). Several studies have, in the past, incriminated the digital media ecosystem – increasingly ubiquitous and accessible – serving as a catalyst to accelerate the same (2; 3). However, the transpiring causal mechanisms instrumental in enabling this phenomenon have been a subject of extensive debate; while the creation of echo chambers and filter bubbles devoid of counter-attitudinal content has long served as the primary hypothesis (4; 5), there is accumulating empirical evidence to suggest otherwise (6). In this study, we address this debate by using an agent-based model to understand how affective polarization may emerge even when individuals are exposed to a diversity of opinions, and do not reside within ideological echo-chambers. To that end, we parameterize two key aspects of how individuals interact with media: one, the affective asymmetry of their engagement with in-party versus out-party content; and two, the proportion of in-party content they are exposed to. We then observe the macro-level changes in affective polarization in the population under various conditions as stipulated by these parameters.

We consider a synthetic group of individuals of size N, each with an allegiance to one of two parties ($a_i \in [1,2]$) that is assumed to remain fixed throughout the course of our study. The individuals are also equipped with feelings of loyalty (in-party affect(IPA_i)) and hostility (out-party affect(OPA_i)) derived from their respective partisan affiliations. Their values are chosen on a feeling thermometer scale of 0 to 100, with values greater than 50 indicating warmth and values less than 50 indicating indifference (1). At every time-step, individuals chosen with a predetermined probability (p) are exposed to pro- and counter-attitudinal content put out by news outlets. The change (Δ_i) in individual feelings upon exposure to news content is determined by accounting for the dissimilarity between the partisan orientation of an individual and the slant of the news content that they are exposed to ($n_i(t)$). The consequent impact of exposure on their feelings towards or against a party is given by equations 1-3,

$$IPA_i(t+1) \;:=\; IPA_i(t) \;+\; \Delta_i(t), \tag{1}$$

$$OPA_i(t+1) \;:=\; OPA_i(t) \;+\; \Delta_i(t), \tag{2}$$

$$\text{where, } \Delta_i(t) \;:=\; -\alpha - \;10 \times \frac{\exp\left(2(a_i - n_i(t) - 0.5)\right) - 1}{\exp\left(2(a_i - n_i(t) - 0.5)\right) + 1}. \tag{3}$$

$\alpha \in [0,10]$ is a constant chosen to vary the severity of individual response to pro- and counter-attitudinal news content. When $\alpha = 0$, individual response towards in-party and out-party content is symmetric, and therefore the exposure to pro- and counter-attitudinal contents enhance and efface the respective feelings by the same amount (Panel A). This is unlike higher values of α that signify different levels of intolerance to content diverse from one's own partisan alignment; when $\alpha = 10$, exposure to content slightly diverse from one's own alignment erodes good-will, with the effect exacerbated upon exposure to counter-attitudinal content (Panel A). Thus, individual's consequent emotional distance towards their in-party and out-party feelings is quantified by an affective distance (AD_i), and affective polarization (AP) or the inter-party

hostility is operationalized as the mean AD for the entire population,

$$AD_i(t+1) := IPA_i(t+1) - OPA_i(t+1), \tag{4}$$

$$AP(t) := \frac{\sum_{i=1}^{N} AD_i(t)}{N}. \tag{5}$$

In our first simulation experiment, we demonstrate trends in the variation of AP due to affective asymmetry of individual engagement with varied news content. To do so, we consider a group of 1000 agents and carry out 100 simulations by randomising initial partisan affiliations and feelings. The results obtained for a few chosen values, $\alpha = \{0, 3, 7, 10\}$, are shown in Panel B. We observe that extreme response to counter-attitudinal news exposure significantly increases AP; a 34% increase in AP when α is increased by 10 can be noticed in Panel B.

In the second simulation experiment, we study the impact of varied levels of selective exposure to news content on individual feelings. For each value of α, we vary the probability (β) of exposure to pro-attitudinal content in a group of size 1000, and tabulate the AP observed from 100 simulation runs. The variation in AP for $\alpha = \{0, 3, 7, 10\}$ and $\beta = \{0, 0.25, 0.5, 0.75, 1\}$ with time is as shown in Panel C. It can be observed that, when $\alpha = 0$, the AP is unaltered due to balanced individual response to pro- and counter-attitudinal content. However, slight asymmetry in response induced when $\alpha = 3$, results in significant variations in AP for different values of β. While exposure to pro-attitudinal content alone ($\beta = 1$) results in minimal AP, counter-attitudinal content ($\beta = 0$) serves to increase hostility by almost 67%. With increased asymmetry in response ($\alpha = \{7, 10\}$), AP is more pronounced and accelerated.

Our proposed model thus shows a potential causal pathway that addresses the empirical puzzle that has been well documented in the literature: the rise of affective polarization in the absence of ideological echo-chambers. Our findings indicate that individuals need not be ensconced within echo-chambers for a population to get affectively polarized. They only need to display affective asymmetry when engaging with partisan content. If out-partisan content invokes a higher degree of outrage than in-partisan content invokes feelings of warmth ((7)), affective polarization may manifest. Additionally, the severity in response to exposure of diverse news content modelled by equation (3) provides for a more nuanced understanding. In our completed study, we will extend this model to account for interpersonal influence, and further factor in skewed sharing patterns of partisan content to more closely replicate the dynamics of affective polarization in the real world.

References

[1] Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. The origins and consequences of affective polarization in the united states. *Annual review of political science* **22**, 129–146 (2019).

[2] Lelkes, Y., Sood, G. & Iyengar, S. The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science* **61**, 5–20 (2017).

[3] Allcott, H., Braghieri, L., Eichmeyer, S. & Gentzkow, M. The welfare effects of social media. *American Economic Review* **110**, 629–676 (2020).

[4] Sunstein, C. R. *# Republic: Divided democracy in the age of social media* (Princeton University Press, 2018).

[5] Dubois, E. & Blank, G. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, communication & society* **21**, 729–745 (2018).

[6] Guess, A., Nyhan, B., Lyons, B. & Reifler, J. Avoiding the echo chamber about echo chambers. *Knight Foundation* **2**, 1–25 (2018).

[7] Brady, W. J. & Crockett, M. J. How effective is online outrage? *Trends in cognitive sciences* **23** (2019).

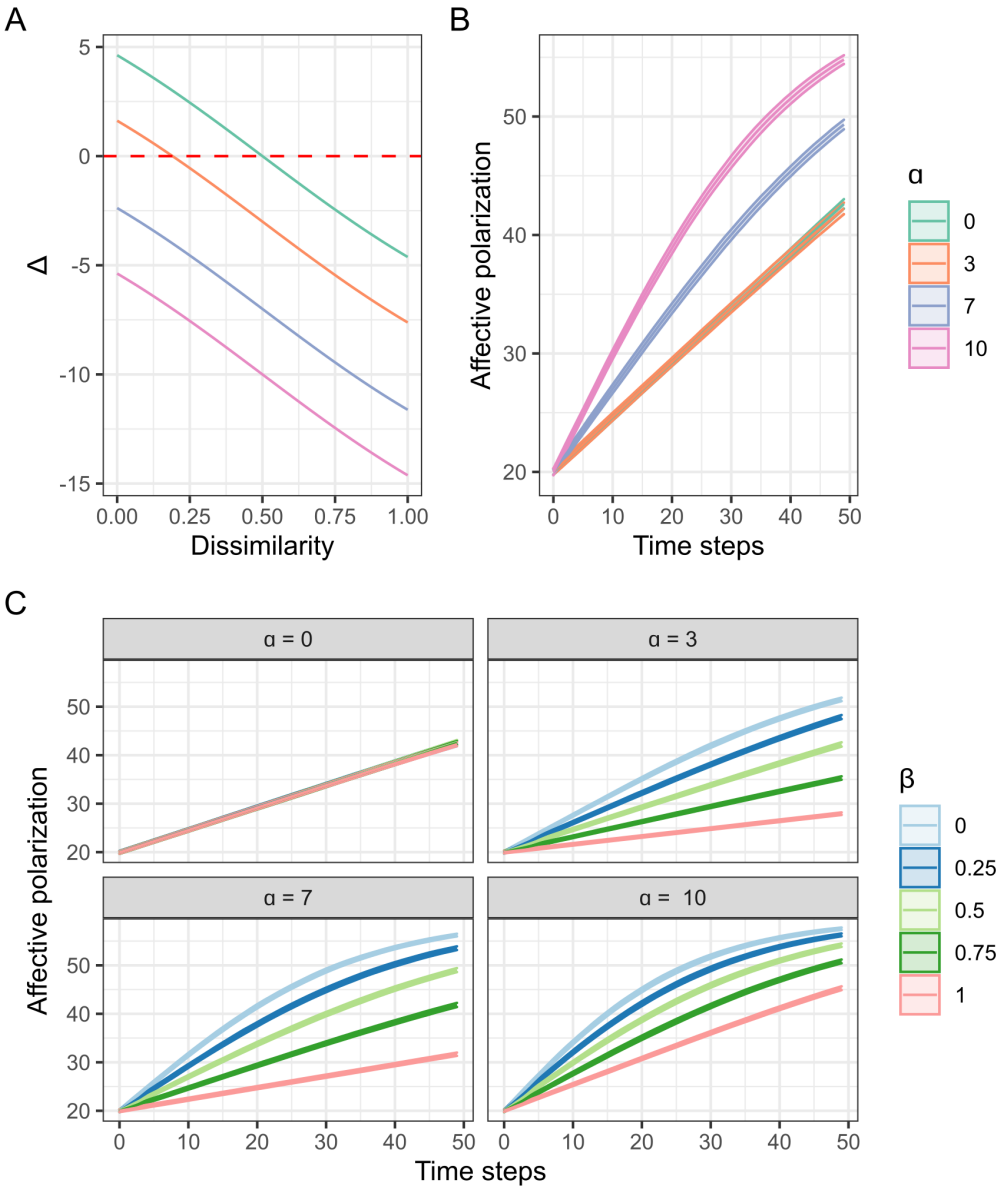

Figure 1: (A) Individual response to news content for varying severity levels; (B) Variation in affective polarization with different levels of severity in individual response to news content; (C) Variation in affective polarization with severity in individual response to news content and selective exposure