UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
A bias-tracking model of rational political polarization.

Permalink
https://escholarship.org/uc/item/3qj5s09p

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Author
Young, David J.

Publication Date
2026

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org
Powered by the California Digital Library
University of California

A bias-tracking model of rational political polarization

David J. Young (d.young@bbk.ac.uk)
School of Psychological Sciences, Birkbeck College, Malet Street
London, WC1E 7HX UK
Department of Psychology, University of Cambridge, Downing Place
Cambridge, CB2 3EB UK

Abstract

Evidence suggests the US population has polarized over politics in recent decades, as the issue positions of America's ideological sub-groups have moved apart. I use simulations to demonstrate a plausible mechanism by which political polarization can occur among 'fully' Bayesian agents. The agents learn from the testimony of competing sources using an empirically-validated Bayesian cognitive model, and do not communicate. Polarization occurs under conditions resembling real-world political debate, and the polarization produced shares several characteristics of real-world US polarization. The core of the mechanism is that agents attempt to infer and account for the bias of the information sources, but doing so creates a feedback spiral where polarized source perceptions drive polarized beliefs and vice versa, unless there is strong disambiguating evidence in one direction or another.

Keywords: rational; bayesian; belief; polarization; bias; simulation.

Introduction

Research on the rationality of belief polarization is at a crossroads. On one hand, cognitive science has increasingly recognized that there are pathways by which belief polarization can occur that are rational (Jern, Chang, & Kemp, 2014; Hahn & Harris, 2014; Bullock, 2009; Olsson & Vallinder, 2013; Benoît & Dubra, 2019), or near-rational, perhaps admitting some reasonable cognitive limitations for the sake of efficiency, like forgetting or processing limits (Singer et al., 2019; Mäs & Flache, 2013; Van Doorn, 2024; O'Connor & Weatherall, 2018). Yet, most accounts of rational belief polarization are only theoretical, with empirical evidence limited (for exceptions see Cook & Lewandowsky, 2016; Jern et al., 2014; Young, Madsen, & de Wit, 2025). Furthermore, rational accounts are often treated as a technicality or curiosity, with most prominent accounts of real-world political belief polarization assuming it emerges from procedurally-irrational thinking (Williams, 2023; Mandelbaum, 2019; Van Bavel, Rathje, Vlasceanu, & Pretus, 2024; Lord, Ross, & Lepper, 1979; Kahan, Peters, Dawson, & Slovic, 2017; Kahan, Hoffman, Braman, Evans, & Rachlinski, 2012). In this article I demonstrate, using simulations, that a plausible, empirically-validated, and relatively simple model of 'fully' Bayesian belief updating in political contexts can, under plausible conditions, give rise to polarization which bears several hallmarks of real-world political issue polarization. I draw on Young and de-Wit (2025)'s recent Bias-and-Expertise Model which proposes a Bayesian Network for how people can rationally infer and account for the biases of political information sources.

Before proceeding further I must disambiguate the sense of polarization to which I am referring, since the term refers to multiple distinct (but related) phenomena in cognitive science. One prominent sense of polarization, which Pallavicini, Hallsson, and Kappel (2021) term 'group escalation', is where the aggregate opinions of groups become amplified by engaging in discussion (e.g. Sunstein, 1999). This is not the sense of polarization I am concerned with. Rather, I am concerned with the sense sometimes known as 'bipolarization' where sub-groups within a superordinate group each update their beliefs in opposing directions in response to an identical stimulus or common treatment, such as a discussion (Mäs & Flache, 2013).

Understanding the drivers of bipolarization is important because it appears the American public's political views have undergone bipolarization in recent decades. For example, the opinions of Democrats and Republicans on key issues have grown further apart, (Webster & Abramowitz, 2017), become less overlapping (Lelkes, 2016), and have become more ideologically-constrained (Dimock, Doherty, Kiley, & Oates, 2014). With access to more recent data, Young et al. (2026) found that the gap between ideological groups in the US increased from 2008-2024, with the average position of liberal-leaning Americans shifting left and while the average position of conservative-leaning Americans has shifted right. Determining whether this bipolarization could be the product of rational reasoning, and if so what its mechanisms are, will influence whether and how any institution who wished to do so should try to depolarize Americans, or prevent further polarization occurring elsewhere.

Numerous previous models of rational bipolarization have pointed to source credibility perceptions as a driving force (Pallavicini et al., 2021; Olsson, 2013; Hahn & Harris, 2014; Cook & Lewandowsky, 2016; Young et al., 2025). Here, the fact that individuals' beliefs grow further apart is blamed on the different levels of trust they place in sources who make opposing claims. One limitation of this work however is that the models of source credibility utilised have overlooked arguably the most important component of credibility in politics, which is *bias*. The rational models of source characteristics used in this prior work assume that agents have a fixed probability of making true claims when discussing different hypotheses - but as Young and de-Wit (2025) have demonstrated, this is not true of political sources. Instead, political information sources usually have particular policies, ideologies, parties, or politicians that their speech is slanted towards supporting, either because they are motivated to behave this way, or possess unintentional biases in their information consumption or reasoning which lead them to form

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

'sincerely biased' beliefs (Wallace, Wegener, & Petty, 2020). This means that political information sources flip whether they will be reliable or unreliable depending on the alignment between the truth and their bias - the same source may be perfectly trustworthy when discussing congenial truths but completely untrustworthy when discussing uncongenial ones. Young and de-Wit (2025) empirically proved their Bias-and-Expertise Bayesian Network could capture the influence of bias in the context of political belief updating, and explain some phenomena which pre-existing Bayesian Network models of source cognition (Bovens & Hartmann, 2003; Harris, Hahn, Madsen, & Hsu, 2016) could not. Notably, the Bias-and-Expertise model disagrees with the models employed by Pallavicini et al. (2021) and Olsson (2013), in that these predict that agents can update in the *opposite* direction to the claims made by sources they distrust, whereas the Bias-and-Expertise model predicts that minimally-credible sources are at worst ignored. Overall, establishing whether and how agents who reason about source bias using the Bias-and-Expertise model undergo polarization is an important next step in the literature on rational polarization. I explore how the Bias-and-Expertise model goes give rise to polarization in the next section using simulations.

I note that Jern et al. (2014) previously discussed perceptions of source bias as a mediating variable in one of the numerous models of rational polarization they proposed. However, their account is specialised to explain the belief polarization reported by Lord et al. (1979), which often fails to replicate (e.g. Anglin, 2019). Moreover their account requires that individuals reason specifically about whether their sources are biased towards the consensus reached by experts on the topic in question, rather than their political bias. This account therefore only applies under specific conditions, and may not generalise to political debates more broadly. In contrast, the Bias-and-Expertise model offers a highly generalisable account of political polarization.

Method

I use simulations where agents attempt to learn about politics from competing information sources by using the Bias-and-Expertise Model to reason. In contrast to many other rational models of bipolarization (e.g., Pallavicini et al., 2021; Assaad & Hahn, 2025; Olsson, 2013; Mäs & Flache, 2013), agents do not communicate with each other. Instead, agents receive identical pieces of testimony and evidence, and employ the same reasoning model, yet still polarize when they start with some higher-order disagreement about source credibility or have some initial disagreement about the ground truth.

Exploring whether individuals can bipolarize even when exposed to identical information is important because some accounts of polarization blame selective information exposure (e.g., right-wingers predominantly consuming right-wing media and left-wingers left-wing media) due to echo chambers, filter bubbles, and algorithmic social media feeds (Flaxman, Goel, & Rao, 2016; Madsen, Bailey, & Pilditch, 2018), with the implication that citizens would be less polarized if asymmetries in media consumption were removed. If however individuals can polarize even when they consume identical information, this would suggest balancing out the ideological slant of consumed media would be insufficient to prevent polarization.

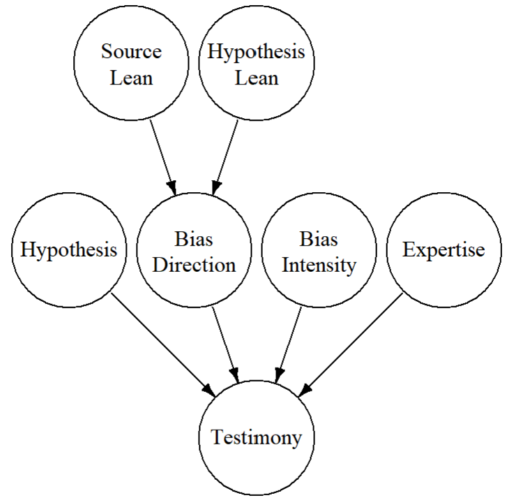

Figure 1: Directed Acylic Graph (DAG) of the Bias-and-Expertise Model's Bayes Net

Furthermore exploring specifically how polarization can occur when agents receive most of their information in the form of source testimony is important because this is how people learn most of the information they receive in many domains (Hahn & Harris, 2014; Henrich, 2016).

I begin with dyadic simulations showing that a pair of agents will bipolarize under certain conditions. I then scale these up to group-level simulations and add in some stochasiticity. I conclude with noting some additional characteristics of the polarization that emerges.

Model

For full details about the Bias-and-Expertise model, see the original paper (Young & de-Wit, 2025). The model employs a Bayesian Network (Pearl, 1988); Bayesian Networks are reasoning tools that can perform inductive inference by applying Bayes' Rule and are therefore well-suited for normative models of learning. I will offer a brief overview of the relevant aspects of it here.

As the diagram in Figure 1 shows, the model supposes that a source's testimony about a hypothesis (i.e., what they claim to be true about the world on a particular topic) is the product of a four-way interaction between the ground truth (Hypothesis: True vs False), the source's Bias Direction (For vs Against the hypothesis), the source's Bias Intensity (Biased vs Impartial), and the source's Expertise (Low vs High). The Bias Direction is in turn determined by an interaction between the 'Leaning' of the Source and the Hypothesis on a common axis of political differentiation, such as left-right - the Bias

Direction is *For* when the leanings match and *Against* when they conflict.

The four-way interaction to determine the source's testimony follows three rules: (1) If Bias Intensity is high (Bias Intensity = *Biased*), the source will just follow the direction of their bias, by claiming the Hypothesis is true if their Bias Direction is *For*, but false if their Bias Direction is *Against*. (2) If Bias Intensity is low (Bias Intensity = *Impartial*) and Expertise is *High*, they will follow the ground truth, by claiming the Hypothesis is true if it is true, and false if it is false. (3) If Bias Intensity is low (*Impartial*) and Expertise is *Low* as well, the source will guess, with a 50:50 chance of saying it is true or false.

The Inference-Influence Spiral

Upon observing a source's testimony, a hearer can apply Bayes' Rule to the network to infer, probabilistically, the likely state of the ground truth, considering their prior beliefs about the source and the ground truth. Figure 2A shows how beliefs about the ground truth are updated as a function of the other variables. Simultaneously, a hearer can probabilistically infer the source's characteristics, taking into account their priors for those characteristics and the ground truth. Figure 2B shows how the source's Bias Intensity, as an example of one source characteristic, is updated as a function of the other variables.

These characteristics of the Bias-and-Expertise Model mean that agents who employ it can bipolarize via *inference-influence* spirals.

Inference: Beliefs about the ground truth affect what we *infer* about sources' credibility when we hear their testimony. Consider what happens when a person hears conflicting testimony about the same hypothesis from two sources who both make claims in line with their Bias Direction (one biased *For* who says it is true, one biased *Against* who says it is false). We can see from Figure 2B that both sources' perceived Bias Intensity is likely to increase, but the amount of increase will be greater for the source whose claim seems less plausible in light of the source's prior (the *For* source if p(Hyp) <0.5, the *Against* source if p(Hyp) >0). Therefore if two people with opposing priors heard the same pair of claims (one p(Hyp) >0.5, one p(Hyp) <0.5), their source credibility perceptions would diverge, as each would update a different source's Bias Intensity by a greater amount. Thus, exposure to conflicting testimony can cause source credibility perceptions to diverge.

Influence: Beliefs about source credibility *influence* what we believe when we hear source testimony. If a person hears two sources give conflicting testimony about a hypothesis as before, they will update their belief in the hypothesis in the direction of the source they deem less biased. This is because, as we see from Figure 2A, they should update less in the direction of each claim in proportion to the source's Bias Intensity. Therefore if two people disagreed about which source was more biased, they would update in opposing direction. Thus, exposure to conflicting testimony can cause beliefs about the ground truth to polarize.

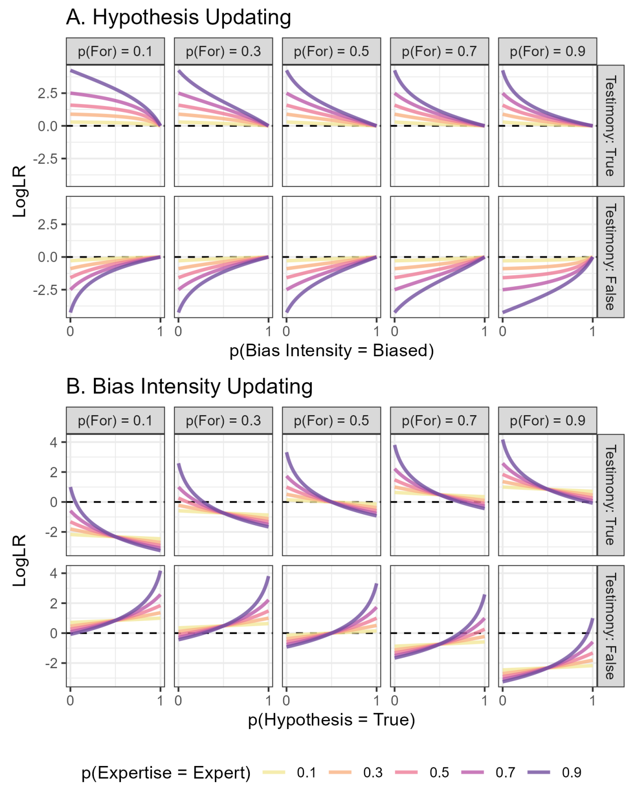

Figure 2: Panel A: Agents update their belief in the hypothesis in the direction of the source's claim, and more so the less they are thought to be biased towards making it. Panel B: Sources who make claims consistent with the presumed direction of their bias suffer are thought to have a higher bias intensity when that claim seems implausible in light of the agent's prior

Putting these dynamics together, we get polarization: if they initially disagree on one issue, this will cause their credibility perceptions of sources who provide competing testimony about it to diverge, meaning if they hear these same sources provide further competing testimony about other issues, their views on those issues will polarize, making them more polarized overall. Similarly, if they hear other pairs of sources provide competing testimony about the same set of hypotheses, their credibility perceptions of them will diverge too, and so on. Thus a minor initial disagreement can spiral into broader polarization about many other hypotheses and sources, driven by normative reasoning about bias and exposure to conflicting testimony along a consistent axis (that is, the hypotheses and sources implicated all have leanings along a common axis). Inference (Beliefs about ground truth → Beliefs about sources) and influence (Beliefs about sources → Beliefs about ground truth) therefore create a kind of feedback loop, which because it must 'move' over different pairs of sources and issues in order to function, I term a spiral.

I now run simulations to show these polarizing mechanisms in action. Code for all simulations is written in *R* and available for download via the OSF: `https://osf.io/ugbts/overview?view_only=e3410a9994b4443f8db3c4b468cbbf02`.

Simulations

Dyadic Simulations

I begin with dyadic simulations involving only two agents. One agent is nominally 'right-wing' and the other 'left-wing'. The left-right axis is used only as a familiar example - this is just an arbitrary label. Both agents hear 32 claims - one from each of four sources about each of eight hypotheses. These sources mirror the kinds of media sources most citizens are reliant upon to learn about politics but may themselves be partisan, like politicians, media institutions, and expert commentators. All hypotheses have a right-wing leaning. Two of the sources have a right-wing leaning, and say all hypotheses are true; two have a left-wing leaning and say all the hypotheses are false. Agents also receive one piece of direct evidence about each hypothesis. Agents use the Bias-and-Expertise Network to reason about each source and hypothesis. The sources, hypotheses, and evidence form one connected network, meaning the order in which the testimony and evidence are presented (which is random) does not affect any of the agents' end-state beliefs.

Across simulations I vary three parameters: (1) The initial level of disagreement between right-wing and left-wing agent, determined by the gap between the right-wing agent's priors for all hypotheses (H_R) and the left-wing agent's priors for all hypotheses (H_L), which are set to, respectively, 0.5 and 0.5, 0.6 and 0.4, or 0.75 and 0.25 (the Source and Hypothesis Leaning variables are set so that 0 = left-wing and 1 = right-wing). (2) The strength and direction of the direct evidence, which I vary by setting the likelihood ratio for each piece of evidence to be either very strong, moderate, or weak evidence for the left (1/8, 1/4, 1/2 respectively), very strong, moderate, or weak evidence for the right (8/1, 4/1, 2/1 respectively) or neutral (1/1); each piece of evidence has the same strength in a given simulation. (3) The gap between the Bias Intensity agents initially ascribe to the source on the same side as them (B_{Own}) and the source on the opposing side (B_{Other}), which is manipulated by setting these priors to either 0.5 and 0.5, 0.4 and 0.6, or 0.2 and 0.8. This gives $3 \times 7 \times 3 = 63$ simulations. The remaining priors are set at 0.8 for the Source Leanings of the right-wing sources and 0.2 for the left-wing sources, 0.6 for all sources' Expertise, and 0.8 for all Hypothesis Leanings.

The results are shown in Figure 3, giving the mean belief across all eight hypotheses for each source at the beginning and end of the simulation. As it shows, polarization (yellow facets) can occur when there is some initial disagreement between the agents as to which side is correct on the hypotheses or which side's sources are less biased, and the direct evidence is weak or neutral. When the agents do not differ at all

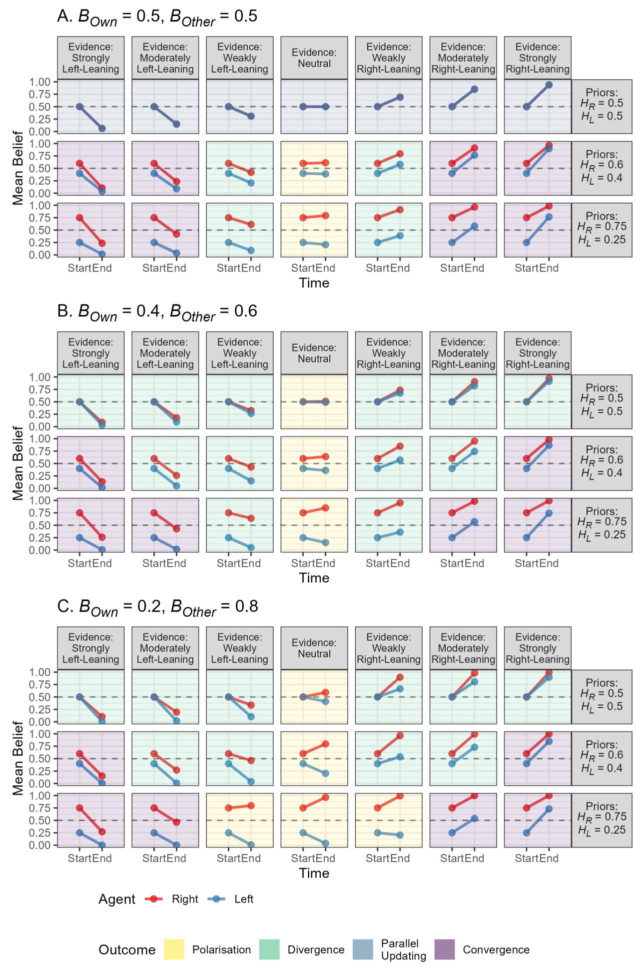

Figure 3: Outcomes of dyadic simulations

to begin with ($B_{Own} = B_{Other} = 0.5$ & $H_R = H_L = 0.5$, top row of panel A), they update in parallel (blue facets). Polarisation can even occur when the direct evidence favours one of the sides, if it is weak and the initial differences between agents are strong (central facets in bottom row of panel C). Otherwise, agents either diverge (green panels), by moving in the same direction but growing further apart, or converge (purple panels), moving closer together.

Group Simulations

Scaling up the dyadic simulations, I present group simulations to demonstrate the phenomenon of bipolarization within groups of n=200 agents. These simulations have the same set-up as the dyadic simulations, but to introduce variance among the agents, I introduce some stochasticity.

As before, all agents begin with priors of 0.8 for the Hypothesis Leanings, and 0.8 for the Source Leaning of the right-wing sources and 0.2 for the left-wing sources. But the remaining priors are determined by a d parameter, which is sampled for each agent i from a normal distribution: $d_i \sim$

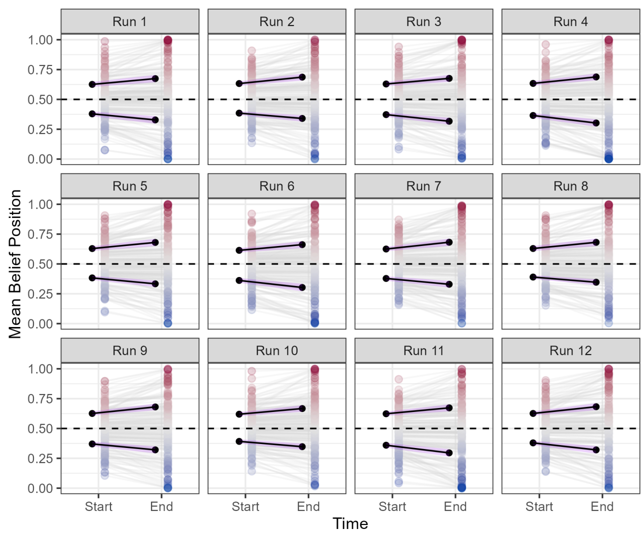

Figure 4: Group Simulations with neutral evidence lead to polarization.

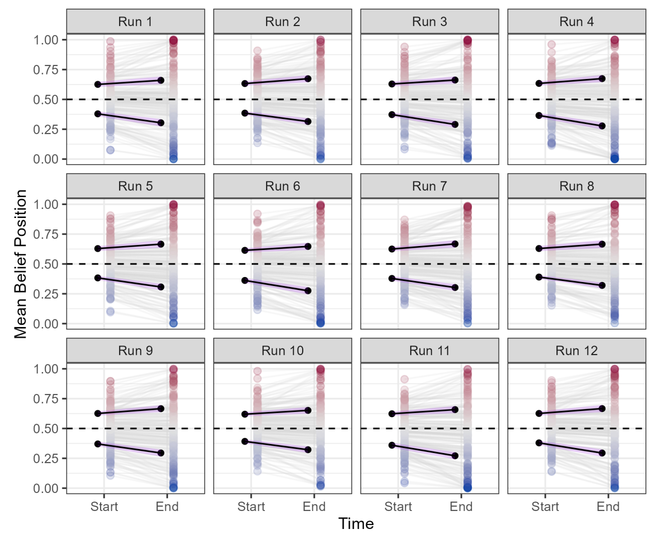

Figure 5: Group Simulations with very weak left-leaning evidence lead to asymmetrical polarization.

$\mathcal{N}(0, 0.15^2)$ with values trimmed to range from -0.49 to +0.49. This d parameter determines how far the agents' priors for these remaining variables differ from 0.5, as well as the direction - positive values indicate right-wing priors and negative values left-wing priors. To obtain the Hypothesis priors for each agent, I sample eight values from a normal distribution with mean d_i and variance 0.15^2, then add 0.5. To obtain the Expertise and Bias Intensity priors, I generate two normally-distributed variables of four values with mean d_i and variance 0.15^2, correlated at $r = 0.5$, where each pair of values corresponds to the Expertise and Bias Intensity of one source. To obtain the priors, I transform the values using the following functions: **Expertise**: *For right-wing sources:* $f(x) = 0.5 + x$ *and for left-wing sources:* $f(x) = 0.5 - x$. **Bias Intensity**: *For right-wing sources:* $f(x) = 0.5 - x$ *and for left-wing sources:* $f(x) = 0.5 + x$. Thus agents with positive d parameters see right-wing sources as having higher Expertise and lower Bias Intensity than left-wing sources, and vice versa. As before, agents also receive pieces of direct observations for each hypothesis. To begin with I set these at $\frac{p(e|H)}{p(e|\bar{H})} = 1$.

I perform twelve runs of the simulation, each with a new random initialization of the d_i parameters. To track polarization, I find the initially 'left-wing' agents with mean Hypothesis priors below 0.5, and the initially 'right-wing' agents with mean Hypothesis priors above 0.5, and calculate their mean Hypothesis priors and posteriors. Polarization result occurs when both groups move in opposing directions with statistical significance (p <.05).

The results are plotted in Figure 4, showing polarization in every run. Collapsing across the twelve runs, there was clear polarisation, with the initially left-wing agents ($\mu_{Prior} = 0.376\,[0.371, 0.382]$) losing confidence in the hypotheses ($\mu_{Posterior} = 0.324\,[0.315, 0.332]$, $t_{Paired}(1193)$=27.161, p<.001) but the initially right-wing agents ($\mu_{Prior} = 0.626\,[0.621, 0.631]$) becoming more confident ($\mu_{Posterior} = 0.678\,[0.669, 0.686]$, $t_{Paired}(1205)$=26.777, p<.001).

Polarization Characteristics

The polarization produced in these Group Simulations can resemble some characteristics of real-world political polarization.

Asymmetry The first characteristic is asymmetry - Young et al. (2026) find left- and right-leaning Americans have grown further apart at different rates, with left-leaning Americans shifting further to the left from 1988 to 2024 than right-leaning Americans shifted to the right. This resembles the kind of asymmetrical polarization seen in Figure 3C's bottom row when there is weak evidence to the left or right. One way the Bias-and-Expertise Model, then, can give rise to this kind of asymmetrical polarization is if the direct evidence people can access, i.e., that which isn't conveyed by partisan sources, itself has a weak leaning. One factor that could explain a greater exposure to liberal influences on Americans' political opinions could be the decline of Americans' religiosity from the '90s onwards (Schwadel, 2013), since arguments against liberal positions on some of the issues Americans have polarized over, like abortion, are often grounded in religion.

In any case, I repeat the Group Simulations above, but alter the direct evidence to extremely weakly favour the left, with each piece of evidence having the likelihood ratio $\frac{0.475}{0.525} \approx 0.90$, therefore weakly implying each (right-leaning) hypothesis is false. Note that this is equivalent to a possibly more realistic case of agents receiving a large amount of mixed evidence which very weakly favours for the left in the aggregate, rather than one single piece of very weak evidence. For direct comparability, I run twelve runs with the same initializations of the d_i parameters as before.

Collapsing across the twelve runs, there was asymmetrical polarisation, with the initially left-wing agents ($\mu_{Prior} = 0.376\,[0.371, 0.382]$) losing confidence in the hypotheses ($\mu_{Posterior} = 0.299\,[0.291, 0.307]$, $t_{Paired}(1193)$=41.857, p<.001) but the initially right-wing agents ($\mu_{Prior} = 0.626\,[0.621, 0.631]$) becoming more confident, though with a smaller magnitude of change ($\mu_{Posterior} = 0.653\,[0.644, 0.662]$, $t_{Paired}(1205)$=13.366, p<.001).

Increasing Constraint DellaPosta (2020) has recently shown that Americans' political opinions have become more ideologically constrained since the mid-20th Century, with citizens increasingly occupying similar ideological positions across different issues. This constraint also emerges in the Group Simulations above, as agents' positions across different issues become more positively correlated. I reanalyze the two preceding simulations, calculating the correlation between agents' positions on each unique pair of hypotheses, then finding the mean. Collapsing across the twelve runs of the simulation with neutral evidence, I find this correlation increases from 0.501 [0.496, 0.506] at the start to 0.758 [0.755, 0.761] at the end. This also occurs for the simulations with very weak evidence to the left, increasing from the same starting point to 0.759 [0.755, 0.762].

Discussion

The results show that Bayesian agents reasoning in accordance with the Bias-and-Expertise Model can polarize even if exposed to identical information when the truth is disputed by sources from opposing political 'sides', in a manner which features some properties of real-world political polarization (like Asymmetry and Increasing Constraint). This suggests that even if we could prevent echo chambers and irrational reasoning among democratic citizens, polarization can still occur. It is surely reasonable for citizens to worry about source bias, but doing so creates feedback loops, via Inference-Influence cycles, which are polarizing.

This work adds to the evidence for rational polarization (Pallavicini et al., 2021; Hahn & Harris, 2014; Cook & Lewandowsky, 2016; Bullock, 2009; Perfors & Navarro, 2019; Assaad & Hahn, 2025; Jern et al., 2014; Olsson, 2013). Furthermore, it adds to the evidence that clear disambiguating evidence can prevent polarization (Perfors & Navarro, 2019), meaning interventions to reduce polarization could focus on improving citizens' ability to access direct evidence about political realities. But several aspects of the present work progress knowledge of rational polarization beyond previous studies. For one, no previous simulation study has used cognitive models which integrate perceptions of source *bias*, despite their importance as a source characteristic in politics (Young & de-Wit, 2025). We now know that perceptions of bias and expertise can mediate bipolarization, despite the Bias-and-Expertise Model not permitting reverse updating (i.e., taking a source's claim that a hypothesis is true as evidence it is false) as assumed in some previous studies (e.g. Pallavicini et al., 2021; Olsson, 2013). Second, this work makes explicit the Inference-Influence spiral and how it leads to polarization. Third, many previous simulations allow, and indeed depend upon, agents communicating with each other (e.g. Pallavicini et al., 2021; Assaad & Hahn, 2025; Perfors & Navarro, 2019; Olsson, 2013). While citizens do communicate, a large fraction of political communication is one-way, with citizens hearing from media institutions, politicians, and expert commentators without being able to meaningfully 'talk back' to them. That polarization can occur when communication is one-way rather than bidirectional is therefore an important result. Fourth, this work shows how the same model can predict, in a unified way, not only polarization but asymmetrical polarization and increasing issue constraint.

One limitation of these results is that polarization requires some initial disagreement between agents about the ground truth or the credibility of sources, but I do not specify where these differences come from. However, it seems inevitable that there will always be 'natural' axes of disagreement within societies due to differences in culture, religion, socioeconomic status, age, ethnicity, gender, sexuality, wealth, occupation, education and so on. These simulations show that when competing sources align along these axes, their conflicting testimony can be polarizing. However, it would be useful to explore situations with multiple axes of disagreement and bias, since political ideologies are multidimensional (Feldman & Johnston, 2014).

The simulations elide numerous aspects of real-world politics, like communication between agents, relationships between hypotheses, and heterogeneity between agents in terms of the testimony and evidence they encounter. Moreover, US polarization has occurred over generations, with a changing population of millions of agents and sources. But the virtue of simple simulations is that they make the mechanism clear: trying to infer and account for sources' bias when they provide conflicting testimony about issues that citizens already disagree about causes polarization. It therefore allows us to identify a simple and plausible 'recipe' for polarization. Complicating the simulations risks undermining clean inference when we already know that, for example, allowing agents to communicate can be polarizing too, depending on the specific rules we introduce for how communication happens (Assaad & Hahn, 2025). While it would be useful in future to perform larger, more complex simulations which include many potential drivers of polarization so that their relative importance and interactions can be assessed, simple, focussed simulations are useful as building blocks in the meantime. Indeed, these simulations demonstrate that bias perceptions may explain some portion of real-world political polarization.

Acknowledgments

Many thanks to three anonymous reviewers who all provided helpful and insightful comments, and to Lee de-Wit for guidance while completing this work.

While working on this manuscript I have been funded by an ESRC DTP Studentship (`https://doi.org/10.17863/CAM.106767`), a postdoc position funded by an award from the Templeton World Charity Foundation to Lee de-Wit and Jens Madsen (`https://doi.org/10.54224/31453`), and finally a second postdoc position funded by the Horizon Europe project as part of the Perycles project.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

The author acknowledges support by the European Union under the Horizon Europe project Perycles (Participatory Democracy that Scales).

Funded by
the European Union

References

Anglin, S. M. (2019, May). Do beliefs yield to evidence? Examining belief perseverance vs. change in response to congruent empirical findings. *Journal of Experimental Social Psychology*, *82*, 176–199. Retrieved 2023-03-15, from `https://linkinghub.elsevier.com/retrieve/pii/S0022103118304529` doi: 10.1016/j.jesp.2019.02.004

Assaad, L., & Hahn, U. (2025). Rational Polarization: Sharing Only One's Best Evidence Can Lead to Group Polarization. *Proceedings of the Cognitive Science Society*, *46*.

Benoît, J., & Dubra, J. (2019). Apparent Bias: What Does Attitude Polarization Show? *International Economic Review*, *60*(4), 1675–1703. doi: 10.1111/iere.12400

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology*. Oxford University Press.

Bullock, J. G. (2009). Partisan Bias and the Bayesian Ideal in the Study of Public Opinion. *The Journal of Politics*, *71*(3), 1109–1124. doi: 10.1017/S0022381609090914

Cook, J., & Lewandowsky, S. (2016). Rational Irrationality: Modeling Climate Change Belief Polarization Using Bayesian Networks. *Topics in Cognitive Science*, *8*(1), 160–179. doi: 10.1111/tops.12186

DellaPosta, D. (2020). Pluralistic Collapse: The "Oil Spill" Model of Mass Opinion Polarization. *American Sociological Review*, *85*(3), 507–536. doi: 10.1177/0003122420922989

Dimock, M., Doherty, C., Kiley, J., & Oates, R. (2014). *Political Polarization in the American Public: How Increasing Ideological Uniformity and Partisan Antipathy Affect Politics, Compromise, and Everyday Life* (Tech. Rep.). Pew Research Center.

Feldman, S., & Johnston, C. (2014). Understanding the Determinants of Political Ideology: Implications of Structural Complexity: Understanding Political Ideology. *Political Psychology*, *35*(3), 337–358. doi: 10.1111/pops.12055

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, *80*(S1), 298–320. doi: 10.1093/poq/nfw006

Hahn, U., & Harris, A. J. (2014). What Does It Mean to be Biased. In *Psychology of Learning and Motivation* (Vol. 61, pp. 41–102). Elsevier. doi: 10.1016/B978-0-12-800283-4.00002-2

Harris, A. J. L., Hahn, U., Madsen, J. K., & Hsu, A. S. (2016). The Appeal to Expert Opinion: Quantitative Support for a Bayesian Network Approach. *Cognitive Science*, *40*(6), 1496–1533. doi: 10.1111/cogs.12276

Henrich, J. P. (2016). *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. Princeton: Princeton University Press. (OCLC: ocn908084074)

Jern, A., Chang, K.-m. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review*, *121*(2), 206–224. doi: 10.1037/a0035941

Kahan, D. M., Hoffman, D. A., Braman, D., Evans, D., & Rachlinski, J. J. (2012). "They Saw a Protest": Cognitive Illiberalism and the Speech-Conduct Distinction. *Stanford Law Review*, *64*, 57.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54–86. doi: 10.1017/bpp.2016.2

Lelkes, Y. (2016). Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly*, *80*(S1), 392–410. doi: 10.1093/poq/nfw005

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109.

Madsen, J. K., Bailey, R. M., & Pilditch, T. D. (2018). Large networks of rational agents form persistent echo chambers. *Scientific Reports*, *8*(1), 12391. doi: 10.1038/s41598-018-25558-7

Mandelbaum, E. (2019). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind & Language*, *34*(2), 141–157. doi: 10.1111/mila.12205

Mäs, M., & Flache, A. (2013). Differentiation without Distancing. Explaining Bi-Polarization of Opinions without Negative Influence. *PLoS ONE*, *8*(11), e74516. doi: 10.1371/journal.pone.0074516

O'Connor, C., & Weatherall, J. O. (2018). Scientific Polarization. *European Journal for Philosophy of Science*, *8*(3), 855–875. (arXiv: 1712.04561) doi: 10.1007/s13194-018-0213-9

Olsson, E. J. (2013). A Bayesian Simulation Model of Group Deliberation and Polarization. In F. Zenker (Ed.), *Bayesian Argumentation* (Vol. 362, pp. 113–133). Dordrecht: Springer Netherlands. (Series Title: Synthese Library) doi: 10.1007/978-94-007-5357-0_6

Olsson, E. J., & Vallinder, A. (2013). Norms of assertion and communication in social networks. *Synthese*, *190*(13), 2557–2571. doi: 10.1007/s11229-013-0313-1

Pallavicini, J., Hallsson, B., & Kappel, K. (2021). Polarization in groups of Bayesian agents. *Synthese*, *198*(1), 1–55. doi: 10.1007/s11229-018-01978-w

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Francisco (CA), US: Morgan Kaufman. doi: 10.1016/B978-0-08-051489-5.50002-3

Perfors, A., & Navarro, D. J. (2019). Why do echo chambers form? The role of trust, population heterogeneity, and objective truth. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *41*.

Schwadel, P. (2013). Changes in Americans' Strength of Religious Affiliation, 1974-2010. *Sociology of Religion*, *74*(1), 107–128. doi: 10.1093/socrel/srs043

Singer, D. J., Bramson, A., Grim, P., Holman, B., Jung, J., Kovaka, K., . . . Berger, W. J. (2019). Rational social and political polarization. *Philosophical Studies*, *176*(9), 2243–2267. doi: 10.1007/s11098-018-1124-5

Sunstein, C. R. (1999). *The Law of Group Polarization* [John M. Olin Law & Economic Working Paper No. 91 (2D Series)].

Van Bavel, J. J., Rathje, S., Vlasceanu, M., & Pretus, C. (2024). Updating the identity-based model of belief: From false belief to the spread of misinformation. *Current Opinion in Psychology*, *56*, 101787. doi: 10.1016/j.copsyc.2023.101787

Van Doorn, M. (2024). How Partisanship Can Moderate the Influence of Communicated Information on the Beliefs of Agents Aiming to Form True Beliefs. *Social Epistemology*, 1–16. doi: 10.1080/02691728.2024.2362680

Wallace, L. E., Wegener, D. T., & Petty, R. E. (2020). Influences of Source Bias That Differ From Source Untrustworthiness: When Flip-Flopping Is More and Less Surprising. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, *118*(4).

Webster, S. W., & Abramowitz, A. I. (2017). The Ideological Foundations of Affective Polarization in the U.S. Electorate. *American Politics Research*, *45*(4), 621–647. doi: 10.1177/1532673X17703132

Williams, D. (2023). The case for partisan motivated reasoning. *Synthese*, *202*(3), 89. doi: 10.1007/s11229-023-04223-1

Young, D. J., Ackland, J., Kapounek, A., Madsen, J. K., Greening, L., & de Wit, L. H. (2026). A new measure of issue polarization using k-means clustering: US trends 1988–2024 and predictors of polarization across the world. *Royal Society Open Science*, *13*(2). doi: https://doi.org/10.1098/rsos.251428

Young, D. J., & de-Wit, L. H. (2025). The Bias-and-Expertise Model: A Bayesian Network Model of Political Source Characteristics. *Cognitive Science*, *49*(11), e70141. doi: 10.1111/cogs.70141

Young, D. J., Madsen, J. K., & de Wit, L. H. (2025). Belief polarization can be caused by disagreements over source independence: Computational modelling, experimental evidence, and applicability to real-world politics. *Cognition*, *259*, 106126. doi: 10.1016/j.cognition.2025.106126