

RUHR
ECONOMIC PAPERS

The Polarization of Political Beliefs: Narratives, Populist Influence and Social Reinforcement

Matthias Reccius and Michael Roos

RUHR UNIVERSITÄT BOCHUM RUB RWI #1234

Imprint

Ruhr Economic Papers

Ruhr Economic Paper #1234 "The Polarization of Political Beliefs: Narratives, Populist Influence and Social Reinforcement"

Responsible Editor: Thomas Bauer

Jointly published by

RWI – Leibniz-Institut für Wirtschaftsforschung e.V.
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Bergische Universität Wuppertal, Schumpeter School of Business and Economics
Gaußstraße 20, 42119 Wuppertal, Germany

Series Coordination:
RWI Press Office, presse@rwi-essen.de, phone +49 (0) 201 8149-213
RWI – Leibniz Institute for Economic Research, Hohenzollernstr. 1–3, 45128 Essen, Germany
www.rwi-essen.de

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.
The Institute has the legal form of a registered association; Vereinsregister, Amtsgericht Essen VR 1784

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors and institutions.

All rights reserved. Essen, Germany, 2026

ISSN 1864-4872 (online)
ISBN 978-3-96973-420-9
DOI https://dx.doi.org/10.4419/96973420

The polarization of political beliefs: narratives, populist influence and social reinforcement *

Matthias Reccius [a] † Michael Roos[a] ‡

November 12, 2026

[a] Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany.

Abstract

We develop an agent-based model of attitude and belief polarization driven by competing economic policy narratives. Citizens form attitudes and update beliefs about policy outcomes through motivated reasoning, while two mainstream political initiators and, in a second phase, a right-wing populist entrant disseminate narratives across a network structured by homophily. Attitude polarization emerges through a phase transition as discounting of others' positions intensifies, with citizens clustering around initiator positions beyond a critical threshold. Populist entry is assumed to impose a one-sided loss frame. It pulls the attitude distribution toward its position, with the resulting shift in polarization depending on discounting intensity and network homophily. Three counter-narrative strategies, pandering, rebalancing, and opportunity framing, are evaluated for their capacity to offset this influence. The model offers a theoretical account of how narrative transmission, motivated reasoning, and elite communication jointly shape polarized public opinion, with implications extending beyond immigration to any policy domain involving an ex ante ambiguous net effect.

Keywords: *narratives*, *policies*, *motivated reasoning*, *networks*, *attitudes*, *values*, *populism*

JEL Classification: D72 – Political Processes: Rent-seeking, Lobbying, Elections, Legislatures, and Voting Behavior D83 – Search; Learning; Information and Knowledge; Communication; Belief; Unawareness, D85 – Network Formation and Analysis: Theory, A13 – Relation of Economics to Social Values, Z13 – Economic Sociology; Economic Anthropology; Language; Social and Economic Stratification

*We thank Ulrich Fritsche and Michelle Alfers for very helpful comments on the draft. We also thank our colleagues from the Chair of Macroeconomics for countless fruitful discussions over coffee and lunch at various stages of the project. Further, we thank the participants of the 2024 International Workshop on Economic Complexity and Macroeconomic Dynamics in Valencia, Spain, and of the 2024 Social Simulation Conference in Krakow, Poland. The work is part of the Narrative Economics Alliance Ruhr (NEAR) project, supported by the Mercator Research Center Ruhr (MERCUR) under project number Ko-2022-0015.

†Corresponding author. Email: matthias.reccius@ruhr-uni-bochum.de

‡Email: michael.roos@ruhr-uni-bochum.de

1 Introduction

In politics, narratives are a key medium through which political actors frame societal issues and advance corresponding policy solutions (Roe, 1994). Research in political science and economics suggests that narratives have the power to influence political attitudes and persuade individuals to shift their beliefs (de Clippel & Zhang, 2022; Lau et al., 1991). At the social level, narratives can also "go viral" and shape collective understandings of political reality by getting shared by individuals in their respective social networks (Roos & Reccius, 2023). While a growing experimental literature examines how individuals update beliefs in response to politically charged messages (Alesina et al., 2018; Cattaneo & Grieco, 2021; Haaland et al., 2023), the broader dynamics of diffusion and reinforcement through social interaction remain less understood. As narratives circulate within social networks, they can gradually align opinions within groups and accentuate differences between them, contributing to political polarization.

Such dynamics are particularly relevant in contemporary political environments. In recent years, right-wing populist actors have achieved and maintained political influence across major democracies, with much of their prominence getting attributed to the strategic use of rhetoric (Bonikowski & Gidron, 2015). This development is troubling, as populist governance has been linked to the erosion of democratic institutions (Galston, 2018; Ruth-Lovell & Grahn, 2022) and economic underperformance (Dornbusch & Edwards, 1991; Funke et al., 2023).

This paper contributes to a better theoretical understanding of attitude and belief polarization through policy narratives and the effects of different communication strategies used by politicians. We propose an agent-based model (ABM) in which agents form attitudes towards a proposed economic policy. These attitudes depend on individuals' beliefs about the policy's positive and negative outcomes and their normative evaluations of these consequences. Private agents communicate their beliefs and attitudes through narratives and revise them in light of information received from others. Simultaneously, two mainstream and one far-right populist politician attempt to shape public attitudes toward the policy. We use the model to answer the following research questions: (1) How do narrative transmission, motivated reasoning, and elite communication interact to produce polarized attitudes and beliefs about economic policy? (2) How does polarization depend on homophily in individuals' discussion networks? (3) Which narrative strategies are effective in counteracting the polarizing influence of populist communication? Our theoretical exploration suggests homophily in the network of narrative interactions and discounting of others' attitudes as important factors.

A classic policy field on which populists influence moderate politicians is immigration

policy. In response to the pressure exerted by populists, moderate politicians limit immigration, even though immigration has been associated with a range of positive economic outcomes. For the sake of expositional clarity, we focus on the example of immigration policy, but our model is more general and can also be applied to other policy domains. The key feature needed for the model is that a policy involves trade-offs, which is likely to be true for most policies.

Traditionally, two forms of political polarization have been at the center of research: ideological polarization, which is the distance between or the drifting apart of political attitudes and preferences, and affective polarization, which is the degree or increase of aversion between political opponents. Based on Lee et al. (2021), Rekker (2022) proposes to add *factual belief polarization* as a third type of political polarization. He argues that this kind of polarization is especially problematic, because it undermines the functioning of democracy. Democratic debate and compromise between political parties requires at least a basic agreement on facts (Haas, 2018). Furthermore, polarization over facts threatens the modern view of policymaking informed by experts and scientific knowledge. The rise of populism in many Western countries might have contributed to a relativization of the importance of scientific knowledge in favor of biased beliefs (see Rekker (2025)) or even conspiracy theories as seen in the case of COVID-19 (Stecula & Pickup, 2021). Empirical evidence supports the relevance of belief polarization about inequality, climate change and immigration in the U.S. and the Netherlands (Rekker, 2024; Rekker & Harteveld, 2022), but according to these authors important questions remain unanswered. While a connection between factual beliefs and policy opinions has been found for fields like redistribution (Alesina & Giuliano, 2011), the direction of causality between ideological polarization and belief polarization is unclear. While it appears plausible that beliefs determine a person's issue attitude, the reverse could also be true as motivated reasoning may lead factual beliefs to follow ideological preferences (Epley & Gilovich, 2016; Kunda, 1990).

Political scientists have struggled to formulate a definition of populism that applies across historical periods and geographical contexts (Mudde & Rovira Kaltwasser, 2017). Nevertheless, several recognized characteristics have contributed to a widely shared minimal definition: one key element is that populist actors rely on an "us-against-them" logic that opposes the "true people" to a supposedly corrupt elite (Mudde & Rovira Kaltwasser, 2012; Müller, 2016). Exclusionary populists, who dominate contemporary political landscapes, expand this core antagonism by introducing a nativist dimension (Obradović et al., 2020). They thus frame the "true people" along two axes: the people versus the elites, and natives versus immigrants. Another element is that populists have a particular communication style (Bracciale & Martella, 2017; Cassell, 2023; Jagers & Walgrave, 2007). They often focus on problems and potential negative outcomes of policies proposed by incumbents as this allows them to take an anti-elitist stance. Using a worldwide data set, Nai (2021) shows that populists use a more negative

communication style than non-populists, which include policy attacks claiming that specific policies are bad or harmful, negative tone and fear-arousing messages. Klinger et al. (2022) find that between 2014 and 2019 negative emotion, negative campaigning and dramatization increased in 12 countries during European Parliament election campaigns on Facebook. In particular, Conservative parties' behavior drove this overall shift, whereas Ecologist parties mainly shifted toward more dramatization. Nationalist parties and to some extent Left Socialist parties used the populist communication style already in 2014 and kept it.

We hence model a populist as a political communicator that does not weigh trade-offs as moderate political parties do, but exclusively talks about the undesired effects of a policy proposed or implemented by moderate incumbents. Furthermore, the populist successfully moralizes the discourse and triggers fear of loss, which changes the trust citizens place in the moderate parties. The evaluations of negative outcomes of the policy become highly salient and dominate the discourse. Take immigration policy as an example. Normally, both left-centrist and right-centrist parties acknowledge that immigration often has effects on both the labor market and on social security systems. A proposed policy might aim at reducing labor market shortages by allowing more migrants to enter the country. This would be the intended or desired effect of the policy. As an undesired side-effect, some migrants may become eligible to social security benefits, such that the policy involves a trade-off between labor market benefits and social security costs. Moderate parties have different political positions about the relative importance of the two effects as well as their (un-)desirability. Left (or liberal) parties tend to emphasize the labor market benefits, whereas right (or conservative) parties focus more on the potential social security costs. What characterizes them as moderates is that neither side denies that there are both costs and benefits and that they deliberate on the policy by evaluating both. In contrast, the populist does not acknowledge that the policy will have benefits, but talks only about the costs. In addition, the costs are highlighted as highly undesirable and dramatized. This allows the populist to use "us-against-them" rhetoric against immigrants, who are causing the costs, and against the moderate parties (or incumbent elites), who obfuscate the truth by their convoluted argumentation and are responsible for welfare losses by wrong policies.

Our model serves to illuminate the mechanisms by which the populist communication style affects both the attitudes and beliefs of citizens. Attitudes (or opinions, which we treat as a synonymous term) are summary evaluations of an object, which can be a policy as in our case. They both influence behavior, e.g. voting, and information processing, in particular the perception, selection and evaluation of information (Wänke & Vogel, 2016). While the influence on behavior is of obvious relevance for political economy, attitudes' impact on information processing is also important and is related to the concept of *motivated reasoning* (Barron et al., 2025; Epley & Gilovich, 2016; Kunda, 1990). Motivated reasoning means that preferences or normative considerations influence

beliefs. It is particularly relevant in political settings, where factual beliefs and values are tightly interlinked. The term *values*, in our model, is used in a narrow and specific sense, denotating preferences over a policy's potential outcome space. The connection between beliefs and values is not straight-forward to measure: While information provisional experiments have been devised to study them (Haaland et al., 2023), there is not yet a universally followed standard on how these experiments are designed to clearly distinguish Bayesian updating from motivated reasoning (Thaler, 2024). Mostly for practical reasons, such experiments are restricted to a single exposure or contain a follow-up to check the durability of measured effect (Armona et al. (2018) and Roth and Wohlfart (2020), among others). While some studies find lasting effects for short-term stimuli, attitudes are generally formed over long periods of time through exposure to media, private discussion, and political canvassing (Broockman & Kalla, 2016; DellaVigna & Ferrara, 2015). In psychology, the effect of prolonged exposure patterns on attitudes and beliefs is sometimes referred to as the *illusory truth effect* (see Hasher et al. (1977) and Udry and Barber (2024) for an overview). While experimental evidence is scarce in economics, several authors study effects of this kind of exposure using observational data (Adena et al., 2015; Martin & Yurukoglu, 2017). In experimental settings, motivated reasoners change their political beliefs even in the absence of information that a Bayesian reasoner would find informative (Barron et al., 2025; Thaler, 2024). Bergman (1998) distinguishes the affective evaluation of an object and the cognitive construction thereof as two important components of attitudes. We focus on attitude change and motivated adjustment of beliefs instead of voting behavior. We propose the combination of attitude change by political as well as private communication and the adjustment of political beliefs based on attitudes as a key channel through which political actors can gain political influence that, in the case of populists, is often much greater than their direct political power.

The mechanisms we propose work as follows. Agents form attitudes about a policy in line with the established *expectancy-value theory* from social psychology (Feather, 1982) according to which individuals assess the desirability of an object by considering the sum of values of its features weighted by their beliefs that the object has these features. Agents talk with others about the policy and hence learn the beliefs, values and attitudes of the others. This social exchange leads agents to update their own attitudes, whereby they discount others' attitudes nonlinearly as a function of the distance between the attitudes (Fink et al., 1983). Agents also respond more strongly to others' attitudes, the more they perceive the others to be similar to themselves. Similarity has long been identified in social psychology as a factor affecting attitude change (Cialdini, 2021; Simons et al., 1970). Attitudinal updating leads to an incongruence between agents' updated attitudes and their prior beliefs. In order to reduce this cognitive dissonance, agents align their beliefs with their updated attitudes, which is a version of motivated reasoning and fits to the previously mentioned impact of attitudes on informational processing. Individuals' attitudes are not only influenced by communication with other individuals, but also by

narratives they receive from political communicators (parties). While moderate political communicators in principle influence individuals in the same way as other individuals, the specific communication style of a populist changes the influence of moderate political communicators on individuals. The emotionalized loss or zero-sum framing employed by the populists dominates the individuals' perception of the political communicators such that they assess similarity in the value dimension only in terms of the evaluations of the negative outcome of the policy. This communication strategy gives the populist a strong influence on attitudes and beliefs and increases polarization. For this reason, we analyze counter strategies that aim at reverting or neutralizing the zero-sum framing set by the populist (rebalancing and opportunity framing).

Our paper is related to the literature on the mainstreaming of populist ideas and discursive practices (Brown et al., 2023; Cammaerts, 2018; Schwörer, 2021) and on discursive shifts due to populism (Ekman & Krzyzanowski, 2018; Krzyzanowski, 2020; Krzyzanowski & Ekström, 2024; Krzyzanowski & Ekström, 2022). A main topic of this literature is to analyze immigration discourses in Western countries to see how they changed after the rise of populist parties. Several papers find that extreme right-wing positions on immigration became more mainstream in the public discourses of many countries. There is some disagreement on whether and why mainstream political parties became more populist. While Manucci and Weber (2017) do not find significant evidence of a "populist Zeitgeist", Schwörer (2021) does report such evidence from mainstream parties' election manifestos. Our theoretical paper cannot resolve the open empirical questions of this literature, but it can suggest new pathways for further empirical research on the impact of populist communication on the public and political discourse.

Another important literature to which our paper is close is the vast literature on opinion dynamics (Grabisch & Rusinowska, 2020; Hassani et al., 2022; Noorazar, 2020). While this literature has produced an impressive array of model specifications and fascinating dynamics, it leaves still ample room for further advances. In his survey of this literature, Noorazar (2020) concludes that there is a gap between the simplicity of the theoretical models and the complexity of humans' behavior. He calls for more realistic interaction structures such as networks with homophily and "more applications of opinion dynamics in real world problems" (p. 521). The survey by Hassani et al. (2022) agrees that similarity-based interaction rules would be interesting to study. Furthermore, these authors suggest to have a fresh look on the influence of dissimilar opinions in future research. Many opinion dynamics models build on the bounded confidence model (Hegselmann & Krause, 2002), in which dissimilar opinions beyond a threshold are ignored and have no impact on the receiver's opinion. The present model deviates in several respects from the opinion dynamics literature and addresses the mentioned issues. First, inspired by narrative economics (Roos & Reccius, 2023) we do not focus solely on the exchange of opinions, but of narratives, which we define as sense-making stories and which are more complex objects than opinions. Narratives serve at least two

functions: they provide meaning and guidance for behavior in a complex and hard-to-understand world and they create group identity, that helps coordinate collective behavior. We conceptualize narratives as stories that transport the senders' attitudes, beliefs and values about policies, which are all highly relevant in political communication. Second, we propose a cognitive model of narrative processing based on communication research (Chung & Fink, 2022; Fink et al., 1983) that aims at more psychological complexity and avoids the discontinuous impact of dissimilar opinions in the bounded confidence models. Finally, we use a network featuring homophily to represent a more realistic interaction structure.

2 A model

The model contains two types of agents: *citizens* and *political initiators* (*PI*s). It describes the evolution of attitudes and beliefs regarding an economic policy domain within a network of citizens. Two mainstream *PI*s each disseminate a narrative promoting their respective policy position. *PI*s seek to expose citizens to their narratives as broadly and frequently as possible, assuming that repeated exposure to a narrative gradually aligns citizens' attitudes and beliefs with the position it promotes. At a fixed point during the simulation, *PI*s and citizens are confronted with a *populist shock*: the entry of a third *PI* with a populist agenda. Citizens in the network have a dual role. On the one hand, they receive policy narratives and are susceptible to political influence; on the other hand, they act as senders, sharing their own evolving narrative with their network neighbors. Over time, citizens continuously update their attitudes and refine their beliefs by considering and making sense of the narratives they receive from political actors and their network peers. Citizens are heterogeneous with respect to their beliefs regarding the policy's outcomes, their policy-relevant values, and their overall attitudes towards the policy. While values are fixed at initialization, beliefs and attitudes evolve endogenously through exposure to narratives from *PI*s and network peers. The model is implemented in NetLogo 7.0.2 (Wilensky, 1999).

2.1 Narratives about a policy trade-off

Let $\mathcal{P}$ denote the spectrum of positions within a given policy domain and let $p_0 \in \mathcal{P}$ be the policy stance currently in effect. We will focus on immigration policy, a prototypical example for a high-stakes domain often subject to populist pressures (Dinas et al., 2019; Edo et al., 2019; Steinmayr, 2021). Against this backdrop, let $\mathcal{O} = \{o^+, o^-\}$ be the set of major economic outcomes typically associated with moving from p_0 to a more liberal policy stance p_1. We define o^+ to be the beneficial labor market effects that may result from firms attracting more migrant workers and o^- to be the fiscal burden that may be sustained should labor market integration fail. Furthermore, let $\delta_{it} \in \{0, 0.25, 0.5, 0.75, 1\}$ denote the relative confidence individual i assigns to o^+ at time point t when evaluating p_1. Conversely, $1 - \delta_{it}$ denotes the relative confidence i assigns to o^- at t. δ_{it} fully describes i's beliefs at time t regarding the consequences of implementing p_1.[1] A primary motive for employing narratives in politics is to reduce the inherent complexity of issues and enable persuasive political messaging. Thus, we argue that, even though $\mathcal{P}$ might actually imply a much higher-dimensional outcome space, the political arena actively reduces it to a simple dichotomy between o^+ and o^-.[2]

[1] Note that p_1 can generate o^+ and o^- concurrently and to varying degrees, so that δ_{it} cannot be interpreted as a Bernoulli probability $P(o^+)$.

[2] The direction of the policy shift is arbitrary. To consider a conservative shift from p_0, we could define o^+ to be the consolidation of the social benefits system and o^- the negative labor market effects of alienating

An individual with $\delta_{it} = 1$ is said to have a *fully optimistic* belief. Such an individual places full confidence in p_1 only resulting in o^+ without any consideration for the possibility of o^- emerging as a side-effect. An individual with $\delta_{it} = 0$ possesses a *fully pessimistic* belief, which predicts p_1 to fully backfire and only result in o^-. Individuals with $\delta_{it} = 0.75$ ($\delta_{it} = 0.25$) are optimistic (pessimistic), while at the same time recognizing the danger (promise) of o^- (o^+). Lastly, an individual with $\delta_{it} = 0.5$ is characterized by a *neutral* belief. The latter weighs both outcomes equally, thereby fully accentuating the uncertainty involved in the trade-off implied by $\mathcal{O}$.

Let $\boldsymbol{\nu}_i = (\nu_i^+, \nu_i^-)$ be the vector of $\mathcal{P}$-relevant values of individual i. It reflects the strength or weakness of an individual's preferences over $\mathcal{O}$. It holds that $\nu_i^+ \in \mathbb{R}_0^+$ and $\nu_i^- \in \mathbb{R}_0^-$, so that all individuals are at least (at most) indifferent to o^+ (o^-). This assumption precludes disagreement about whether elements in $\mathcal{O}$ represent benefits or costs. Because we conceptualize values to primarily represent deep-seated ideological preferences, $\boldsymbol{\nu}_i$ is fixed over time. This also means that, while they are functionally connected to the *economic* outcomes of p_1, values are influenced by *ideological* openness to migrants. However, depending on individual factors, material self-interest could come into play. For example, ν_i^+ may be higher for individuals who benefit from immigration filling gaps in domestic labor supply. This could be the case for employers or consumers of services such as care work whose demand would otherwise remain unmet.

A full policy narrative $N_{it} = (\delta_{it}, \boldsymbol{\nu}_i, A_{it})$ consists of the set of i's current beliefs $(\delta_{it}, 1 - \delta_{it})$ over potential policy outcomes $\mathcal{O}$, the vector of i's $\mathcal{P}$-relevant values $\boldsymbol{\nu}_i$ and i's current policy attitude A_{it}. The latter represents that individuals' current summary evaluation of the policy change to p_1. The dynamic adjustments of citizens' attitudes and beliefs are subject to their exposure to narratives and will be discussed later. Before any narratives are shared, however, i's initial policy attitude A_{i0} is calculated as a convex combination of i's values, weighted by i's beliefs:

$$A_{i0} = \delta_{i0}\nu_i^+ + (1 - \delta_{i0})\nu_i^- \tag{1}$$

Figure 1 is a graphical illustration of how p_1 maps to i's initial policy attitude A_{i0}.

*PI*s choose narratives as their mode of political communication because it aligns with how citizens naturally make sense of the world. Through their chosen narratives, *PI*s aim to achieve public resonance and, consequently, boost the persuasive potential of their positions. Accordingly, the structure of N_{it} and the computation of A_{i0} apply equally to both agent types.

Citizens' value components are independent and identically distributed (iid) draws from common scaled and shifted Beta(α, β) distributions. Independence holds both within value

migrant workers. If the signs of relevant parameters are flipped accordingly, the model stays conceptually identical.

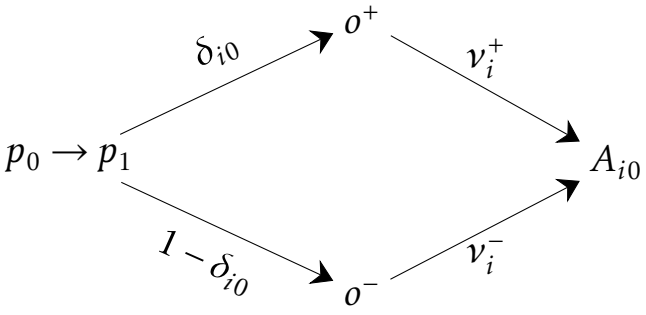
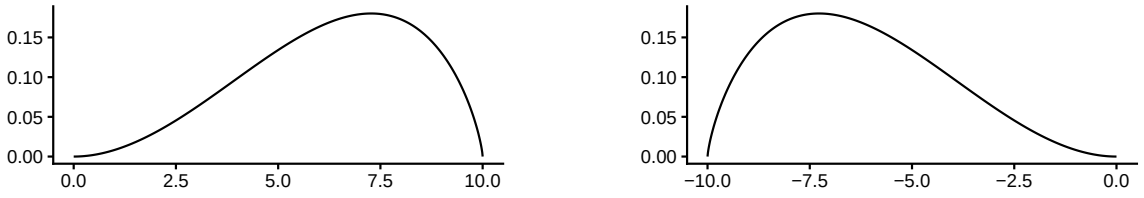


Figure 1: **The Narrative Diamond**: the cognitive mapping for individual i from a suggested policy change (p_0 to p_1) to its two potential outcomes (using initial beliefs) and from outcomes to the initial policy attitude A_{i0} (using values).



(a) ν^+ (b) ν^-

Figure 2: Probability Density Functions (PDFs) of values among citizens.

dimensions and across citizens:

$$\begin{aligned} \nu_i^+ &= 10 \cdot X, & X &\sim \text{Beta}(3, 1.75) \\ \nu_i^- &= 10 \cdot Y - 10, & Y &\sim \text{Beta}(1.75, 3) \end{aligned} \tag{2}$$

The theoretical probability density functions (pdfs) of these distributions are visualized in Figure 2. Citizens' values are bounded, with $\nu_i^+ \in (0, 10)$ and $\nu_i^- \in (-10, 0)$. Both distributions are skewed towards larger absolute values. First and foremost, this implies that citizens tend to not be indifferent about $\mathcal{O}$. Given that citizens in our model are assumed to regularly engage in a form of political discussion, such individuals can reasonably be assumed to likely have distinct political preferences.

Citizens start out with the neutral belief $\delta_{i0} = 0.5$ so that all initial variation in A_i is caused by the heterogeneity in values. Citizens' policy attitudes correspond to their level of support for p_1. This means that, in a public opinion poll or a referendum held at t, any citizen C_i with $A_{it} > 0$ would vote in favor of implementing p_1. Note that, as implied by Equation 2, the upper (lower) attitude limit can only be reached with a fully optimistic (pessimistic) belief and a maximally liberal $\boldsymbol{\nu}_i = (10, 0)$ (conservative $\boldsymbol{\nu}_i = (0, -10)$).

Since *PI*s are goal-oriented actors, they operate with fixed sets of values and belief. At the

start of the simulation, we model two PIs, endowing the leftist PI_L with $\boldsymbol{\nu}_L = (7, -3) \wedge \delta_L = 0.75$ and the conservative PI_C with $\boldsymbol{\nu}_C = (3, -7) \wedge \delta_C = 0.25$. Hence, following Equation 1, $A_L = 4.5$ and $A_C = -4.5$. Across agent types and $\forall t$, it holds that $A_{it} \in (-10, 10)$, so that A_L and A_C can be considered moderate policy positions held by mainstream political actors. Since implementing policies has opportunity costs, attitudes of PIs can also be understood as the political capital they are willing to invest in pushing or opposing p_1. Note that N_L and N_C are mirror-images of each other, with PI_L voicing more optimism than PI_C about the labor market potential encoded in o^+ and less concern about the downside of o^- . The related symmetry between value profiles $\boldsymbol{\nu}_L$ and $\boldsymbol{\nu}_C$ also suggests an ideological equilibrium. The moderate center-left $\leftrightarrow$ center-right dynamic inherent in this parameterization has been characteristic for the political dynamics in most Western democracies in the second half of the 20th century. It has also served as a reference point for other economic models of politics, most notably the median-voter model (Downs, 1957). The directional tension between the components of N_L and N_C still gives each PI a distinct position on p_1. Since the value distributions in Figure 2 are symmetric, the population is ideologically balanced and neither PI_L nor PI_C has any systematic advantage over its competitor, on average.

It is worth noting that our treatment of $\boldsymbol{\nu}_i$ departs from the conceptualization of narratives established by Eliaz and Spiegler (2020), who model them as causal beliefs over outcomes, and from Eliaz et al. (2024), who acknowledge the importance of preferences but regard them as external to the narrative. In contrast, we contend that policy narratives are inextricably tied to values. Political persuasion involves more than the technocratic task of explaining prospective outcomes. Few, if any, economic policies are Pareto efficient, so that policy-making always requires politicians to balance diverging interests. By including $\boldsymbol{\nu}_i$ as part of the narrative, we acknowledge how politicians articulate values and priorities to justify their policy positions in addition to describing the effects of their proposals.

2.2 A political discussion network

The set of citizens $C = \{1, ..., 100\}$ is involved in an undirected political discussion network (C, g) where a link $ij \in g$ defines a channel through which citizens i and j can exchange narratives. In total, g encompasses 250 links, so that the average number of links in the network — its average degree — is 5. We assume that the social structure of political interactions is inertial relative to the timescale implied by the model. Hence, g is grown at initialization before any narrative interactions take place and remains fixed for the duration of each simulation.

The network formation process is subject to homophily.[3] Generally speaking, citizens form

[3] see Jackson (2010) for a survey on homophily

links in mutual agreement based on a pairwise comparison of each other's value systems. Let $s_{ij} \in [s_{\min}, 1]$ denote the value-similarity between citizens i and j, with $s_{ij} = 1$ indicating identical value sets and $s_{\min}$ greatest possible value distance. Until the number of links in g reaches 250, links are established iteratively according to the following algorithm: A pair of unconnected citizens i and j is selected at random and form a link with probability

$$P(ij) = {s_{ij}}^{\lambda}. \tag{3}$$

The hyperparameter $\lambda \geq 0$ governs the extent of value-based homophily in g. We assume that $s_{\min} > 0$, so that even highly dissimilar citizens have a strictly positive probability of forming a link, that decreases towards 0 as λ increases. Similarity is defined using the radial similarity function

$$s_{ij} \;=\; \exp\left(-\left(\frac{d_{ij}}{\sigma}\right)^{\beta}\right) \tag{4}$$

applied to the Euclidean distance between value vectors

$$d_{ij} \;=\; \sqrt{(\nu_i^+ - \nu_j^+)^2 + (\nu_i^- - \nu_j^-)^2}. \tag{5}$$

We set $\sigma = 9$ and $\beta = 2.5$ to attain a reasonable decay of s_{ij} given the range of $\boldsymbol{\nu}$. Figure 3 illustrates the full range of value similarities produced by this parametrization when comparing any value profile $\boldsymbol{\nu}_j$ to that of a hypothetical citizen C_i with $\boldsymbol{\nu}_i = (10, 0)$. While our model concerns a single policy field, we use d_{ij} as a proxy for overall ideological distance. This is because individuals' policy preferences tend to be correlated across domains.

Our goal is to attain a network where assortativity and clustering occur as a result of shared values. Rather than replicating a certain degree distribution through preferential attachment (Barabási & Albert, 1999) or assuming clustering by imposing a small-world network structure (Watts & Strogatz, 1998), our approach emphasizes that individuals primarily discuss politics with those whose fundamental political values resemble their own. Accordingly, the mechanism implied by Equation 3 models the process of political acquaintanceship: In the case of $\lambda = 0$, every encounter results in a link and g is a random network. As λ increases, links become increasingly dependent on shared values. Given the probabilistic nature of link creation, higher values of λ results in delayed convergence of network formation: Citizens i and j might meet several times before establishing a connection, even if their value profiles closely match. The number of meetings required follows a geometric distribution with $p = s_{ij}^{\lambda}$. This means that if, for example, $s_{ij} = 0.95$, the expected number of meetings required for a link to form is approximately 1.05 with

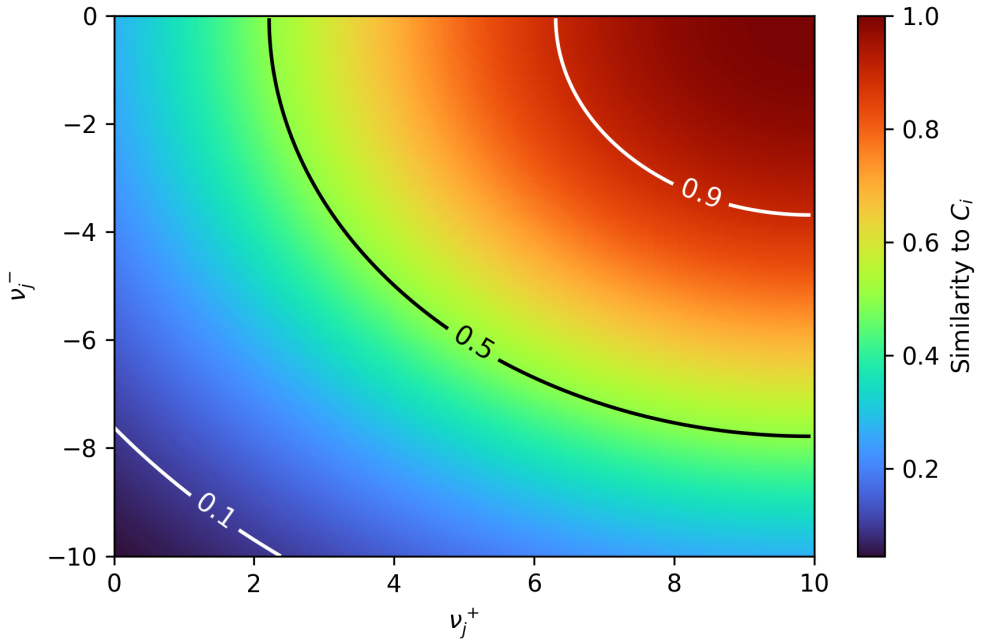


Figure 3: **Value Similarity:** A heatmap displaying the value similarity of a citizens C_j with an arbitrary value profile to a hypothetical citizen C_i with $\boldsymbol{\nu} = (10, 0)$

$\lambda = 1$, it is 2.79 with $\lambda = 20$, and 13 with $\lambda = 50$.

Network parameterization. Empirical political discussion networks generally exhibit large amounts of clustering. We vary network parameterizations to span a broad range of different clustering coefficients (CCs) and average path lengths (APLs). The λ values are spaced to reflect the non-linearity in Equation 3. With $\lambda > 400$, the network formation process does not reliably converge, resulting in increasing numbers of isolated citizens. As a result, the maximum value of λ we use is 400. Table 1 reports average CCs and average APLs for all parameterizations, ranging from random graphs ($\lambda = 0$) through various levels of homophily. While CCs increase monotonically in λ, average APL starts to increase only with $\lambda \geq 50$. Figure 7 displays the value-space and link-structure of an example network, featuring identical citizen sets, comparing $\lambda = 0$ to $\lambda = 100$.

At the highest level of homophily, with $\lambda = 400$, we attain networks with a mean CC of 0.439 and a mean APL of 4.701. This combination of substantially increased CC relative to random networks and shorter average path length than regular lattices with similar node count ($APL^{\text{Lattice}} \approx 9$), reflects the *small-world property*. It is a known characteristic of empirical political discussion networks Jürgens et al. (2011) and of social and economic networks more broadly. [4]

[4] See Uzzi et al. (2007) for a review.

Metric	$\lambda = 0$	$\lambda = 10$	$\lambda = 25$	$\lambda = 50$	$\lambda = 100$	$\lambda = 200$	$\lambda = 400$
$\overline{\text{CC}}$	0.049	0.066	0.095	0.135	0.199	0.300	0.439
	(0.010)	(0.012)	(0.016)	(0.020)	(0.022)	(0.035)	(0.036)
$\overline{\text{APL}}$	3.006	2.995	3.088	3.304	3.645	4.141	4.701
	(0.029)	(0.058)	(0.083)	(0.139)	(0.204)	(0.246)	(0.399)

Table 1: **Summary of network statistics:** Means of clustering coefficients and average path lengths for all network parameterizations, one the basis of 700 runs $\forall\lambda$. Standard deviations in parentheses.

2.3 Attitude and belief updating

From this point forward, for clarity, we denote the individual communicating (or "sending") a narrative with the subscript s, and the individual receiving the narrative with the subscript r.

While the reach of citizens is confined to their respective neighborhoods, PIs distribute their narratives globally. Within our framework, we consider formats such as TV commercials, social media campaigns, and talk radio interviews to be interchangeable media for narrative dissemination. At the start of each period t, every PI broadcasts their narrative to each citizen independently with pairwise probability $p_B = 0.5$. All narrative interactions are memoryless which means that citizens are very likely to receive the same policy narratives from the same PIs repeatedly over time. After top-down political influence is exerted in each t, every citizen who possesses one or more links fulfills the function of narrative sender and tells their narrative N_s to one randomly selected neighbor C_r. This dual role of citizens as senders and receivers of narratives highlights once more the distinction between citizens and PIs. In contrast to PIs, citizens are *sense-makers* who not only voice their own dispositions, but use their social neighborhoods to reconcile them with the views of PIs and fellow citizens. The random nature of C_s choosing C_r, along with heterogeneity in node degree, creates substantial variation in the number of narratives citizens receive in each period.

Influence over attitudes: As a first-order effect, receiving narratives influences a receivers attitude A_{rt} according to the Psychological Discrepancy Discounting (PDD) model (Fink et al., 1983). When comparing attitudes between two individuals, the PDD model differentiates between *positional discrepancy* and *psychological discrepancy*. The former is simply the objective numerical difference between A_{rt} and the senders' attitude A_{st}:

$$d_{srt} = |A_{rt} - A_{st}| \tag{6}$$

To assess how disruptive A_{st} feels to C_r, Fink et al. (1983) scale d_{srt} by the range of attitudes deemed plausible by C_r, denoted R_{r0}. This normalization yields the psychological discrepancy ψ between A_{rt} and A_{st}:

$$\psi_{srt} = \frac{d_{srt}}{R_{r0}}. \tag{7}$$

We define R_{r0} as the range of initial attitudes that occur in C_r's neighborhood at $t = 0$. This assumption is supported by the empirical literature on *perceived polarization* which finds that individuals whose social environment features a narrow range of political attitudes tend to perceive the same attitudinal difference as larger compared to those with broader exposure (Steiglechner et al., 2025). Since this perception is triggered when the salience of policy issues is high (Bordalo et al., 2022), it applies naturally to our policy domain $\mathcal{P}$. Elevated levels of perceived polarization contribute to the erosion of trust in diverging information individuals receive from political actors (Enders & Armaly, 2018) and their social environments (Lee, 2022). Equation 7 captures this mechanism.

In any period t, C_r may receive a total of M narratives. If $M > 1$, C_r does not iteratively update attitudes between narrative interactions but instead evaluates every N_m independently against A_{rt}, storing the resulting M *provisional updates* until the end of the period. This means that citizens have the capacity to ponder different, possibly diametrically opposed, attitudes and narratives in the short term without adapting to them instantaneously. Consider also that some narrative interactions within a given period will, by chance, be bidirectional, meaning a pair of citizens (C_r, C_s) could exchange sender-receiver roles. The short-term tolerance for cognitive dissonance that underlies provisional updating therefore enables us to naturally accommodate not only one-sided broadcasting, but also genuine, reciprocal political discussion through narratives.

The receiver's mth provisional update in period t is computed as

$$A_{rt}^{m} = \frac{A_{rt} + t_{srt} \cdot \Delta(\psi_{srt}) \cdot A_{st}}{1 + t_{srt} \cdot \Delta(\psi_{srt})} \tag{8}$$

where t_{srt} is a weight indicating trust in the sender. $\Delta(\cdot)$ is the effective discount factor applied to A_{st}, based on the psychological discrepancy it presents to C_r. The latter is defined as

$$\Delta(\psi_{srt}) = \exp(-\gamma \cdot \psi_{srt}). \tag{9}$$

The global discount rate $\gamma \geq 0$ indicates citizens' sensitivity to attitude discrepancies. If $\gamma = 0$, no discounting occurs and Equation 8 reduces to a standard DeGroot-style

weighted average between A_{st} and A_{rt} (Degroot, 1974). Recalling Equation 7, it is now clear that higher levels of perceived polarization result in stronger discounting of any $A_{st} \neq A_{rt}$. Note that the model excludes the possibility of a *backfire* or *boomerang effect* (van der Linden, 2021; Wood & Porter, 2018), where very large attitude discrepancies could cause the receiver to be repelled and update in the opposite direction. Instead, influence is strictly positive, decaying smoothly and monotonically in ψ_{srt}.

When all interactions in t have occurred and after having received M narratives in total, C_r's updated attitude is calculated as

$$A_{rt+1} = \frac{1}{M} \sum_{m} A_{rt}^{m}. \tag{10}$$

Discounting without directional motives: Consider the first round of updates: At the end of period $t = 0$, the attitude revision specified in Equation 10 yields a new A_{r1}. Any non–zero change in attitude breaks the original coherence among A_{r0}, $\boldsymbol{\nu}_r$, and δ_{r0} defined in Equation 1. At the end of every period, to alleviate this emergent cognitive dissonance, C_r performs a second-order sense-making step, re-evaluating belief δ_{it}. The search is restricted to the status-quo δ_{rt} and the two adjacent beliefs:

$$\Theta_{rt+1} = \left\{\delta_{rt} - 0.25,\ \delta_{rt},\ \delta_{rt} + 0.25\right\}. \tag{11}$$

Given the attitude update, the revised δ_{rt+1} corresponds to the belief that minimizes the *cognitive dissonance* between A_{rt+1} and any candidate belief in Θ_{rt+1} :

$$\delta_{rt+1} \;=\; \underset{\tilde{\delta}_{rt+1} \in \Theta_{rt+1}}{\arg\min} \left| A_{rt+1} - \left[\tilde{\delta}_{rt+1}\, \nu_r^{+} + \left(1 - \tilde{\delta}_{rt+1}\right) \nu_r^{-}\right]\right|. \tag{12}$$

This motivated reasoning heuristic captures the idea that individuals seek coherence between beliefs, values and attitudes. In conjunction with the narrative-driven first-order attitude update, it closes an endogenous feedback loop through which adaptation of policy attitudes translates into gradual shifts of beliefs.

Pre-shock narrative trust: While the PDD mechanism formalized in Equation 9 determines how the receiver is influenced by incoming *attitudes*, the narrative trust variable t_{srt} in Equation 8 captures the receivers (perceived) alignment of factual beliefs and political preferences with the sender. As with value similarity in Equation 4, trust is computed with a radial similarity function. Its argument is the weighted Euclidean distance between the receiver's vector $\begin{pmatrix} \delta_{rt} \\ \boldsymbol{\nu}_r \end{pmatrix}$ and the sender's vector $\begin{pmatrix} \delta_{st} \\ \boldsymbol{\nu}_s \end{pmatrix}$:

$$t_{srt} \;=\; \exp\left(-\sqrt{[10(\delta_{rt}-\delta_{st})]^2+\|\boldsymbol{\nu}_r-\boldsymbol{\nu}_s\|^2}/10\right). \tag{13}$$

Low trust induces motivated skepticism when r evaluates an incoming narrative (Taber & Lodge, 2006) [5]. Depending on the regime and the identity of the sender, $\boldsymbol{\nu}$ may contain both value dimensions or only one: For all interactions among citizens, it holds that $\|\boldsymbol{\nu}_r-\boldsymbol{\nu}_s\|^2=\left(\nu_r^{k_1}-\nu_s^{k_1}\right)^2$, which means that C_r selects the value dimension along which she is more closely aligned with the sender. We refer to this value as the k_1 value-component, following the terminology used by Iijima and Kamada (2017). The k_1 component should not be confused with the *first* component of a value vector, ν^+. Formally:

$$k_1=\underset{j\in\{+,-\}}{\arg\min}\left|\nu_r^j-\nu_s^j\right|. \tag{14}$$

This specification has two implications. First, comparing values dimension by dimension implies that citizens evaluate each value as an independent trait. Second, in peer interactions between citizens, the receiver's focus on the k_1 component deliberately emphasizes the "common ground" in political preferences. Accordingly, when $\nu^k=\nu^{k1}$, we refer to the resulting trust measure as *common-ground-trust*:

$$t_{srt}=t_{srt}^{\text{common}}=\;\exp\left(-\sqrt{[10(\delta_{rt}-\delta_{st})]^2+(\nu_r^{k1}-\nu_s^{k1})^2}/10\right)\quad\forall\; s\notin\{C,L,P\} \tag{15}$$

2.4 A populist shock

At $t=5000$, the model is subjected to a shock, triggered by the entry of a populist political actor PI_P, who is characterized by values $\boldsymbol{\nu}_P=(0,-10)$ and belief $\delta_P=0$. The fully pessimistic belief $\delta_P=0$ assigns no credence to o^+. It yields a one-sided narrative that exclusively emphasizes the risks and costs of immigration. The values $\boldsymbol{\nu}_P$ also disregard o^+. Positive labor market effects thus do not enter the populist's normative evaluation. This parameterization captures three central features of exclusionary populism: First, the extreme value structure is consistent with the belief. A populist who expects no benefits from p_1 has no reason to value such benefits. Second, within a populist logic, supply side benefits are not seen as normatively compelling if they are perceived to benefit "elites" and corporations while harming the earning potential of "the true people". This framing serves both dimensions of the populist *us-against-them* logic. It evokes the conspiratorial notion of mainstream politicians plotting against their own people which acts as the narrative glue between anti-elitist and xenophobic sentiment.

[5] Note that the belief-term in Equation 13 is scaled up by a factor of 10 to match the belief-dimension to the range of the value dimensions.

Third, emphasizing fiscal risks through ν^- triggers loss-aversion, which complements the broader populist agenda. It provides an economic justification for a platform that prioritizes cultural preservation and ethno-national identity over economic efficiency or aggregate welfare.

A key part of populist politics is the moralization of issues, which entails putting the central normative components of populist position to the forefront of the discourse. While a full risk assessment of $\mathcal{O}$ would, for example, account for the long-term risks of no immigration, the loss-averse evaluation in $\boldsymbol{\nu}_P$ resonates by tapping into fears of societal and personal decline. Accordingly, beyond the populists mere entry into the model, the shock triggers a more fundamental change in how the globally distributed PI-narratives are perceived by citizens. We model this as a switch of the value component that is salient when computing t_{rst} in PI-citizen interactions. Only for the latter, $\|\boldsymbol{\nu}_r - \boldsymbol{\nu}_s\|^2 = (\nu_r^- - \nu_s^-)^2$ now holds, reflecting a loss of awareness of the labor market opportunity o^+. Upon populist entry we therefore adapt Equation 13:

$$t_{srt} = t_{srt}^{\text{zero-sum}} = \exp\left(-\sqrt{[10(\delta_{rt} - \delta_{st})]^2 + (\nu_r^- - \nu_s^-)^2}/10\right) \quad \forall\, s \in \{C, L, P\} \tag{16}$$

The newly established salience of ν^- fully aligns with the priorities reflected in $\boldsymbol{\nu}_P$ and δ_P. Intuitively, the populist succeeds in dominating p_1-related discourse by imposing components of N_P on all PI-citizen communication.

Interventions: A key question for moderate political actors and institutions is how to limit the impact of populist narratives. We analyze three alternative interventions that are put into practice at $t = 5300$. The post-shock latency of 300 periods reflects the time required for the populist threat to be recognized and for counter-measures to be devised.

The first intervention is a material change of N_C. As the results in section 3 will show, the populist entry causes a downward shift in the average attitude among citizens, while pulling citizens away from the conservative's belief. When observing this drift, the conservative may be inclined to conclude that his own moderate belief δ_C lacks a clear profile and that less ambiguous conservative messaging could counteract the populists downward pressure on attitudes and beliefs. The *pandering* intervention therefore encompasses the PI_C strategically adopting the rhetoric of the fully pessimistic populist belief $\delta_C = \delta_P = 0$.

The second and third interventions do not entail PIs changing policy positions. Instead, they represent collective PI-efforts to counteract $t_{rst}^{zero-sum}$, which was imposed by the populist in Equation 16. Both strategies counter the populist by redirecting the relative salience of outcomes in PI-citizen interactions. Specifically, *rebalancing* restores t_{srt} for PI-citizen interactions, thereby undoing the populists' broader discursive effect. Hence, after rebalancing, the populist becomes a normal political actor without the capacity to

consistently re-frame the discourse. Intuitively, rebalancing might entail expanding the loss frame to account for the expected costs of no immigration, such as labor supply bottlenecks and strained pension systems.

The final intervention is *opportunity framing*, which entails anchoring v^+ as the salient value component in PI-citizen interactions:

$$t_{srt} = t_{srt}^{\text{opportunity}} = \exp\left(-\sqrt{[10(\delta_{rt} - \delta_{st})]^2 + (v_r^+ - v_s^+)^2}/\sigma\right) \quad \forall\, s \in \{C, L, P\} \tag{17}$$

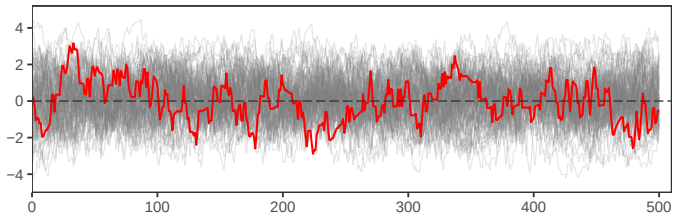
Establishing trust in o^+, to the extent that it surpasses the loss-aversion associated with o^-, arguably represents a more fundamental paradigm shift than the first two interventions. Intuitively, it could be achieved through a credible institutional commitment to policies like integration programs, language training and supply-side reforms, such as cutting red tape around the recognition of professional credentials.

3 Results

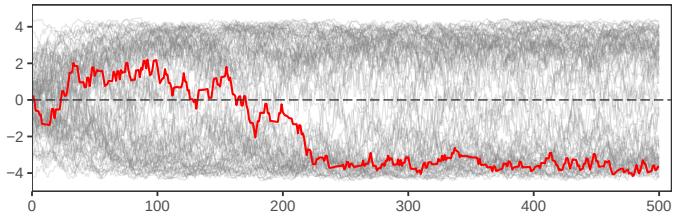
We present our results in three steps. First, we analyze how attitudes form prior to the entry of a populist and how the attitude distribution depends on network homophily and the intensity of attitude discounting. Second, we consider a populist shock: the arrival of a populist political actor that broadcasts a narrative focusing on the negative outcome of the policy only and influences the discourse accordingly. Finally, we consider three counter-reactions of the established mainstream political actors: conservative pandering, balance restoration and opportunity framing. Analyses were conducted with the statistical software package R (R Core Team, 2025).

3.1 Attitude and belief formation before the populist shock

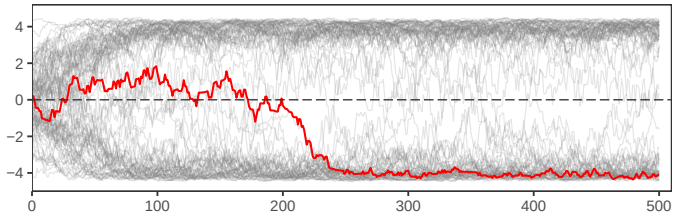
In the initial setup with two mainstream PIs, model dynamics vary significantly across parameter combinations. For a first set of analyses, we aim for a systematic exploration of the full parameter space. Thus, we also include very large values of the discount rate $\gamma > 1$, which represent extreme attitude discounting that can effectively turn citizens into semi-stubborn network nodes.



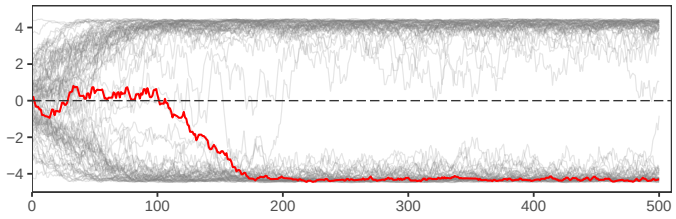
(a) $\gamma = 0.2$



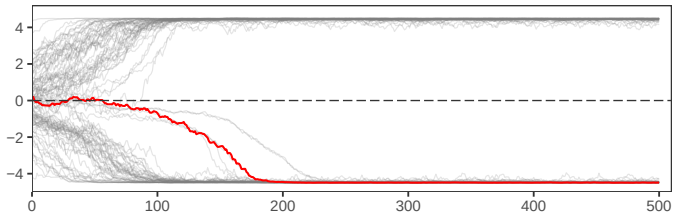
(b) $\gamma = 0.4$



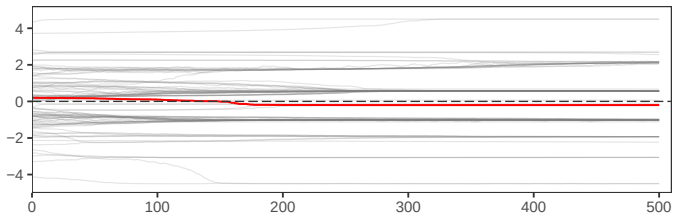
(c) $\gamma = 0.5$



(d) $\gamma = 0.6$



(e) $\gamma = 1$



(f) $\gamma = 10$

Figure 4: **Individual attitude trajectories of citizens for** $1 \leq t \leq 500$ **with** $\lambda = 100$ for selected values of γ. Red trajectory marks an example citizens. The latter and the random seed are identical across configurations.

Attitude polarization. Polarization can be measured in multiple ways (Bramson et al., 2017). To characterize how the shape of the attitude distribution and its tendency to exhibit polarization vary across (λ, γ), we first consider histograms of A_i. Because single runs are subject to stochastic variation, we average each histogram across runs to obtain a stable picture of the typical distributional shape for each parameter combination. Figure 19 in the appendix shows a grid of these average histograms across combinations of λ and γ at $t = 4999$. Averaging can, however, obscure distributional heterogeneity if individual runs diverge toward opposing equilibria (Lee et al., 2015). We therefore supplement this aggregate view with time-series of A_i for individual example runs.

Figure 4 shows all individual attitude trajectories of citizens in the model's burn-in period for representative runs and selected γ settings in homophilous, $\lambda = 100$ networks. It also indicates how attitudes are distributed upon network initialization. This distribution is independent of λ and γ and, as a result, fairly stable. Since $\delta_0 = 0.5$ and given the symmetry and independence of the value distributions, initial attitudes are dispersed roughly symmetrically around 0. In general and for all parameter settings, attitudes tend to fall between -4.5 and 4.5. While lower and higher attitudes can occur at different t for individual citizens, the frequent dissemination of the fixed PI-narratives establish de-facto bounds on A_i.

When discounting is kept minimal by restricting $\gamma \leq 0.2$, citizens appear to lean centrist according to Figure 19 at any given t, regardless of the level of homophily in the network. A comparison with Figure 4a reveals that this pattern is not reflective of reliably centrist attitudes in individual trajectories. Instead, citizens are swayed to change A_i drastically according to received narratives, even if incoming attitudes present large positional discrepancies. Since $p_B = 0.5$, citizens are frequently confronted with PI-narratives, which are equally distributed between PIs. The combination of these mechanisms cause A_i to resemble a random walk that is roughly bounded by PI attitudes. At any t, these dynamics yield the unimodal, mean-0 attitude distribution over citizens observed in Figure 19. Increasing γ strengthens and stabilizes attraction to PIs: As Figure 4e indicates, setting $\gamma = 1$ causes attitudes to diverge from the center and cluster tightly around those of the PIs. At this intensity, discounting is too strong for casual narrative encounters between citizens to exert any meaningful influence at any homophily level. Due to the constant exposure to PI narratives, this results in a stepwise assimilation towards one of the PI's attitudes. With $\gamma \geq 5$, the same centrist pattern observed for low γ emerges in the average, but for opposite reasons: With sender attitudes getting heavily discounted, citizens barely change their attitudes in response to incoming narratives. This effect increases in λ, as citizens are also increasingly shielded from contrarian narratives. Figure 4f illustrates this dynamic for $\gamma = 10$: Given their stubbornness, most citizens are essentially too centrist to be influenced by moderate PIs, resulting in an attitude distribution that settles to multiple local modes which remain virtually unchanged over time.

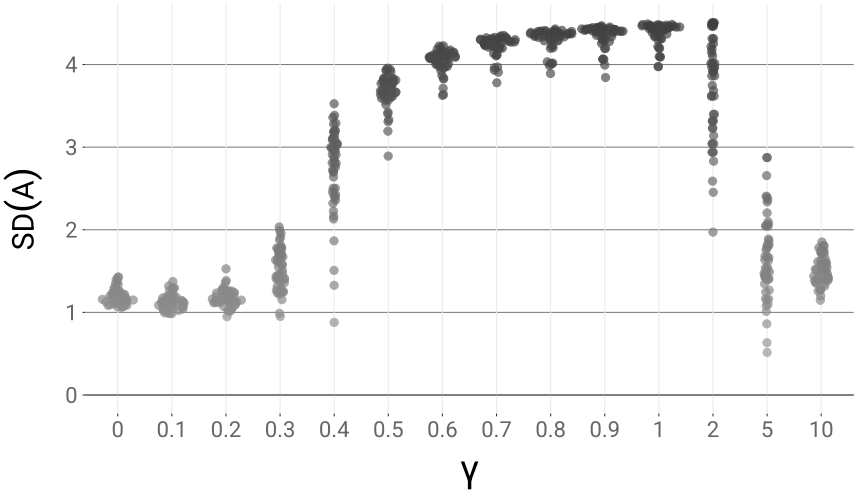


Figure 5: **Attitude polarization across runs** at $\lambda = 100$ measured as its standard deviation.

Given the patterns analyzed, the most plausible attitude adjustment patterns occur for mid-range selections of $0.3 \leq \gamma \leq 1$. The progression of polarization upon increasing λ is qualitatively similar across this γ-range, but the onset of bimodality shifts to progressively lower degrees of homophily: at $\gamma = 0.3$, the attitude distribution starts hollowing out for $\lambda \geq 200$. At $\gamma = 0.6$, the critical point is at $\lambda \geq 10$. For $0.7 \leq \gamma \leq 1$, bimodality is apparent even in non-homophilous networks, with the amount of concentration that is observable around the modes increasing in λ.[6]

In addition to the modal structure apparent in the histograms, we follow Santos et al. (2021) and rely on the standard deviation to measure attitude polarization. Figure 5 provides a detailed analysis of how polarization in individual runs contributes to the aggregated distributions. The $\lambda = 100$ case serves as an example, but patterns are similar across the range of λ values. Standard deviations are tightly clustered around a low mean of around 1 for $\gamma \leq 0.3$. At $0.3 \leq \gamma \leq 0.5$, the model transitions from a low-variance regime to a high-variance regime with SD(A) ≈ 4. Beyond that threshold, polarization saturates and stays high. For extreme discounting at $\gamma \geq 2$, polarization gradually dissolves, as overall narrative influence on attitudes in the network is reduced towards 0.

At the critical value $\gamma = 0.4$, variability between runs is large, indicating sensitivity to slight probabilistic variations in initial conditions. One of these initial conditions is the value-homogeneity within neighborhoods, which is described in the model by citizens' initial attitude range R_0. The third column of Table 2 in the appendix regresses SD(A) on its per-run average $\overline{R_0}$ for $\lambda = 100$ at the critical point $\gamma = 0.4$. The coefficient is negative and significant on a $\alpha = 1\%$ level of significance, so that an increased range of attitudes in neighborhoods is shown to decrease the amount of polarization that emerges. The R^2

[6] Figure 15 in the appendix demonstrates that the observed polarization is not a result of PI-independent, within-network dynamics. It plots the trajectories of individual attitude in a single network without PI influence. All citizens eventually reach a consensus attitude. However, this consensus is affected by extreme, stubborn individual attitudes affecting the average before converging towards it.

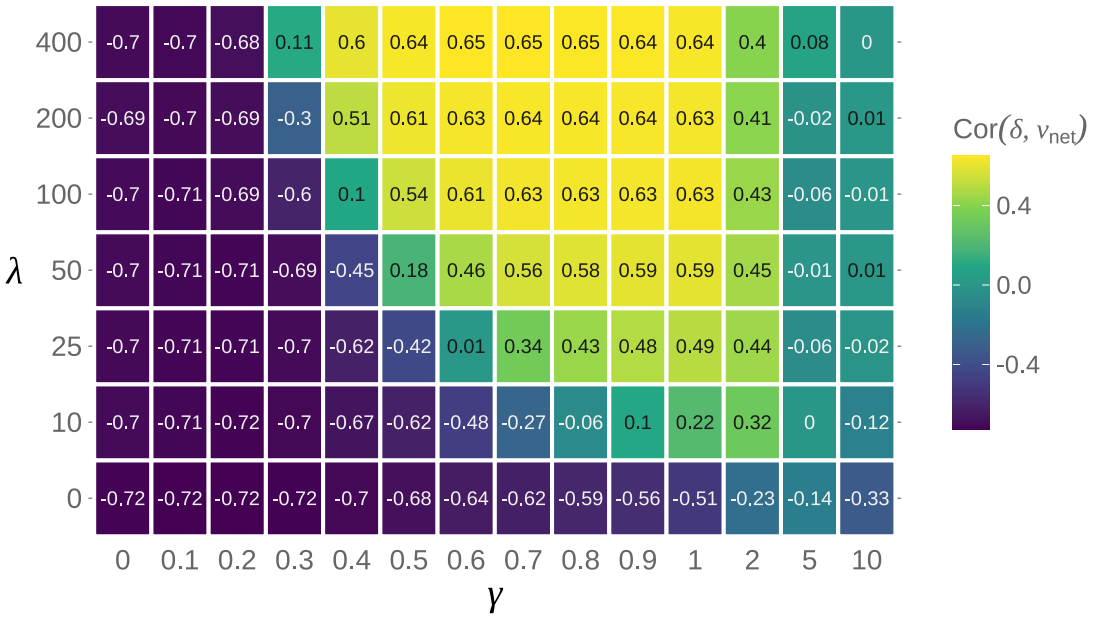
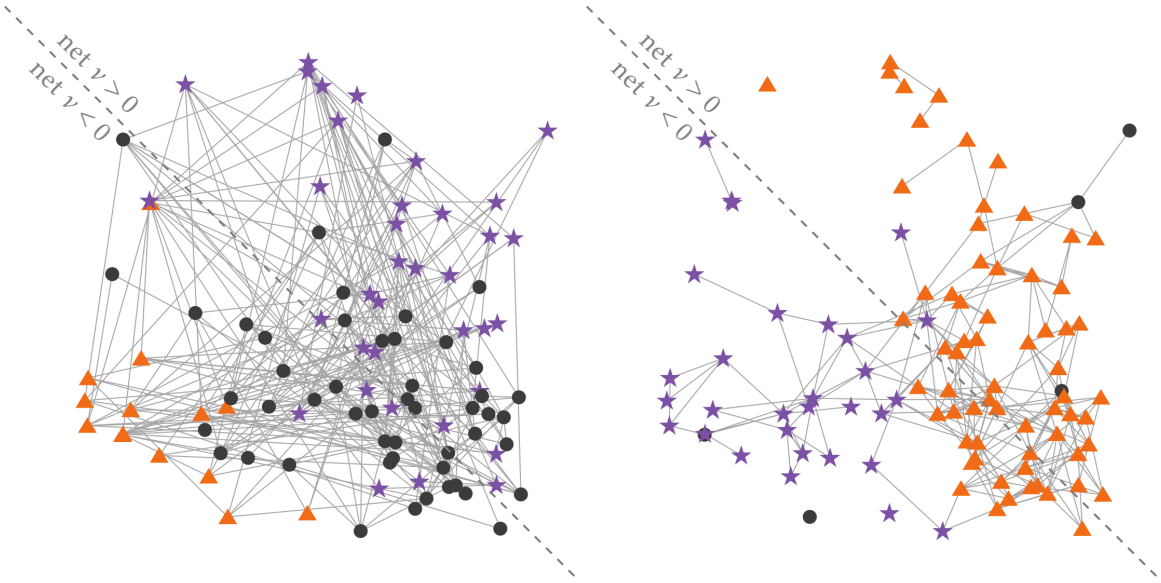


Figure 6: **Heatmap of Belief-value correlations:** Mean values for parameter combinations across $N = 200$ runs.

value shows that roughly 20% of the variation in polarization is linearly explained by these differences. Two effects coincide: First, higher R_0 mechanically increases openness to discrepant narratives by reducing their psychological discrepancy, given in Equation 7. Second, since $\delta_0 = 0.5$, the variation in R_0 is driven by the value profiles of neighbors. Thus, a more heterogenous neighborhood naturally exposes citizens to a wider array of values while simultaneously expanding the influence of discrepant attitudes.

Belief-value correlations and belief polarization. Over t, the model exerts correlations between citizens' beliefs and values. Since citizens' initial belief is $\delta_0 = 0.5$, this indicates an adaptation of beliefs according to their (fixed) value profiles over time. Barron et al. (2025) document positive belief-value correlations under motivated political reasoning, which would predict co-occurrence of optimistic (pessimistic) beliefs with liberal (conservative) net-values for our model. However, Figure 6 shows that the sign of these correlations depends on discounting intensity and the degree of homophily in the network. For moderate to high discounting ($0.4 \leq \gamma \leq 1$), positive correlations emerge across most network configurations, with higher λ reducing the threshold value of γ at which the phase transition from negative to positive occurs. Contrarily, given no or weak discounting with $\gamma < 0.3$, belief-value correlations are negative for all λ. Random networks without homophily ($\lambda = 0$) also exhibit negative belief-value correlations for all γ.

Figure 14 in the appendix plots per-run correlations for the $\lambda = 100$ case across the γ range. This run-level breakdown shows a pattern similar to one documented for polarization:

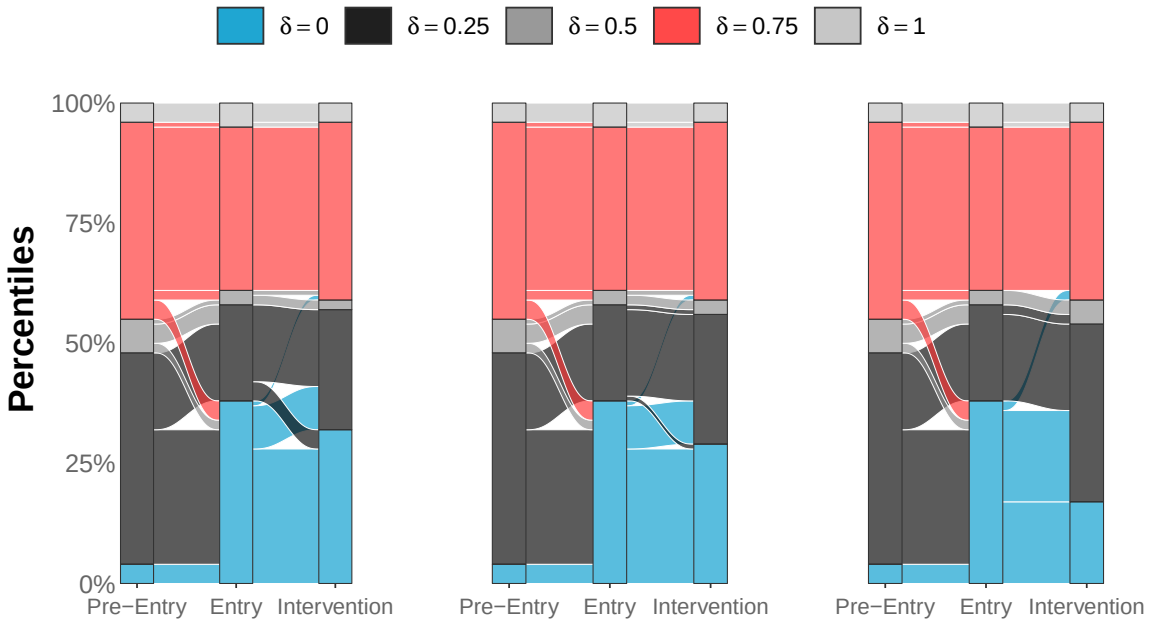


(a) $\lambda = 0$, $\gamma = 0.6$ (b) $\lambda = 100$, $\gamma = 0.6$

Figure 7: **Homophily induces positive belief-value correlations:** Two value-spaces of networks with identical citizen sets at $t = 4999$. Triangles represent optimistic beliefs $\delta_i \in \{0.75, 1\}$, stars are pessimistic $\delta_i \in \{0, 0.25\}$, gray dots represent $\delta_i = 0.5$.

While large negative and positive average correlations are stable, the cases where average correlation is close to 0 result from large between-run variation. For extreme discounting at $\gamma \geq 5$, this no average correlation result emerges in all homophilous networks while being subject to high variance. For $\gamma = 10$, between-run deviations from 0 are much less pronounced, indicating the overall lack of significant updating.

The phase transition in belief-value correlations can be explained by the belief-formation process: Since values are fixed, citizens adjust δ_i to minimize cognitive dissonance upon changing A_i. Because beliefs and values enter as complements in Equation 12, low-discounting citizens in diverse neighborhoods have to adjust δ_i counter to their values. Thus, the baseline tendency is a negative belief-value correlation. As shown in Figure 7, homophily counteracts this by exposing citizens predominantly to signals from value-aligned peers, while more restrictive discounting reduces the impact of distant attitudes. The interaction of these mechanisms shifts beliefs toward optimism for liberal citizens and toward pessimism for conservative ones. Sustained homophilous reinforcement therefore reconciles the model's predictions with the positive belief-value correlations documented empirically by Barron et al. (2025). When comparing the patterns emerging in Figure 4 and Figure 6, it also becomes obvious that belief-value correlations and attitude polarization are tightly interwoven phenomena: In parameter regimes where many individual citizens fluctuate and remain non-committal, their attitudes are not universally strong enough to translate into positive correlations. Once polarization clearly emerges, motivated reasoning becomes a mass phenomenon, causing beliefs to correlate positively with values.



(a) Pandering (b) Rebalancing (c) Opportunity framing

Figure 8: **Transition dynamics of beliefs** with $\lambda = 100$ and $\gamma = 0.6$. Beliefs are measured at $t = 4999$, $t = 5299$ and $t = 5600$.

The link between polarized attitudes and belief-value correlations suggests that beliefs may be polarized as well. Since a useful analysis of belief sorting requires some degree of stability of beliefs over time, we stick to restricting the analysis to $\lambda = 100$ and $\gamma = 0.6$. Given that beliefs follow attitudes, as specified in Equation 12, and given the moderate, mean-zero starting distribution of attitudes, polarized beliefs necessarily emerge in some form wherever attitudes are polarized. The pre-entry columns of Figure 8 display the distribution of beliefs at $t = 4999$ in a typical run that produces roughly symmetrical attitude polarization: Beliefs are indeed concentrated, as the majority of citizens is split almost evenly between $\delta_i = 0.25$ and $\delta_i = 0.75$. The neutral category and the two extreme beliefs remain marginalized. Thus, the population sorts into two internally cohesive camps of moderate optimists and moderate pessimists, each aligned with one of the mainstream *PI*s. This bimodal polarization pattern is distinct from radicalization, where citizens would concentrate in the fully optimistic and fully pessimistic belief categories.

Because beliefs act as weights on values, and values are moderate for most citizens and both mainstream *PI*s alike, the model's influence structure renders extreme beliefs rare. Most neutral-belief citizens, visualized in Figure 7 as gray dots, are located at the margins of the value distribution. Citizens with such outlier value profiles tend to retain non-partial belief, because they do not require developing strong beliefs to reconcile a highly partial attitude with their values.

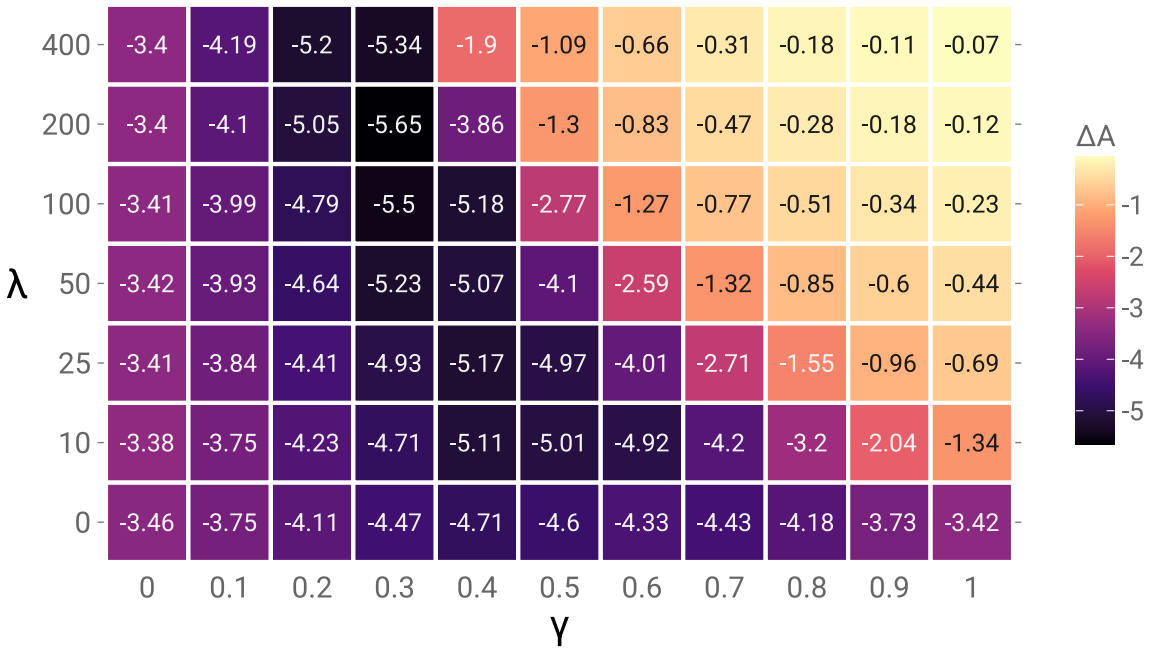


Figure 9: Attitude drop caused by the populist within 300 periods upon entry as a function of the discount factor γ, by homophily level λ.

3.2 The populist entry

A populist shock. In simulation period $t = 5000$, the populist enters the model and immediately imposes loss framing on the discourse such that citizens base their trust in PIs only on the closeness of ν^-. Figure 9 reports the average changes in attitude resulting from the populist's entry. Figure 10 reports changes in attitude polarization, measured as the percentage differences in SD(A), across combinations of (γ, λ).[7]

At low values of γ, citizens are highly susceptible to the new populist narrative, irrespectively of their own attitudes. Therefore, as shown on a micro-level in Figure 11a, the populist draws the entire population toward its position. The resulting mean attitude drop toward the populist peaks within a critical region around $\gamma \approx 0.3$, where ΔA reaches -5.65 for $\lambda = 200$. This is a shift of more than 25% of the full attitude range modeled. Below $\gamma = 0.3$, attitude displacement rises as γ increases. Above it, the collapse of mean attitudes attenuates strongly. The location of the peak in ΔA is not fixed, but shifts toward lower discounting values as λ increases. Additionally, in highly homophilous populations, mean attitudes recover within a narrow band of γ. In networks with less homophily, they rise more uniformly in γ.

Differences in polarization reflect these changes: The attitude drift caused by the populist lessens the influence of the progressive PI, who's attitude is now farther removed from citizens' attitudes, leaving the mean attitude in between those of the conservative and

[7] Starting with this analysis, we truncate γ above 1.

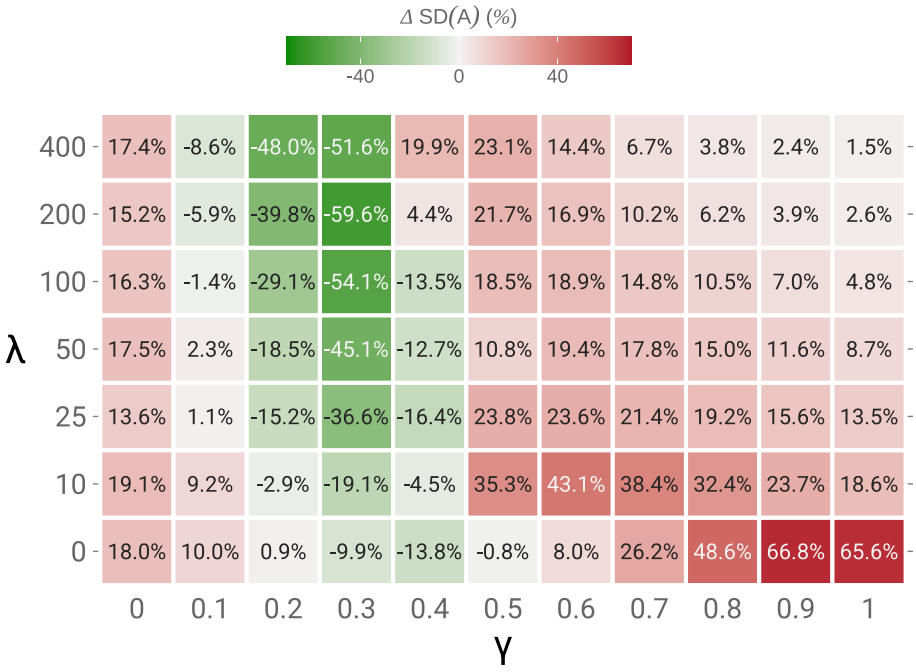
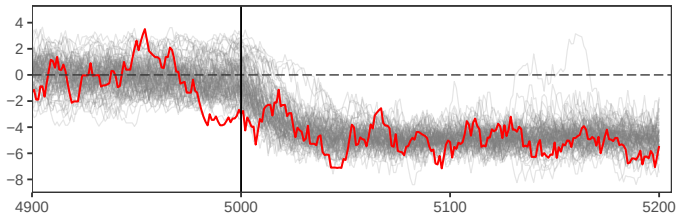


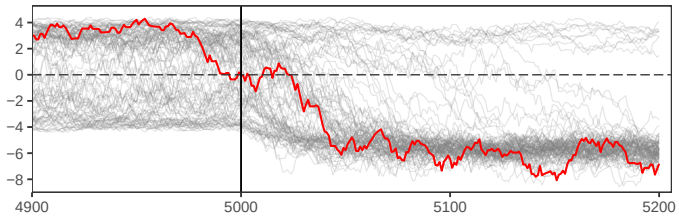
Figure 10: **Polarization effect of the populist** as a function of the discount factor γ, by homophily level λ.

populist PIs. The universal displacement of attitudes compresses its distribution and produces a substantial decline in SD(A) of up to −59.6%, on average, for the same $\lambda = 200$ scenario where $|\Delta A|$ peaks. This compression intensifies as a function of γ, reflecting the self-reinforcing reduction in influence of the progressive PI. Beyond the critical point around $\gamma = 0.3$, the populist's initial influence on mean attitudes starts subsiding, making polarization a non-monotonic function of γ: Figure 11 shows an increasing number of citizens becoming resistant or immune to the populist narrative, causing them to remain at high A_i as γ increases. As a side-effect, this increased resilience to populist narratives increases polarization, as the populist near-consensus gives way to a bipolar distribution. Increasing γ beyond this critical point reduces the post-entry gain in polarization, while also reducing the drop in mean attitudes for low-A_i citizens. For $\gamma \geq 1$, this results in nearly unchanged distributions post-entry.

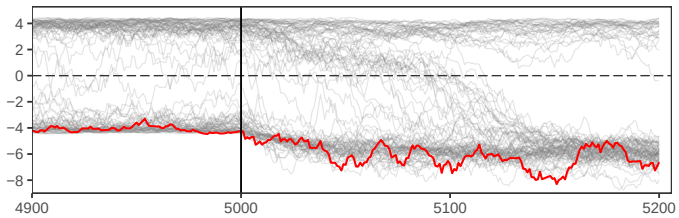
Considering the entry column of Figure 8, the populist naturally causes a large swing towards fully pessimistic beliefs at moderate discounting and homophily levels. Post-entry, the $\delta_i = 0$ category accounts for almost 50% of all citizen beliefs in the representative run. Transition into this category is primarily fed by the adjacent moderately pessimistic category, while a smaller but non-negligible share draws on formerly neutral and moderately optimistic citizens. In line with the stark increase in attitude polarization observed for this model configuration in Figure 10, transition from the moderate optimistic category either feeds $\delta = 0$ or $\delta = 0.5$, with no transition occurring to moderately pessimistic beliefs, reflecting a bipolar attitude pattern without uncommitted citizens.



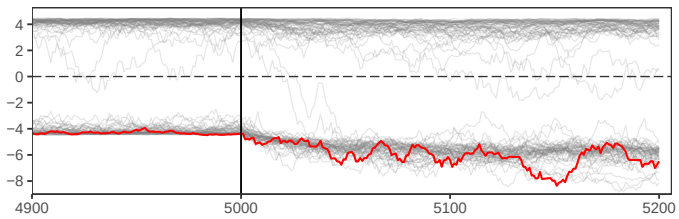
(a) $\gamma = 0.2$



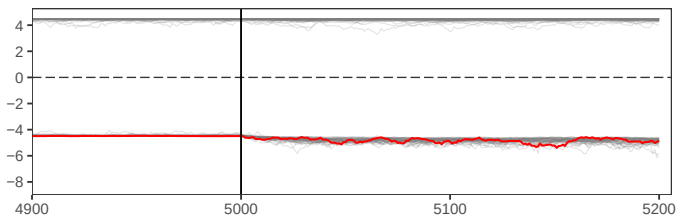
(b) $\gamma = 0.4$



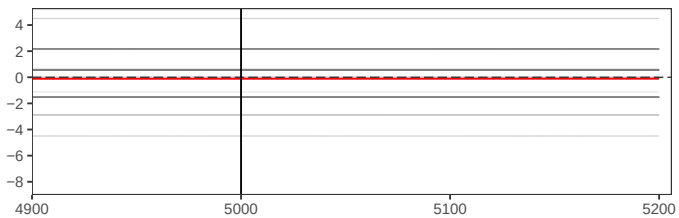
(c) $\gamma = 0.5$



(d) $\gamma = 0.6$



(e) $\gamma = 1$



(f) $\gamma = 10$

Figure 11: **Individual attitude trajectories of citizens for** 300 **periods post populist entry with** $\lambda = 100$ for selected values of γ. Red trajectory marks an example citizens. The latter and the random seed are identical across configurations.

3.3 Interventions

In Figure 12, we analyze the impacts of the three intervention strategies on mean attitudes and attitude polarization. For each λ, the top row of the graphic displays the recovery of mean attitudes each strategy induces across the γ-range, relative to the no-intervention baseline. For example: a reading of 10% would mean that an average of 10% of the populist-induced attitude drop that would have prevailed without intervention was prevented by the strategy. All differences are measured between $t = 4999$ and $t = 5600$, from right before the entry to 300 periods after the intervention. [8]

Across all γ values, the most effective intervention with respect to reversing the attitude drop brought about by the populist is opportunity framing. Its recovery curves stay positive for all λ, with a large peak that gets more pronounced in λ. With moderate to large homophily and $\gamma \leq 0.3$, recovery is almost uniformly between 10% and 20%, then increases sharply within a narrow γ range, before slightly tapering off at high γ values. Peak recovery under opportunity framing occurs at $\gamma = 0.5$ and exceeds a magnitude of 40%, which is roughly three times the maximum of either alternative strategy. Homophily strengthens this effect, as higher λ shifts recovery upward across the mid-to-high γ region, with the largest gains concentrated near $\gamma \in [0.5, 0.6]$. With little and no homophily, the curve is smoother, which is true for all strategies.

The other strategies are markedly less effective with regards to limiting the attitude drop. Rebalancing produces uniformly smaller recovery of 0 to a little over 10%. In contrast to what we see with opportunity framing, homophily has little moderating role in shaping the peak amplitude of recovery, but only shifts its location to smaller γ as λ increases.

Pandering yields a more complex pattern: At low-to-mid γ values below 0.4, recovery is 0 for all λ, before rising. The shape of this increase in recovery in γ is heavily dependent on λ: For higher values, the peak is sudden, while turning more gradual at lower homophily levels. Similar to opportunity framing, the most pronounced peak occurs for maximum $\lambda = 400$ with a mean of roughly 15% at $\gamma = 0.4$. For this network configuration, the strategy offsets the populist's impact to a slightly larger degree than rebalancing. However, in high homophily setups, recovery deteriorates sharply as γ increases beyond that peak and reverses signs. This sign-switch results in an almost symmetric curve around 0-recovery: at $\gamma \geq 0.7$, the highest-λ runs exhibit a mean deterioration of more than 10%, meaning the intervention increases the average drop in attitudes relative to doing nothing.

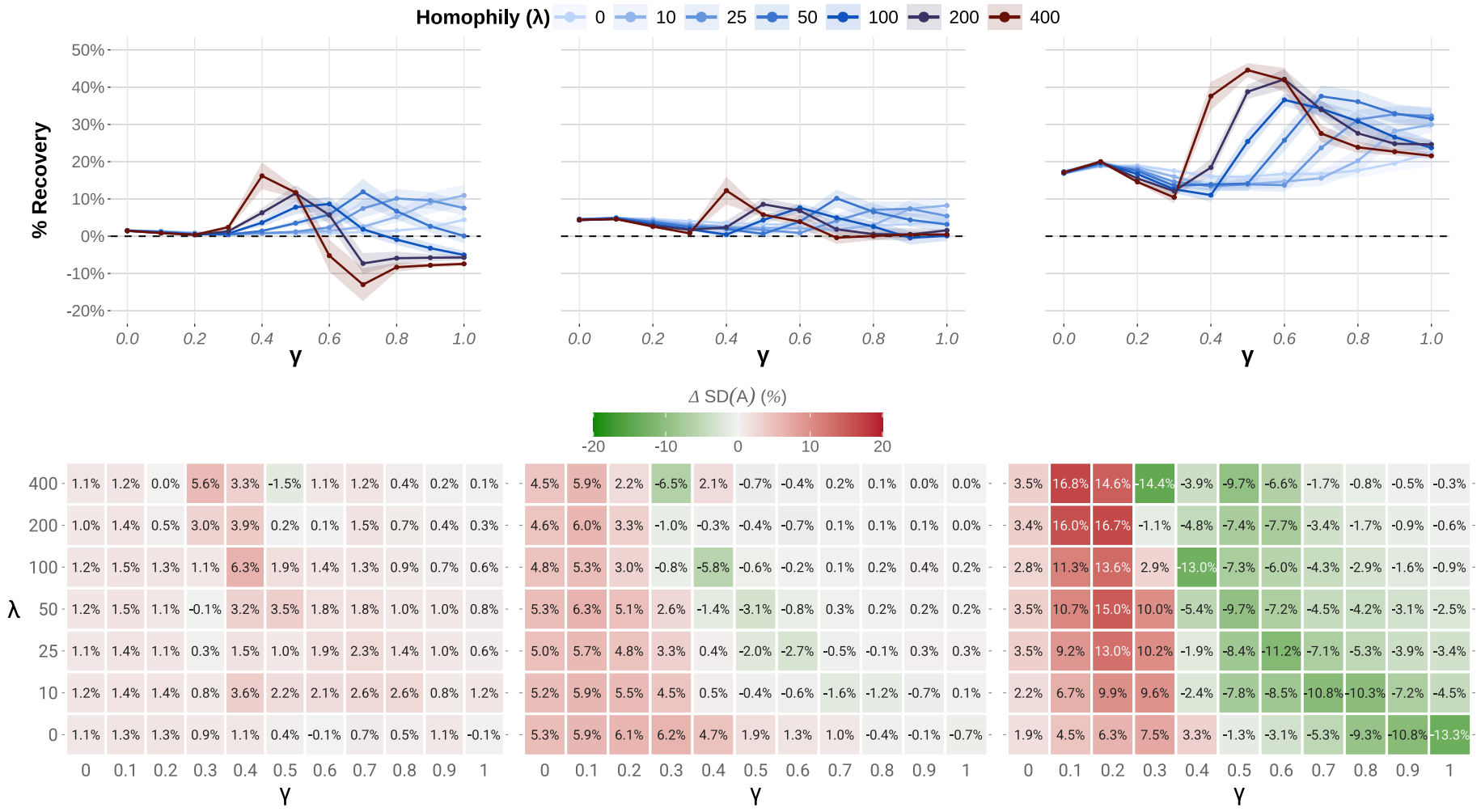
Analogously, the bottom row of Figure 12 shows the effects of the strategies on attitude polarization relative to the no-intervention baseline. We use the $\lambda = 400$ case as an

[8] For each intervention and parameter combination, we regress the normalized attitude displacement of treated runs on a binary treatment indicator using OLS, clustered on random seeds. Shaded bands in Figure 12 indicate 95% confidence intervals based on HC1 cluster-robust standard errors.

example, where attitude recovery is the most pronounced: Under opportunity framing, polarization rises for low γ, where effects on mean attitudes are only moderate. As γ increases, a sign switch to lower polarization occurs at $\gamma = 0.3$, *before* the strategy peaks in terms of attitude recovery. Notably, at peak recovery, with $\gamma = 0.5$, a large degree of attitude recovery coincides with lowered polarization. This pattern is similar for all homophilous networks with at least moderate discounting, indicating that opportunity framing mitigates the populist's attitude effect on a critical share of citizens.

For the pandering and rebalancing strategies, mean-attitude recovery and significant polarization recovery do not coincide. Across the parameter grid, pandering is neutral to slightly aggravating regarding polarization, while rebalancing displays a qualitatively similar, but significantly weaker version of a pattern discussed for opportunity framing: Its peak in polarization-reduction indicates around half of the effect-size documented for opportunity framing.

Belief transitions. The intervention column of Figure 8 allows for a comparison of the belief dynamics induced by the strategies for $\lambda = 100$ and $\gamma = 0.6$. In this general range of the parameter space, results are rather stable, causing qualitatively similar belief dynamics between runs. Slight deviations are the result of by pre-entry differences. The three strategies affect beliefs mostly in the range below $\delta = 0.5$ with neither strategy causing large changes in the optimistic categories. Under pandering, exchange between $\delta = 0$ and $\delta = 0.25$ is bilateral with a surplus of citizens returning to moderately pessimistic belief. This net loss in the fully pessimistic camp corroborates Figure 12, where attitude recovery is slightly positive for the given parameter configuration. While rebalancing presents a similar dynamic with a slightly reduced transitions towards $\delta = 0$, opportunity framing is drastically different. Under opportunity framing, the fully pessimistic belief category is reduced by over half in the given network. While a small but non-negligible portion of pessimists switches to moderate optimism, the majority of transitions occur into the moderately pessimistic group. As a net effect, while pandering and rebalancing largely fail to affect the belief distribution, opportunity framing restores the moderately pessimistic belief-category, as the number of citizens holding this belief now roughly corresponds to the number of moderate optimists. While the reduction of influence exhibited by the populist does not translate to a full restoration of the pre-entry belief distribution, the pre-shock symmetry between moderate pessimists and moderate optimists is re-attained, relegating fully pessimistic belief to little over 10% of citizens.



(a) Pandering (b) rebalancing (c) Opportunity framing

Figure 12: **Impact of intervention strategies on mean attitudes and polarization.** Columns are strategies. The top row shows average recovery of $\overline{A}$, the bottom row shows the percentage Δ in $\overline{\mathrm{SD}}(A)$, both vs. the no-intervention counterfactual. 50 runs per cell, paired on random-seeds.

4 Discussion

The model's strength lies in jointly accounting for two phenomena that prior frameworks have largely treated separately. First, it provides a detailed, mechanistic account of how diverging political narratives translate into the polarization of attitudes and beliefs. Polarization emerges endogenously from the interaction of narrative discounting, homophily, and motivated reasoning. Second, the model offers a mechanism for the formation of belief-value correlations, a pattern documented empirically by Barron et al. (2025). In the model, this correlation arises as a byproduct of citizens reconciling evolving attitudes with fixed values through motivated reasoning. Homophilous exposure determines whether this reconciliation pulls beliefs into alignment with values.

While we make detailed assumptions on the inner structure of narratives and their effects on attitudes, several aspects of narrative dynamics fall outside the scope of the present model and are left for future research. First, we assume constant and uniformly distributed exposure of citizens to all political initiators without allowing citizens to select their political diets in any way. In a related assumption, we abstract from media as intermediaries between political actors and the public in narrative transmission. While the homophilous neighborhoods in our model do entail self-reinforcing loops, future work might consider echo-chambers in the context of mass media and social media. Both media types act as a second-order supplier of economic policy narratives (Eliaz & Spiegler, 2024) that can either amplify or undermine a political agenda. We also deliberately leave out personality traits of political actors as affecting trust, focussing entirely on the policy content of economic policy narratives.

The outcome space ascribed to narratives in our modeling framework is a construction that reflects the discourse rather than empirical facts. For example, in this paper, we assume that citizens have a single, overall attitude towards immigration. While "irregular" migration and qualified immigration of workers are certainly distinct issues, various spillover effects are documented in the literature that justify modeling immigration attitudes as one-dimensional: First, the public tends perceive immigrants as one homogeneous out-group, not differentiating between different groups or skill levels (Alesina et al., 2022; Blinder, 2013). Second, xenophobic sentiment and policies directed at limiting migration are known to deter all groups of potential migrants, including skilled workers, from immigrating (Docquier & Rapoport, 2025; Slotwinski & Stutzer, 2019). Lastly, the opposition that is often made between skilled and irregular migrants as disjoint sets is, on the whole, invalid, and causes adverse selection (Coniglio et al., 2006; Düvell, 2011; Purkayastha et al., 2023). These features of public immigration attitudes also bear directly on the pandering strategy. A political actor who panders may rationalize doing so by invoking the distinction between irregular and skilled migration and framing a crackdown on the former as compatible with attracting the latter. Since this distinction does not resonate in public perception, pandering shifts overall

immigration attitudes rather than selectively targeting irregular flows.

Immigration policy serves as a useful example for laying out our model. However, we believe that it is portable to many policy domains. The basic requirement for applying our modeling logic to a policy is for it to have desirable main-effects and potentially undesirable side-effects, resulting in a net effect that is apriori unclear and thus amenable to various narratives. We believe this applies to most if not all facets of economic policy.

5 Conclusion

We develop an agent-based model of attitude and belief formation around immigration policy. Two mainstream political initiators (PIs), representing a center-right to center-left ideological spectrum, disseminate competing narratives to sway the policy attitudes of a network of politically interested citizens. Citizens receive and exchange these narratives among themselves, discounting incoming attitudes according to the PDD model (Fink et al., 1983), and aligning their beliefs regarding the policy's impact with their evolving attitudes through motivated reasoning.

In this first phase of the simulation, attitude polarization emerges in the network through a phase transition: When discounting is weak, citizen attitudes are highly volatile but not polarized. Around a critical discounting value, citizen attitudes start clustering around PI attitudes and polarization emerges. Beliefs follow attitudes, so that citizens, at the critical value, adopt value-aligned beliefs that correspond to those held by the PIs. The resulting belief polarization is bimodal but not extreme, while value-based homophily in the network lowers the polarization threshold.

We subject the model to a political shock through the entry of a populist PI who advocates a one-sided, negative immigration narrative that denies the economic potential of immigration. The populist changes the discourse by imposing his loss frame on all PI-citizen interaction and thereby pulls citizens toward its position. The resulting attitude drop peaks at intermediate discounting intensities, where citizens are not fully anchored to the mainstream PIs and susceptible to the populist narrative. In regimes where the populist succeeds in moving the mass of citizens, the attitude distribution compresses and polarization decreases. In regimes with high discounting intensity or a high degree of homophily, high-attitude citizens are shielded from the populist's influence and uphold their attitudes, resulting in increased attitude polarization. Fringe but non-negligible populist influence therefore results in high degrees of polarization. On the belief level, the populist triggers a large-scale shift toward fully pessimistic beliefs, more than halving the share of moderate pessimists in the population for most parameter regimes.

We implement three intervention strategies initiated by one of or both mainstream PIs.

All are designed to shift which policy outcome is salient when citizens evaluate *PI* narratives. The choice of strategy is decisive in mitigating the populist's impact: Pandering entails the center-right *PI* adopting the rhetoric of the populist's pessimistic belief. It is a high-risk, low reward strategy: Under strong homophily and moderate-to-weak discounting, it partially recovers mean attitudes and beliefs. At only moderately increased discounting intensity, it deepens the attitude drop while staying neutral regarding polarization. Rebalancing entails restoring balanced outcome salience in *PI*-citizen interactions, thereby undoing the loss frame imposed by the populist. This reestablished balance recovers only a small share of the populist-induced attitude drop across all parameter configurations. Opportunity framing counters the populist's loss frame by swinging outcome salience in the opposite direction, establishing the labor market potential of immigration as the dominant value dimension for citizens. It dominates both alternatives, recovering up to 40% of the populist-induced attitude drop, while lowering polarization.

The results carry a clear strategic implication: In a discursive environment shaped by populist loss framing, capitulation and pandering accelerate the erosion that it tries to contain. The only effective counter is a concerted effort to shift salience toward the policy's gains, which the populist narrative suppresses.

References

Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the rise of the nazis in prewar germany. *The Quarterly Journal of Economics*, *130*(4), 1885–1939. https://doi.org/10.1093/qje/qjv030

Alesina, A., & Giuliano, P. (2011). Preferences for redistribution. In *Handbook of social economics* (pp. 93–131). Elsevier. https://doi.org/10.1016/b978-0-444-53187-2.00004-8

Alesina, A., Miano, A., & Stantcheva, S. (2022). Immigration and redistribution. *The Review of Economic Studies*, *90*(1), 1–39. https://doi.org/10.1093/restud/rdac011

Alesina, A., Stantcheva, S., & Teso, E. (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review*, *108*(2), 521–554. https://doi.org/10.1257/aer.20162015

Armona, L., Fuster, A., & Zafar, B. (2018). Home price expectations and behaviour: Evidence from a randomized information experiment. *The Review of Economic Studies*, *86*(4), 1371–1410. https://doi.org/10.1093/restud/rdy038

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barron, K., Becker, A., & Huck, S. (2025). Motivated political reasoning: On the emergence of belief-value constellations. *European Economic Review*, *172*, 104929. https://doi.org/10.1016/j.euroecorev.2024.104929

Bergman, M. M. (1998). A theoretical note on the differences between attitudes, opinions, and values. *Swiss Political Science Review*, *4*(2), 81–93.

Blinder, S. (2013). Imagined immigration: The impact of different meanings of 'immigrants' in public opinion and policy debates in britain. *Political Studies*, *63*(1), 80–100. https://doi.org/10.1111/1467-9248.12053

Bonikowski, B., & Gidron, N. (2015). The populist style in american politics: Presidential campaign discourse, 1952–1996. *Social Forces*, *94*(4), 1593–1621. https://doi.org/10.1093/sf/sov120

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2022). Memory and probability. *The Quarterly Journal of Economics*, *138*(1), 265–311. https://doi.org/10.1093/qje/qjac031

Bracciale, R., & Martella, A. (2017). Define the populist political communication style: The case of italian political leaders on twitter. *Information, Communication & Society*, *20*(9), 1310–1329. https://doi.org/10.1080/1369118X.2017.1328522

Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., Flocken, C., & Holman, B. (2017). Understanding polarization: Meanings, measures, and model evaluation. *Philosophy of science*, *84*(1), 115–159.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*(6282), 220–224. https://doi.org/10.1126/science.aad9713

Brown, K., Mondon, A., & Winter, A. (2023). The far right, the mainstream and mainstreaming: Towards a heuristic framework. *Journal of Political Ideologies*, *28*(2), 162–179.

Cammaerts, B. (2018). The mainstreaming of extreme right-wing populism in the low countries: What is to be done? *Communication Culture & Critique*, *11*, 7–20.

Cassell, K. J. (2023). The populist communication strategy in comparative perspective. *The International Journal of Press/Politics*, *28*(4), 725–746. https://doi.org/10.1177/19401612211055695

Cattaneo, C., & Grieco, D. (2021). Turning opposition into support to immigration: The role of narratives. *Journal of Economic Behavior amp; Organization*, *190*, 785–801. https://doi.org/10.1016/j.jebo.2021.08.015

Chung, S., & Fink, E. L. (2022). Mathematical models of message discrepancy: Previous models and a modified psychological discounting model. *Communication Theory*, *32*(4), 471–487. https://doi.org/10.1093/ct/qtac010

Cialdini, R. B. (2021). *Influence: The psychology of persuasion* (New and Expanded) [Literaturverzeichnis: Seite 491-543]. Harper Business.

Coniglio, N. D., Arcangelis, G. D., & Serlenga, L. (2006). *Intentions to return of undocumented migrants: Illegality as a cause of skill waste* (tech. rep. No. 2356). IZA Discussion Papers.

de Clippel, G., & Zhang, X. (2022). Non-bayesian persuasion. *Journal of Political Economy*, *130*(10), 2594–2642. https://doi.org/10.1086/720464

Degroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121. https://doi.org/10.1080/01621459.1974.10480137

DellaVigna, S., & Ferrara, E. L. (2015, July). *Economic and social impacts of the media* (tech. rep.). National Bureau of Economic Research. https://doi.org/10.3386/w21360

Dinas, E., Matakos, K., Xefteris, D., & Hangartner, D. (2019). Waking up the golden dawn: Does exposure to the refugee crisis increase support for extreme-right parties? *Political Analysis*, *27*(2), 244–254. https://doi.org/10.1017/pan.2018.48

Docquier, F., & Rapoport, H. (2025). The vicious circle of xenophobia: Immigration and right wing populism. *Economic Policy*, *40*(122), 551–573. https://doi.org/10.1093/epolic/eiaf001

Dornbusch, R., & Edwards, S. (1991). The macroeconomics of populism. In R. Dornbusch (Ed.). S. Edwards (Ed.), *The macroeconomics of populism in latin america*. University of Chicago Press. https://doi.org/10.7208/9780226158488-004

Downs, A. (1957). *An economic theory of democracy*. Harper Row.

Düvell, F. (2011). Paths into irregularity: The legal and political construction of irregular migration. *European Journal of Migration and Law*, *13*(3), 275–295. https://doi.org/10.1163/157181611x587856

Edo, A., Giesing, Y., Öztunc, J., & Poutvaara, P. (2019). Immigration and electoral support for the far-left and the far-right. *European Economic Review*, *115*, 99–143. https://doi.org/10.1016/j.euroecorev.2019.03.001

Ekman, M., & Krzyzanowski, M. (2018). A populist turn? news editorials and the recent discursive shift on immigration in sweden. *Nordicom Review*, *42*(S1), 67–87.

Eliaz, K., Galperti, S., & Spiegler, R. (2024). False narratives and political mobilization. *Journal of the European Economic Association*. https://doi.org/10.1093/jeea/jvae047

Eliaz, K., & Spiegler, R. (2020). A model of competing narratives. *American Economic Review*, *110*(12), 3786–3816. https://doi.org/10.1257/aer.20191099

Eliaz, K., & Spiegler, R. (2024, April 16). *News media as suppliers of narratives (and information)* (tech. rep.).

Enders, A. M., & Armaly, M. T. (2018). The differential effects of actual and perceived polarization. *Political Behavior*, *41*(3), 815–839. https://doi.org/10.1007/s11109-018-9476-2

Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, *30*(3), 133–140. https://doi.org/10.1257/jep.30.3.133

Feather, N. T. (Ed.). (1982). *Expectations and actions: Expectancy-value models in psychology* [Literaturangaben]. Lawrence Erlbaum.

Fink, E. L., Kaplowitz, S. A., & Bauer, C. L. (1983). Positional discrepancy, psychological discrepancy, and attitude change: Experimental tests of some mathematical models1. *Communication Monographs*, *50*(4), 413–430. https://doi.org/10.1080/03637758309390178

Funke, M., Schularick, M., & Trebesch, C. (2023). Populist leaders and the economy. *American Economic Review*, *113*(12), 3249–3288. https://doi.org/10.1257/aer.20202045

Galston, W. A. (2018). The populist challenge to liberal democracy. *Journal of Democracy*, *29*(2), 5–19. https://doi.org/10.1353/jod.2018.0020

Grabisch, M., & Rusinowska, A. (2020). A survey on nonstrategic models of opinion dynamics. *Games*, *11*(4), 1–29.

Haaland, I., Roth, C., & Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, *61*(1), 3–40. https://doi.org/10.1257/jel.20211658

Haas, M. (2018). *Why democracies flounder and fail: Remedying mass society politics*. Springer.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 107–112. https://doi.org/10.1016/s0022-5371(77)80012-1

Hassani, H., Razavi-Far, R., Saif, M., Chiclana, F., Krejcar, O., & Herrera-Viedma, E. (2022). Classical dynamic consensus and opinion dynamics models: A survey of recent trends and methodologies. *Information Fusion*, *88*, 22–40.

Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, *5*(3).

Iijima, R., & Kamada, Y. (2017). Social distance and network structures: Social distance and network structures. *Theoretical Economics*, *12*(2), 655–689. https://doi.org/10.3982/te1873

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jagers, J., & Walgrave, S. (2007). Populism as political communication style: An empirical study of political parties' discourse in belgium. *European Journal of Political Research*, *46*, 319–345. https://doi.org/10.1111/j.1475-6765.2006.00699.x

Jürgens, P., Jungherr, A., & Schoen, H. (2011). Small worlds with a difference: New gatekeepers and the filtering of political information on twitter. *Proceedings of the 3rd International Web Science Conference*, 1–5. https://doi.org/10.1145/2527031.2527034

Klinger, U., Koc-Michalska, K., & Russmann, U. (2022). Are campaigns getting uglier, and who is to blame? negativity, dramatization and populism on facebook in the 2014 and 2019 ep election campaigns. *Political Communication*, *40*(3), 263–282. https://doi.org/10.1080/10584609.2022.2133198

Krzyzanowski, M. (2020). Discursive shifts and the normalisation of racism: Imaginaries of immigration, moral panics and the discourse of contemporary right-wing populism. *Social Semiotics*, *30*(4), 503–527.

Krzyzanowski, M., & Ekström, H. (2024). "no longer the haven of tolerance?" the press and discursive shifts on immigration in sweden 2010 - 2022. *Social Semiotics*, *35*(3), 472–494.

Krzyzanowski, M., & Ekström, M. (2022). The normalization of far-right populism and nativist authoritarianism: Discursive practices in media, journalism and the wider public sphere. *Discourse & Society*, *33*(6), 719–729.

Kunda, Z. (1990). The case of motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Lau, R. R., Smith, R. A., & Fiske, S. T. (1991). Political beliefs, policy interpretations, and political persuasion. *The Journal of Politics*, *53*(3), 644–675. https://doi.org/10.2307/2131574

Lee, A. H.-Y. (2022). Social trust in polarized times: How perceptions of political polarization affect americans' trust in each other. *Political Behavior*, *44*(3), 1533–1554. https://doi.org/10.1007/s11109-022-09787-1

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, G., Sun, Z., & Parker, D. C. (2015). The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation*, *18*(4). https://doi.org/10.18564/jasss.2897

Lee, N., Nyhan, B., Reifler, J., & Flynn, J. (2021). More accurate, but no less polarized: Comparing the factual beliefs of government officias and the public. *British Journal of Political Science*, *51*(3), 1315–1322.

Manucci, L., & Weber, E. (2017). Why the big picture matters: Political and media populism in western europe since the 1970s. *Swiss Political Science Review*, *23*(4), 313–334.

Martin, G. J., & Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, *107*(9), 2565–2599. https://doi.org/10.1257/aer.20160812

Mudde, C., & Rovira Kaltwasser, C. (2012). Exclusionary vs. inclusionary populism: Comparing contemporary europe and latin america. *Government and Opposition*, *48*(2), 147–174. https://doi.org/10.1017/gov.2012.11

Mudde, C., & Rovira Kaltwasser, C. (2017). *Populism: A very short introduction*. Oxford University Press. https://doi.org/10.1093/actrade/9780190234874.001.0001

Müller, J.-W. (2016, January). *What is populism?* University of Pennsylvania Press. https://doi.org/10.9783/9780812293784

Nai, A. (2021). Fear and loathing in populist campaigns? comparing the communication style of populists and non-populists in elections worldwide. *Journal of Political Marketing*, *20*(2), 219–250. https://doi.org/10.1080/15377857.2018.1491439

Noorazar, H. (2020). Recent advances in opinion propagation dynamics: A 2020 survey. *The European Physical Journal Plus*, *135*, 521.

Obradović, S., Power, S. A., & Sheehy-Skeffington, J. (2020). Understanding the psychological appeal of populism. *Current Opinion in Psychology*, *35*, 125–131. https://doi.org/10.1016/j.copsyc.2020.06.009

Purkayastha, D., Bircan, T., Yar, A. W. A., & Ceylan, D. (2023). Irregular migration is skilled migration: Reimagining skill in eu's migration policies. *Humanities and*

Social Sciences Communications, *10*(1). https://doi.org/10.1057/s41599-023-01885-7

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rekker, R. (2022). Political polarization over factual beliefs. In *Knowledge resistance in high-choice information environments* (pp. 222–236). Routledge.

Rekker, R. (2024). Factual belief polarization between democrats and republicans: Source or epiphenomenon of ideological and affective polarization? *Frontiers in Political Science*, *6*. https://doi.org/10.3389/fpos.2024.1254826

Rekker, R. (2025). A four-level model of political polarization over science: Evidence from 10 european countries. *Public Understanding of Science*, *34*(4), 424–445. https://doi.org/10.1177/09636625241306352

Rekker, R., & Harteveld, E. (2022). Understanding factual belief polarization: The role of trust, political sophistication, and affective polarization. *Acta Politica*, *59*(3), 643–670. https://doi.org/10.1057/s41269-022-00265-4

Roe, E. (1994, November). *Narrative policy analysis: Theory and practice*. Duke University Press. https://doi.org/10.1515/9780822381891

Roos, M., & Reccius, M. (2023). Narratives in economics. *Journal of Economic Surveys*, *38*(2), 303–341. https://doi.org/https://doi.org/10.1111/joes.12576

Roth, C., & Wohlfart, J. (2020). How do expectations about the macroeconomy affect personal expectations and behavior? *The Review of Economics and Statistics*, *102*(4), 731–748. https://doi.org/10.1162/rest_a_00867

Ruth-Lovell, S. P., & Grahn, S. (2022). Threat or corrective to democracy? the relationship between populism and different models of democracy. *European Journal of Political Research*, *62*(3), 677–698. https://doi.org/10.1111/1475-6765.12564

Santos, F. P., Lelkes, Y., & Levin, S. A. (2021). Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences*, *118*(50). https://doi.org/10.1073/pnas.2102141118

Schwörer, J. (2021). *The growth of populism in the political mainstream: The contagion effect of populist messages on mainstream parties' communication*. Springer.

Simons, H. W., Berkowitz, N. N., & Moyer, R. J. (1970). Similarity, credibility, and attitude change: A review and a theory. *Psychological Bulletin*, *73*(1), 1–16. https://doi.org/10.1037/h0028429

Slotwinski, M., & Stutzer, A. (2019). The deterrent effect of an anti-minaret vote on foreigners' location choices. *Journal of Population Economics*, *32*(3), 1043–1095. https://doi.org/10.1007/s00148-019-00729-6

Stecula, D. A., & Pickup, M. (2021). How populism and conservative media fuel conspiracy beliefs about covid-19 and what it means for covid-19 behaviors. *Research amp; Politics*, *8*(1). https://doi.org/10.1177/2053168021993979

Steiglechner, P., Smaldino, P. E., & Merico, A. (2025). How opinion variation among in-groups can skew perceptions of ideological polarization (A. Szolnoki, Ed.). *PNAS Nexus*, *4*(7). https://doi.org/10.1093/pnasnexus/pgaf184

Steinmayr, A. (2021). Contact versus exposure: Refugee presence and voting for the far right. *The Review of Economics and Statistics*, *103*(2), 310–327. https://doi.org/10.1162/rest_a_00922

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*.

Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics*, *16*(2), 1–38. https://doi.org/10.1257/mic.20220146

Udry, J., & Barber, S. J. (2024). The illusory truth effect: A review of how repetition increases belief in misinformation. *Current Opinion in Psychology*, *56*, 101736. https://doi.org/10.1016/j.copsyc.2023.101736

Uzzi, B., Amaral, L. A. N., & Reed-Tsochas, F. (2007). Small-world networks and management science research: A review. *European Management Review*, *4*(2), 77–91. https://doi.org/10.1057/palgrave.emr.1500078

van der Linden, S. (2021). The gateway belief model (gbm): A review and research agenda for communicating the scientific consensus on climate change. *Current Opinion in Psychology*, *42*, 7–12. https://doi.org/10.1016/j.copsyc.2021.01.005

Wänke, T., & Vogel, M. (2016). *Attitudes and attitude change*. Psychology Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Wilensky, U. (1999). NetLogo [http : / / ccl . northwestern . edu / netlogo/]. http://ccl.northwestern.edu/netlogo/

Wood, T., & Porter, E. (2018). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y

6 Appendix

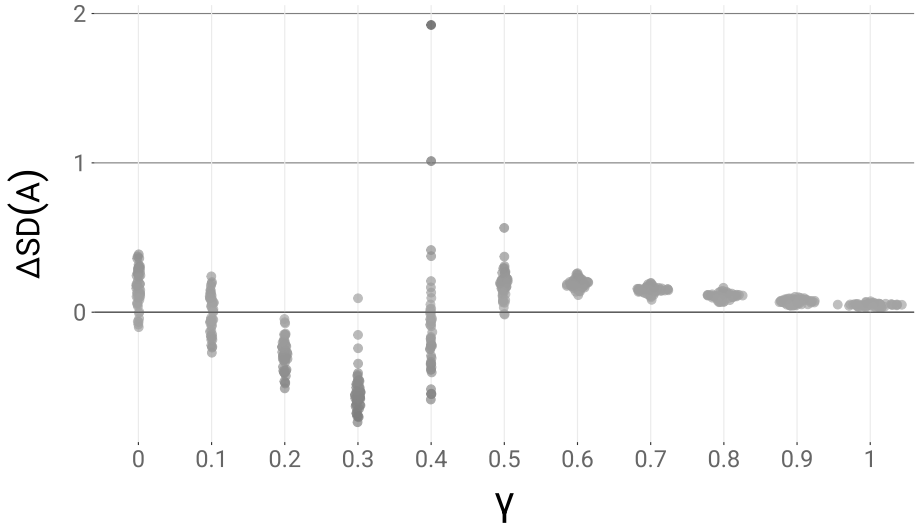


Figure 13: **Change of SD(A) due to the populist entry for** $\lambda = 100$**:** Delta is computed between $t = 4999$ and $t = 5299$, spanning the period just before the entry to immediately prior to interventions. A y-axis reading of 1 implies a 100% increase.

Table 2: Network statistics as possible determinants of attitude polarization ($SD(A)$)

	Dependent variable:		
	$SD(A)$		
	(1)	(2)	(3)
Clustering coefficient	3.187		
	(3.842)		
APL		0.439	
		(0.494)	
$\overline{R_{r0}}$			-4.675^{**}
			(1.703)
Constant	2.109^{**}	1.141	9.568^{***}
	(0.773)	(1.826)	(2.475)
Observations	50	50	50
R^2	0.018	0.029	0.205
Note:	*p<0.05; **p<0.01; ***p<0.001 Cluster–robust standard errors in parentheses.		

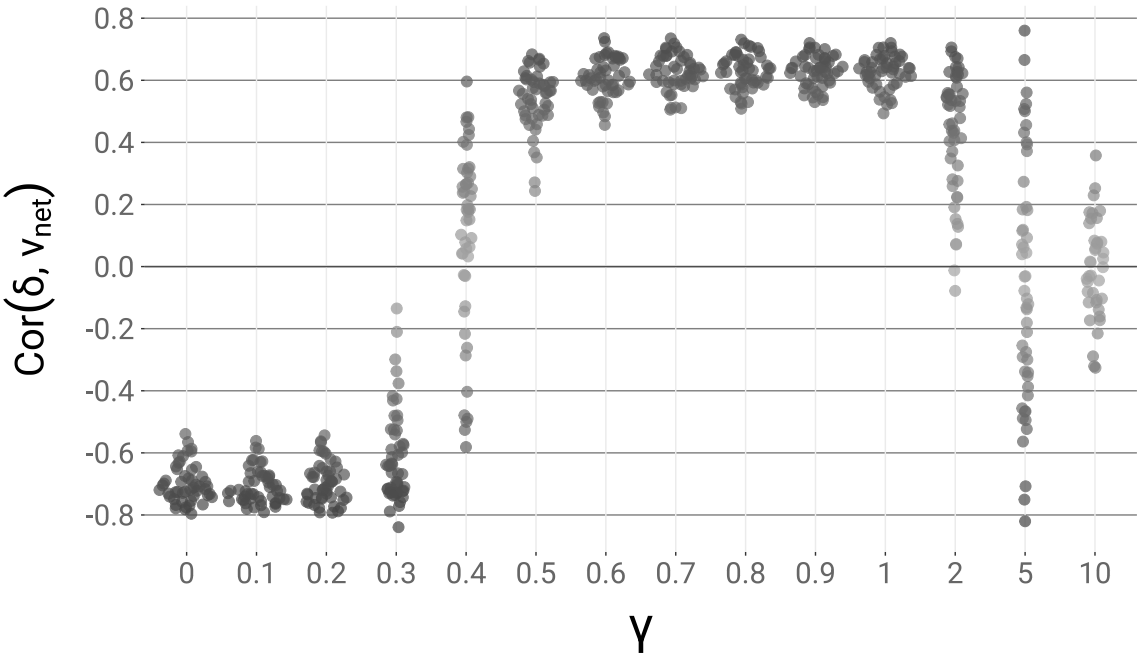
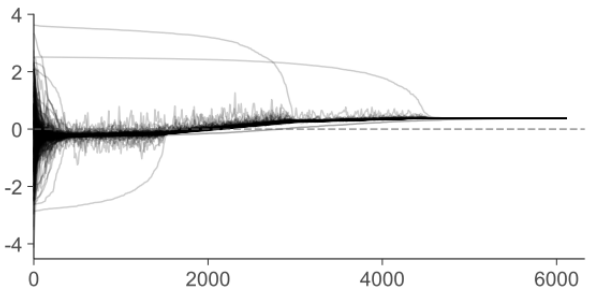
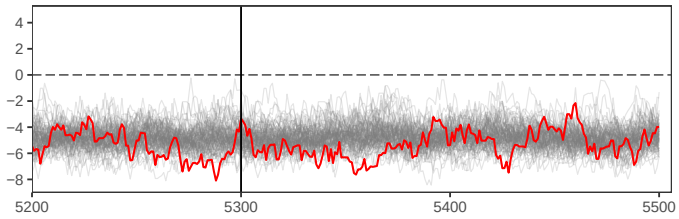


Figure 14: Sensitivity of the correlation between beliefs and values to different configurations of γ. For each γ-value, $N = 100$ networks were simulated with $\lambda = 100$. Note that DeGroot-influence ($\gamma = 0$) precludes belief-value constellations from developing.

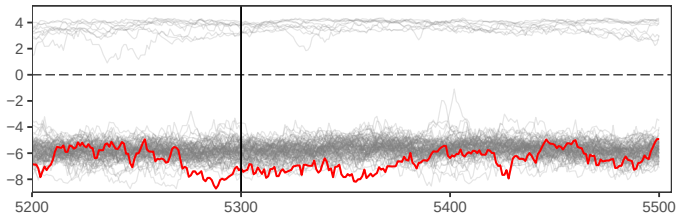


(b) $\gamma = 3$

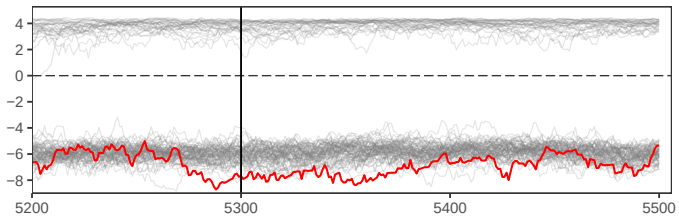
Figure 15: Evolution of attitudes without *PI*s



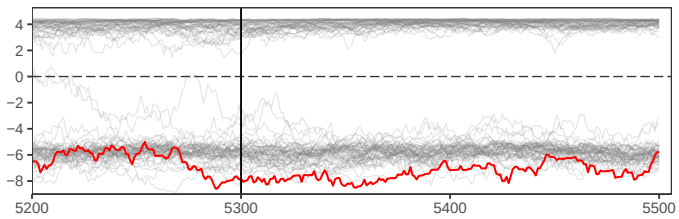
(a) $\gamma = 0.2$



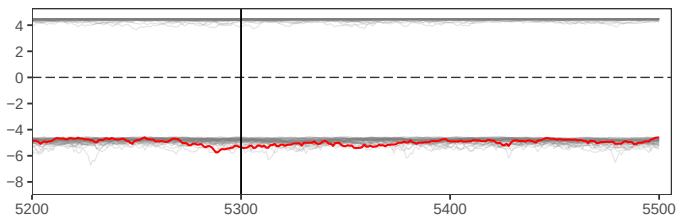
(b) $\gamma = 0.4$



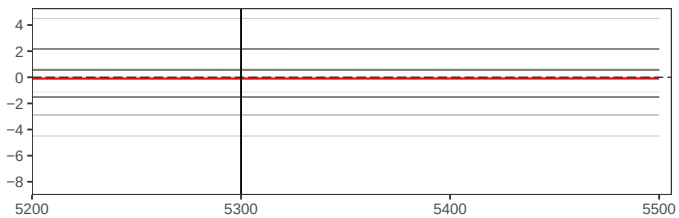
(c) $\gamma = 0.5$



(d) $\gamma = 0.6$

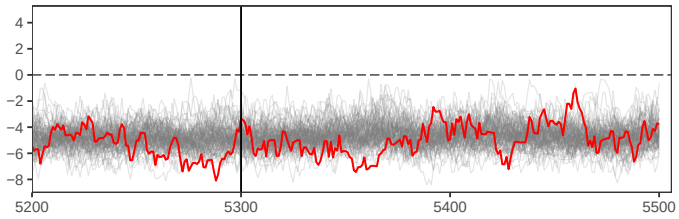


(e) $\gamma = 1$

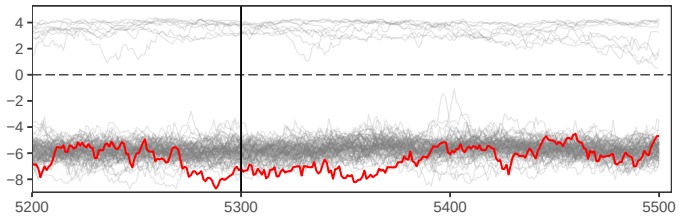


(f) $\gamma = 10$

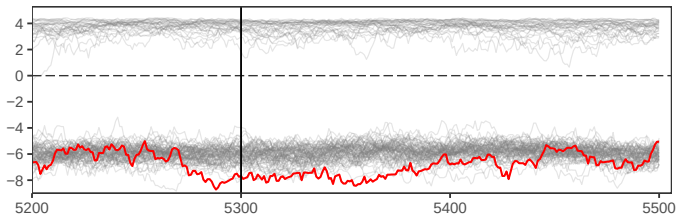
Figure 16: **Individual attitude trajectories of citizens in pandering regime for** $5200 \leq t \leq 5600$ **with** $\lambda = 100$ for selected values of γ. Red trajectory marks an example citizens. The latter and the random seed are identical across configurations.



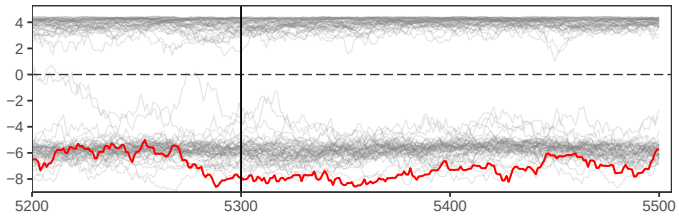
(a) $\gamma = 0.2$



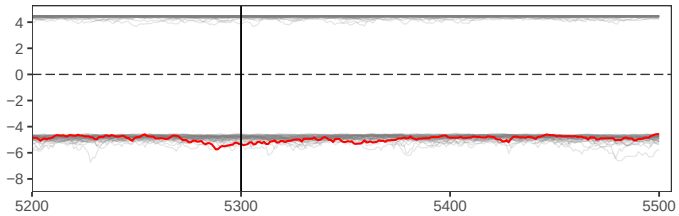
(b) $\gamma = 0.4$



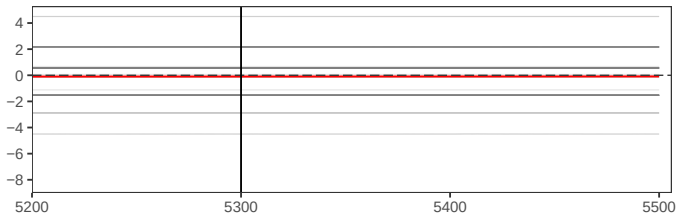
(c) $\gamma = 0.5$



(d) $\gamma = 0.6$

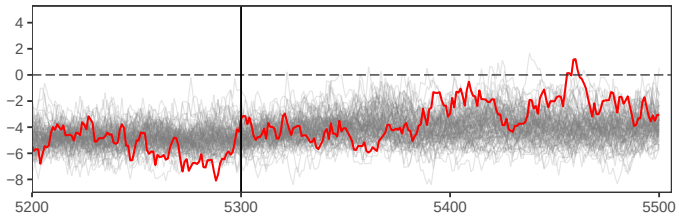


(e) $\gamma = 1$

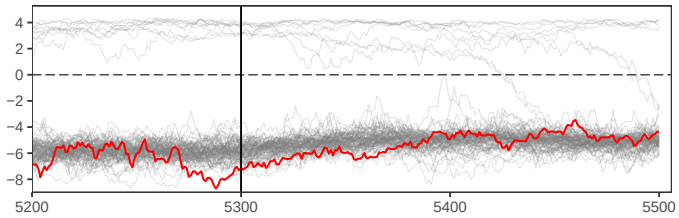


(f) $\gamma = 10$

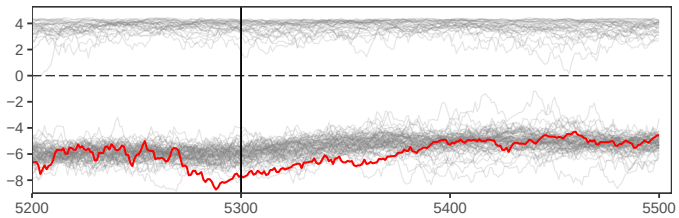
Figure 17: **Individual attitude trajectories of citizens in rebalancing regime for** $5200 \leq t \leq 5600$ **with** $\lambda = 100$ for selected values of γ. Red trajectory marks an example citizens. The latter and the random seed are identical across configurations.



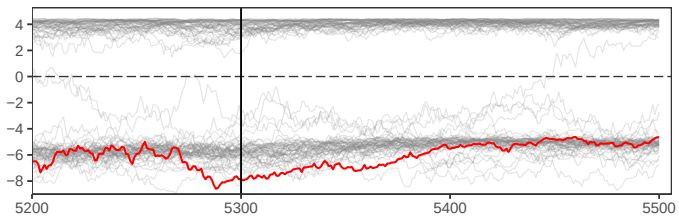
(a) $\gamma = 0.2$



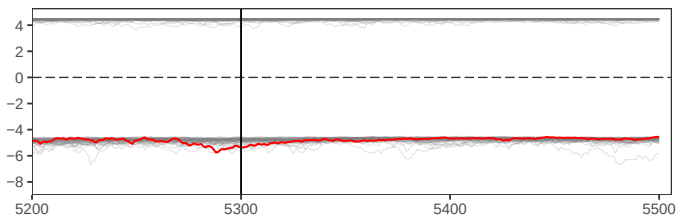
(b) $\gamma = 0.4$



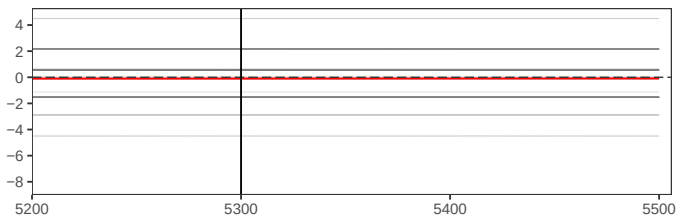
(c) $\gamma = 0.5$



(d) $\gamma = 0.6$



(e) $\gamma = 1$



(f) $\gamma = 10$

Figure 18: **Individual attitude trajectories of citizens in opportunity framing regime for** $5200 \leq t \leq 5600$ **with** $\lambda = 100$ for selected values of γ. Red trajectory marks an example citizens. The latter and the random seed are identical across configurations.

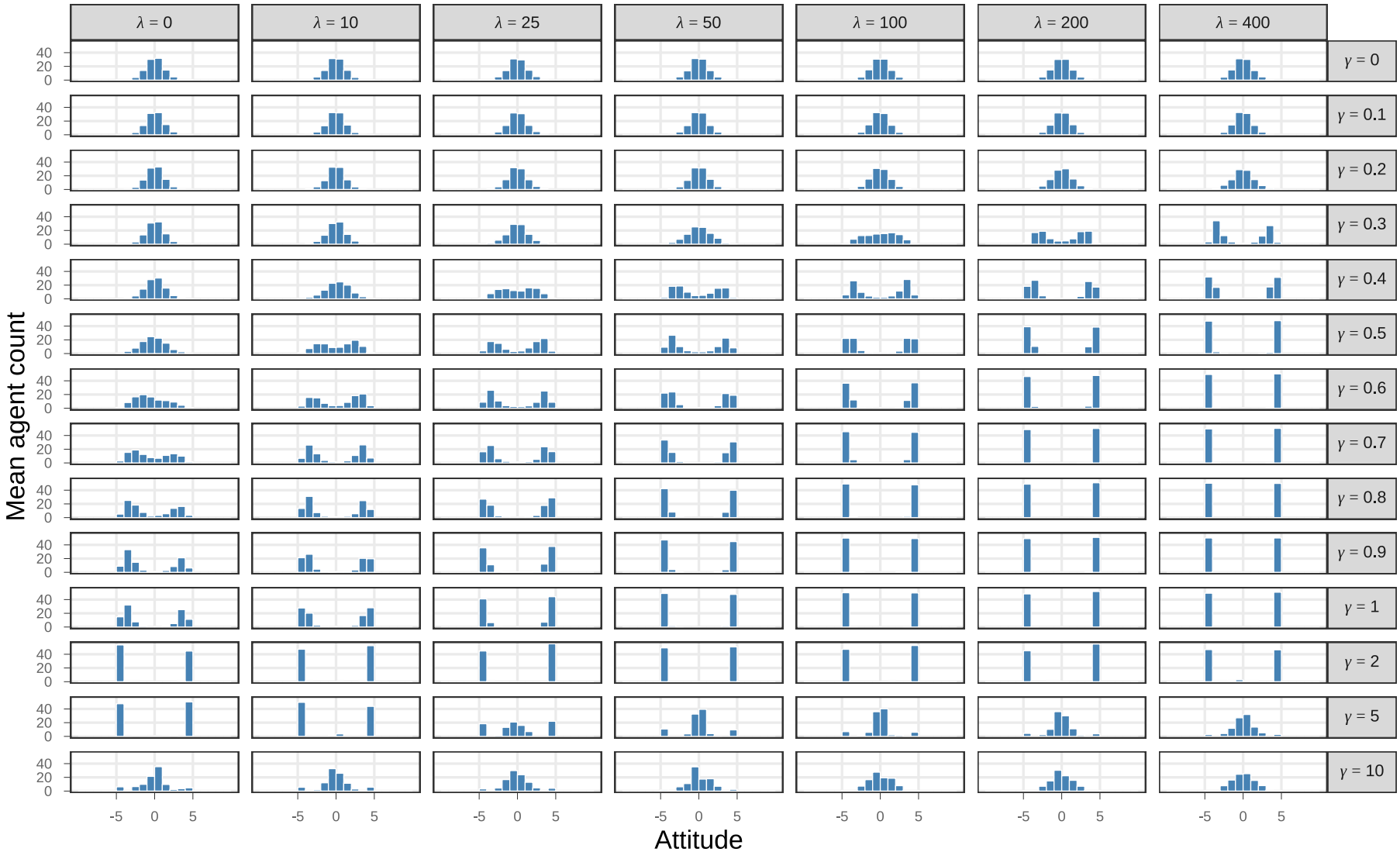


Figure 19: **Histograms of average attitude distributions** for all (λ, γ) combinations at $t = 4999$, before the populist shock. $N = 200$ $\forall(\lambda, \gamma)$.