

Coupled opinion-environmental dynamics in polarized and prejudiced populations

Cameron Kerr^{a,*}, Madhur Anand^b, Chris T. Bauch^a

*^aDepartment of Applied Mathematics, University of Waterloo, 200 University Ave
W, Waterloo, N2L 3G1, Ontario, Canada*

*^bSchool of Environmental Sciences, University of Guelph, 50 Stone Rd E, Guelph, N1G
2W1, Ontario, Canada*

^{*}Corresponding author, *E-mail address*: c5kerr@uwaterloo.ca

Abstract

Public opinion on environmental issues remains polarized in many countries, posing a significant barrier to the implementation of effective policies. Behind this polarization, empirical studies have identified social susceptibility, personal prejudice, and personal experience as dominant factors in opinion formation on environmental issues. However, current coupled human-environment models have not yet incorporated all three factors in polarized populations. We developed a stylized coupled human-environment model to investigate how social susceptibility, personal prejudice, and personal experience shape opinion formation and the environment in polarized populations. Using analytical and numerical methods, we characterized the conditions under which polarization, consensus, opinion changes, and cyclic dynamics emerge depending on the costs of mitigation, environmental damage, and the factors influencing opinion formation. Our model shows that prejudice is the key driver of persistent polarization, with even slightly prejudiced populations maintaining indefinite polarization independent of their level of objectivity. We predict that polarization can be reduced by decreasing the role of prejudice or increasing the willingness to consider opposing opinions. Finally, our model shows that cost reduction methods are less effective at reducing environmental impact in prejudiced populations. Our model generates thresholds for when reducing costs or emissions is more useful depending on the factors which influence the population’s opinion formation. Overall, our model provides a framework for investigating the importance of cognitive and social structures in determining human-environment dynamics.

Keywords: coupled human-environment systems, opinion dynamics, polarization, social learning

1. Introduction

Environmental protection is a critical issue for which consensus is needed for collective action. Public opinion is a significant determinant of public policy (Burstein, 2003). However, there is a lack of consensus in public opinion on a variety of environmental issues (Chiapella et al., 2018; Funk et al., 2020; Guo et al., 2019; Leiserowitz et al., 2023). For example, surveys show that only 50-60% of individuals in the United States believe that climate

action is a high priority, with a slightly higher 60-70% in Canada expressing the same view (Leiserowitz et al., 2023). Furthermore, environmental issues have become increasingly politicized, deepening polarization in public opinion (Chinn et al., 2020; Smith et al., 2024).

Polarization and disagreement in public opinion on environmental issues are governed by a variety of social, psychological, and experiential factors. For example, social identities are formed around agreeing individuals, further strengthening polarization (Bliuc et al., 2015; Irene et al., 2024; Iyengar et al., 2012). The integration of social identity into opinion formation was shown to be the main factor in determining participation in climate strikes among youth (Brügger et al., 2020). An individual’s experiences with negative environmental issues also plays a role in their opinion formation process (Bergquist et al., 2019; Papp, 2022; Sloggy et al., 2021). For example, experience with natural disasters, temperature anomalies and poor air quality were associated with environmental concern and voting for environmentally-motivated platforms (Hoffmann et al., 2022; Papp, 2022). In the United States, experiencing extreme climate events like Hurricane Irma was shown to increase belief in climate change (Bergquist et al., 2019). However, individual’s experiences with the environment are often shaped by their prior beliefs and prejudices (Chiapella et al., 2018; Denny et al., 2022; Guo et al., 2019; Howe and Leiserowitz, 2013; Myers et al., 2013). For example, a United States survey found that individuals who were dismissive of climate change were less inclined to report higher-than-normal temperatures during the summer (Howe and Leiserowitz, 2013). Current investigations into these factors (social influence, prejudice, and experience) have treated them independently. Further research should aim to understand the interplay of these factors in determining the environmental dynamics in polarized populations and populations in disagreement.

Despite the importance of social and psychological factors in environmental issues, many models treat humans and the environment as uncoupled systems. Over longer time scales, such as in many environmental systems, human decision making becomes an essential factor. Recent coupled human-environment system (CHES) models aim to incorporate the effects of human decision making on the environment and vice versa. These models display a wider variety of qualitative regimes when compared to either uncoupled case (Tilman et al., 2020). CHES models have been applied towards a variety of environmental phenomena including forest conservation (Satake et al., 2007; Sigdel et al., 2017), fisheries management (Bailey et al., 2019; Tilman et al.,

2017), and climate change (Beckage et al., 2018; Bury et al., 2019; Moore et al., 2022).

One approach used in CHES models relies on evolutionary game theory (Bury et al., 2019; Tilman et al., 2020; Weitz et al., 2016). In these models, the proportion of mitigating individuals is dependent on the strategy’s associated payoff. This framework has allowed for analytical approaches to study the long-term behavior of various human-environment dynamics (Tilman et al., 2020). Stability analysis of the dynamics demonstrates that long-term behavior is dependent on the incentives of mitigation and the rate of environmental feedback. To further incorporate social factors, models have factored in the benefit of complying with social norms into the net cost of mitigation (Bury et al., 2019).

Opinion dynamics models provide an approach more tailored towards analyzing the opinion-formation process of individuals. Opinion dynamics models are a type of agent-based model used to study how a population’s opinions evolve over time according to social and psychological influences (Noorazar et al., 2020). Each agent’s opinion is shaped by the opinions of the agents they interact with. The set of interacting agents is determined by a social network or by a confidence bound (meaning agents only interact with agents whose opinions are near their own). Different models incorporate additional mechanisms. In the Friedkin-Johnsen (FJ) model (Friedkin and Johnsen, 1990), each agent has a prejudice that continues to influence them during their opinion formation process. The relative weight between internal bias and social susceptibility can vary across agents or populations. In the Hegselmann-Krause (HK) model with truth seekers (Hegselmann and Krause, 2006), agents adjust their opinions based on a fixed ‘true’ opinion. The degree to which agents prioritize the ‘truth’ versus social influence can significantly affect how opinions evolve in the system.

Despite the abundance of research on opinion dynamics models and their relevance to the factors governing public opinion on environmental issues, coupled opinion-environmental models are a recent development. A recent study coupled an adapted FJ model with climate dynamics to investigate the emergence of polarization (Kumar et al., 2025). They found that opinions favored mitigation after steep temperature increases. Despite an initially uniform distribution of opinions the model demonstrated that polarization pushed agent’s opinions to either extreme.

High-dimensionality makes analyzing coupled opinion-environmental models difficult. One approach to this problem has been to reduce the model to a

2-dimensional system by considering the dynamics at consensus (Couthures et al., 2025a,b). There is a need to apply this method to a more realistic setting which reflects the disagreement and polarization seen in real populations. Under this setting, we could better investigate the social, psychological, and experiential factors which lead to polarization.

In this paper we develop an analytical understanding of opinion-environmental dynamics in polarized populations. We conduct stability and numerical bifurcation analysis to answer the following research questions: i) In what scenarios does polarization exist? Can polarization be reduced? ii) What is the long-term behavior of the opinion-environmental system in polarized populations? How does prejudice change this behavior? iii) What changes can be made to reduce the environmental impact in prejudiced and polarized populations? iv) In what scenarios does the population converge on the objective best response? Is it possible for a prejudiced population to converge on the objective best response?

2. Materials and Methods

2.1. Model

In this paper, simplified models are used to describe opinion and environmental dynamics. This allows us to develop an analytical understanding of how the factors influencing public opinion and the environment interact. These stylized models can generate interesting effects that warrant further study. The population in our model consists of N agents $A = \{1, \dots, N\}$. Each agent $i \in A$ has an opinion $o_i(t) \in [-1, 1]$, we define the opinion profile as $o(t) = (o_i(t))_{i \in A}$. An agent is a mitigator if their opinion is closer to 1, and is a non-mitigator if their opinion is closer to -1.

Environmental dynamics. We define the environmental state $e(t) \in \mathbb{R}_{\geq 0}$ representing the level of a resource that naturally decays, such as a pollutant. The rate that the pollutant naturally decays is given by γ . The population's emission rate at time t is $u(o(t))$. The function $u : [-1, 1]^N \rightarrow \mathbb{R}_{\geq 0}$ maps the opinion profile of the population to their emission rate. We define τ_e as the relative speed of the environmental dynamics. For simplicity, we assign $\tau_e = 1$. The environmental state is governed by:

$$\tau_e \dot{e}(t) = -\gamma e(t) + u(o(t)) \quad (1)$$

Opinion dynamics. An agent's opinion is determined by a weighted combination of agents they interact with, personal prejudice, and the objec-

tive best response corresponding to the current environmental state. Each of these factors have been shown to influence public opinion on environmental issues in empirical studies (Bergquist et al., 2019; Bliuc et al., 2015; Brügger et al., 2020; Chiapella et al., 2018; Guo et al., 2019; Hoffmann et al., 2022; Howe and Leiserowitz, 2013; Papp, 2022; Sloggy et al., 2021). Agent interaction is dependent on a confidence bound, r . Agents $i, j \in A$ interact if and only if the distance between their opinions is less than or equal to r ($|o_i(t) - o_j(t)| \leq r$). We define the set of agents that interact with agent i at time t to be $N_i(t) = \{j \in A : |o_i(t) - o_j(t)| \leq r\}$.

An agent's prejudice is defined to be their initial opinion $o_i(0)$. The objective best response in our model is dynamic since it is dependent on the environmental state. The objective best response is defined by a function $h : \mathbb{R}_{\geq 0} \rightarrow \{-1, 0, 1\}$ that maps the environmental state to the opinion with the highest payoff. Combining these factors, we model the opinion of each agent by:

$$\tau_o \dot{o}_i(t) = -o_i(t) + \frac{\lambda(1 - \omega)}{|N_i(t)|} \sum_{j \in N_i(t)} o_j(t) + \lambda \omega o_i(0) + (1 - \lambda)h(e(t)) \quad (2)$$

The relative strength of social susceptibility, prejudice, and objectivity in opinion formation are given by the influence coefficients $\lambda(1 - \omega)$, $\lambda\omega$, and $1 - \lambda$ respectively. The influence coefficients lie on the simplex Δ^2 since $\lambda \in [0, 1]$ and $\omega \in [0, 1]$. τ_o represents the speed of opinion dynamics. Similar to previous studies (Bury et al., 2019; Kumar et al., 2025) we will assume that the opinion dynamics evolve slower than the environmental dynamics, $\tau_o > \tau_e = 1$. For all analysis and simulations discrete versions of Eq (1) and (2) were used.

2.2. Emission and response functions

For the results presented in Sections 3.2, 3.3, 3.4, 3.5 it is necessary to define appropriate functions h and u . We assume that a population full of non-mitigators ($o_i(t) = -1 \quad \forall i \in A$) emit the pollutant at a rate α and a population of mitigators ($o_i(t) = 1 \quad \forall i \in A$) does not emit. Assuming that the emission rate is linearly dependent on the opinion profile we define $u(o(t)) = \frac{1}{2}\alpha(1 - \frac{1}{N} \sum_{i=1}^N o_i(t))$. The best response function h equals 1 or -1 depending on whether the mitigator or non-mitigator opinion has a higher payoff, respectively. For simplicity, we define the payoff of switching from a non-mitigator to a mitigator opinion to be $-\beta + e(t)$ where β is the net cost

of mitigating and $e(t)$ is the level of pollutant, which harms an individual's health or psychology. Therefore, $h(e(t)) = \text{sign}(-\beta + e(t))$.

2.3. Visualization and numerical bifurcation analysis

Heat maps were used to visualize time-averaged values for $e(t)$ and $o(t)$ over the cost-emissions (β, α) parameter space. The parameter space was divided into a 100×100 grid spanning $\beta \in [-10, 10]$ and $\alpha \in [0, 2]$. These values were chosen to adequately sample all regions with qualitatively distinct long-term behavior. The model was simulated for $T = 1000$ steps and the first 900 time steps were removed to eliminate transients. The remaining values for $e(t)$ and $o(t)$ were averaged and used for visualization. All simulations were initialized with $a(0) = 1, b(0) = -1, \gamma = 0.2$.

Similarly, heat maps were used to visualize time-averaged values for $e(t)$ and $o(t)$ over the influence coefficients simplex. Parameter values were probed from a 100×100 grid spanning $\lambda \in [0, 1]$ and $\omega \in [0, 1]$, each associated with a unique triplet $(\lambda\omega, \lambda(1 - \omega), 1 - \lambda)$ in the simplex Δ^2 . As before, the model was simulated for $T = 1000$ steps and the last 100 were used to obtain average values for $e(t)$ and $o(t)$. All simulations were initialized with $a(0) = 1, b(0) = -1, \gamma = 0.2$.

Heat maps were used to visualize time-averaged values for polarization $P(t)$ over the prejudice-confidence bound $(\lambda\omega, r)$ parameter space. The parameter space was divided into a 300×300 grid spanning $\lambda\omega \in [0, 0.5]$ and $r \in [0, 2]$. As before, the polarization map was simulated for $T = 1000$ steps and the last 100 were used to obtain average values for $P(t)$. All simulations were initialized with $a(0) = 1, b(0) = -1, \gamma = 0.2, 1 - \lambda(1 - \omega) = 0.5$.

Numerical bifurcation analysis was used to investigate dependence of the behavior of $e(t)$ and $o(t)$ on the relative cost of mitigation (θ), the net cost of mitigation (β), and the emission rate (α). In each case, 10000 values were sampled from $\theta \in [0, \pi]$, $\beta \in [-10, 10]$ and $\alpha \in [0, 2]$, respectively. Since θ represents an angle, a magnitude of 5 was chosen to specify a unique point in polar coordinates in the (β, α) parameter space. The model was simulated for $T = 1000$ time steps and the first 900 time steps were removed to eliminate transients. The remaining values are visualized in the resulting numerical bifurcation diagram.

All simulations and visualizations were created in Python 3.13.5 using the NumPy, Matplotlib, and Mpltern packages. For further details on the numerical analysis and code see: <https://github.com/CameronKerr/PolDynamics>.

2.4. Reduction of dynamics in polarized populations

We focus our analysis of the opinion-environmental dynamics on the case in which the population is polarized. In a polarized population the system presented in Eq (1) and (2) reduces to three dimensions.

Lemma 1: For $i, j \in A$, if $o_i(0) = o_j(0)$ then $o_i(t) = o_j(t) \forall t$.

Proof: We proceed by induction, the base case ($t = 0$) is given. Assume that for some $t \geq 0$, $o_i(t) = o_j(t)$. Therefore, for any $k \in A$, $|o_i(t) - o_k(t)| \leq r$ iff $|o_j(t) - o_k(t)| \leq r$. This implies $N_i(t) = N_j(t)$. Therefore:

$$\begin{aligned} o_i(t+1) &= (1 - \frac{1}{\tau_o})o_i(t) + \frac{1}{\tau_o}(\frac{\lambda(1-\omega)}{|N_i(t)|} \sum_{k \in N_i(t)} o_k(t) + \lambda\omega o_i(0) + (1-\lambda)h(e(t))) \\ &= (1 - \frac{1}{\tau_o})o_j(t) + \frac{1}{\tau_o}(\frac{\lambda(1-\omega)}{|N_j(t)|} \sum_{k \in N_j(t)} o_k(t) + \lambda\omega o_j(0) + (1-\lambda)h(e(t))) \\ &= o_j(t+1) \blacksquare \end{aligned}$$

Similarly to the synchronization manifold presented in (Couthures et al., 2025a) we introduce the polarization manifold, L . The polarization manifold consists of all opinion profiles and environmental states in which the opinion profile can be split into two groups where each group consists of agents with the same opinion.

$$L = \{(o, e) \in [-1, 1]^N \times \mathbb{R}_{\geq 0} \mid o = (a\mathbf{1}_n, b\mathbf{1}_{N-n}), a, b \in [-1, 1], n \leq N\}$$

Proposition 1: If $(o(0), e(0)) \in L$, then $(o(t), e(t)) \in L \forall t$.

Proof: We proceed by induction, the base case ($t=0$) is given. Since $(o(0), e(0)) \in L$ there exists n such that $o_i(0) = o_j(0) \forall i, j \in \{1, \dots, n\}$ and $o_i(0) = o_j(0) \forall i, j \in \{n+1, \dots, N\}$. By Lemma 1, $o_i(t) = o_j(t) \forall i, j \in \{1, \dots, n\}$ and $o_i(t) = o_j(t) \forall i, j \in \{n+1, \dots, N\}$ for all t . Therefore, $\forall t$ there exists $a(t), b(t) \in [-1, 1]$ such that $o(t) = (a(t)\mathbf{1}_n, b(t)\mathbf{1}_{N-n})$. $\blacksquare$

We can redefine the dimension $N+1$ system defined in Eq (1) and (2) to a 3-dimensional system in terms of the opinion of the two polarized groups $(a(t), b(t))$ and the environmental state $(e(t))$. Note that the dynamics change depending on whether the polarized groups are within the bound of confi-

dence r . If $|a(t) - b(t)| > r$:

$$\begin{aligned} a(t+1) &= (1 - \frac{1}{\tau_o})a(t) + \frac{1}{\tau_o}(\lambda(1 - \omega)a(t) + \lambda\omega a(0) + (1 - \lambda)h(e(t))) \\ b(t+1) &= (1 - \frac{1}{\tau_o})b(t) + \frac{1}{\tau_o}(\lambda(1 - \omega)b(t) + \lambda\omega b(0) + (1 - \lambda)h(e(t))) \\ e(t+1) &= (1 - \gamma)e(t) + u([a(t)\mathbf{1}_n, b(t)\mathbf{1}_{N-n}]^T) \end{aligned}$$

If $|a(t) - b(t)| \leq r$:

$$\begin{aligned} a(t+1) &= (1 - \frac{1}{\tau_o})a(t) + \frac{1}{\tau_o}(\lambda(1 - \omega)(pa(t) + (1 - p)b(t)) + \lambda\omega a(0) + (1 - \lambda)h(e(t))) \\ b(t+1) &= (1 - \frac{1}{\tau_o})b(t) + \frac{1}{\tau_o}(\lambda(1 - \omega)(pa(t) + (1 - p)b(t)) + \lambda\omega b(0) + (1 - \lambda)h(e(t))) \\ e(t+1) &= (1 - \gamma)e(t) + u([a(t)\mathbf{1}_n, b(t)\mathbf{1}_{N-n}]^T) \end{aligned}$$

where p is the proportion of agents in group $a(t)$, $p = \frac{n}{N}$.

3. Results

3.1. Long-term existence of polarization

In this section, we show that polarization persists in a population whenever prejudice is present. Furthermore, the level of polarization at equilibrium is proportional to the strength of prejudice in the population. To reduce polarization, one must either reduce the level of prejudice or broaden the range of tolerance in which agents in each polarized group consider the opposing opinion.

To formalize the notion of polarization, we define the polarization function as the distance between the two polarized opinion groups:

$$P(t) = |a(t) - b(t)|$$

Without loss of generality, we assume $a(0) > b(0)$. From this assumption it follows that $a(t) \geq b(t)$ for all t (see Appendix A.1), allowing us to simplify the polarization function to $P(t) = a(t) - b(t)$. By substituting the expressions for $a(t)$ and $b(t)$ we can determine the difference equations governing polarization. It is important to note that any environmental influence is symmetric across both groups and therefore cancels out in the calculation of polarization.

When $P(t) > r$:

$$P(t+1) = \frac{\tau_o - 1 + \lambda(1 - \omega)}{\tau_o} P(t) + \frac{\lambda\omega}{\tau_o} P(0)$$

When $P(t) \leq r$:

$$P(t+1) = \frac{\tau_o - 1}{\tau_o} P(t) + \frac{\lambda\omega}{\tau_o} P(0)$$

Proposition 2: If $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} < r$ then $\lim_{t \rightarrow \infty} P(t) = \lambda\omega P(0)$. If $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} > r$ then $\lim_{t \rightarrow \infty} P(t) = \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)}$. *Proof:* See Appendix A.2

This proposition, along with the heatmap of long-term polarization (Figure 1), characterizes the long-term behavior of polarization in terms of the factors influencing opinion formation. Polarization converges to 0 if and only if the population is non-prejudiced ($\lambda\omega = 0$) and has a non-zero objectivity ($1 - \lambda > 0$). Conversely, if the population is prejudiced ($\lambda\omega > 0$), polarization will persist indefinitely ($\lim_{t \rightarrow \infty} P(t) > 0$). However, polarization will always decrease from its initial value $P(0)$ as long as the population exhibits some degree of objectivity.

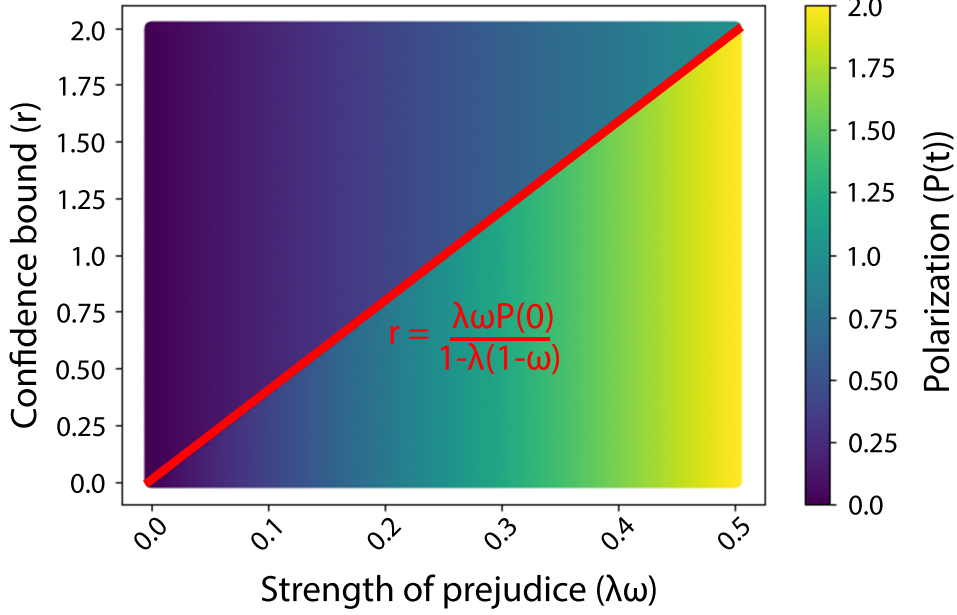


Figure 1: Long-term behavior of polarization. Long-term behavior of the polarization $P(t) = |a(t) - b(t)|$ between the opinion groups over the strength of prejudice ($\lambda\omega$) and the confidence bound (r). The sharp line represents the theoretical threshold ($r = \frac{\lambda\omega P(0)}{1 - \lambda(1 - \omega)}$) separating the regions of high and low polarization described in Proposition 2.

In prejudiced populations, polarization can be reduced by increasing the confidence bound r above the threshold $\frac{\lambda\omega P(0)}{1 - \lambda(1 - \omega)}$. Above this threshold, the two opinion groups consider each other's opinions at equilibrium. Doing so reduces polarization by a factor of $\frac{1}{1 - \lambda(1 - \omega)}$. In the non-prejudiced case, polarization always converges to zero, although increasing r increases the rate of convergence (for details see Appendix A.3).

3.2. Behavior of the opinion-environmental system in non-prejudiced populations

In this section, we show that the opinion profile and level of pollutant in non-prejudiced populations hinges on whether the cost of mitigation is higher or lower than the maximum environmental cost.

To begin, we analyze the dynamics in non-prejudiced populations ($\lambda\omega = 0$). According to Proposition 2, the polarized groups in a non-prejudiced population converge to a consensus. Therefore, we can reduce our analysis

by considering a population at consensus, let the consensus opinion be $c(t)$. Since there is no prejudice, agents value the opinion of others with a strength $\lambda \in [0, 1]$ and the objective best response with strength $1 - \lambda$. Incorporating our functions h and u , our model reduces to:

$$\begin{aligned} c(t+1) &= \frac{\tau_o - 1 + \lambda}{\tau_o} c(t) + \frac{1 - \lambda}{\tau_o} \text{sign}(-\beta + e(t)) \\ e(t+1) &= (1 - \gamma)e(t) + \frac{1}{2}\alpha(1 - c(t)) \end{aligned}$$

Proposition 3: The point $(c, e) = (1, 0)$ is a stable fixed point when $\beta < 0$ and $(c, e) = (-1, \frac{\alpha}{\gamma})$ is a stable fixed point when $\beta > \frac{\alpha}{\gamma}$. *Proof:* See Appendix B.1.

This proposition describes the long-term behavior in two regions of the cost-emissions (β, α) parameter space. In the first region, there is a net benefit to mitigation ($\beta < 0$) making it the objective best response ($h(e(t)) = 1$). In this region, the proposition implies that the population holds a purely mitigative opinion ($c(t) = 1$). In the second region, where the cost of mitigating exceeds the maximum environmental cost ($\beta > \frac{\alpha}{\gamma}$), not mitigating is the objective best response ($h(e(t)) = -1$). Here, the population holds a purely non-mitigative opinion ($c(t) = -1$). We refer to the region where mitigation is the best response as the mitigative region, and the region where non-mitigation is the best response as the non-mitigative region.

Figure 2a presents a numerical heatmap that confirms these analytical predictions. It shows sharp boundaries defining the mitigative and non-mitigative regions, consistent with the theoretical thresholds. The opinion profile and environmental state observed in each region are consistent with the equilibrium described in Proposition 3.

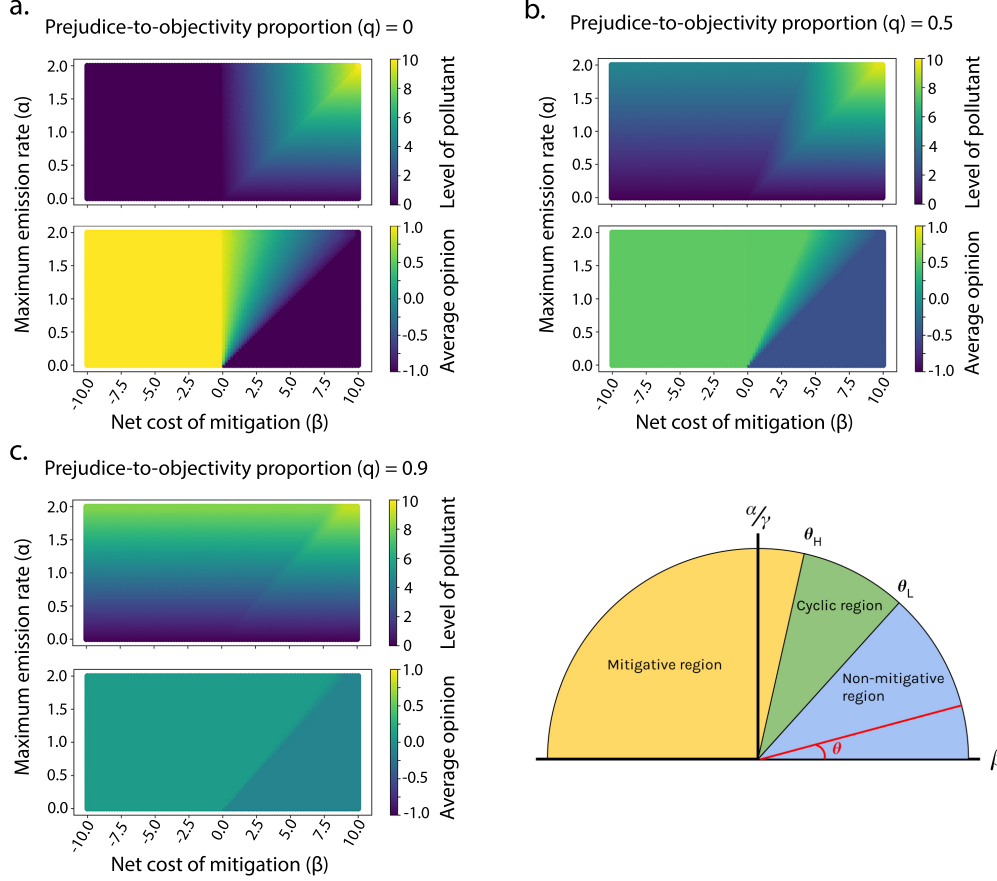


Figure 2: Long-term behavior over the cost-emissions parameter space. Effect of the net cost of mitigation (β) and the maximum emission rate (α) on the level of pollutant and average opinion in the population. Heat maps visualize the long-term behavior in populations with (a) no prejudice ($q = 0$), (b) equal prejudice and objectivity ($q = 0.5$), and (c) high levels of prejudice ($q = 0.9$). In each case, the behavior is separated into three regions (mitigative, cyclic, and non-mitigative regions) dependent on the angle to the β -axis, $\theta = \tan^{-1}(\frac{\alpha}{\beta\gamma})$. This behavior is constant along lines from the origin in the (β, α) parameter space shown in (a-c).

We can define the region in between these cases the cyclic region. In the non-prejudiced case, this region occurs when $0 < \beta < \frac{\alpha}{\gamma}$. Here, there is a cost to mitigate, but it is 'worth it' because the potential environmental damage outweighs the cost. Numerical bifurcation diagrams (Figure 3a) show that the model exhibits cycles in this region. Specifically, the environmental state $e(t)$ oscillates above and below the cost of mitigation β and the opinions

fluctuate around the value $c(t) = 1 - \beta \frac{2\gamma}{\alpha}$. The numerical heatmap in Figure 2a supports these predictions by confirming the presence of the cyclic region in the corresponding parameter space.

Proposition 3 and the numerical heat map in Figure 2a demonstrate that the mitigative, non-mitigative, and cyclic regions of the (β, α) parameter space are defined by their angle to the β -axis. Let us denote this angle by θ , defined as $\theta = \tan^{-1}(\frac{\alpha}{\beta\gamma})$. Intuitively, θ represents how the net cost of switching to a mitigator strategy (β) compares to the maximum environmental cost ($\frac{\alpha}{\gamma}$). A low value indicates that mitigation is a costly strategy relative to the environmental damage, while a high value suggests the opposite. Therefore, we refer to θ as the *relative cost of mitigation*.

Our previous results can be restated in terms of the relative cost of mitigation. The non-mitigative region occurs in non-prejudiced populations when $0 < \theta < \frac{\pi}{4}$, the cyclic region when $\frac{\pi}{4} < \theta < \frac{\pi}{2}$, and the mitigative region when $\frac{\pi}{2} < \theta < \pi$. The numerical bifurcation diagram over θ shown in Figure 3a confirms these boundaries.

3.3. Behavior of the opinion-environmental system in prejudiced populations

In this section, we find that in a prejudiced population, each individual's opinion converges to a weighted average between their own prejudice and the objective best response.

To analyze the qualitative behavior of the system in prejudiced populations, we begin by computing numerical bifurcation diagrams with respect to the relative cost of mitigation θ . These diagrams (Figure 3b and 3c) reveal that while the overall structure of the mitigative, non-mitigative, and cyclic regions is preserved from the non-prejudiced case, the boundaries and system behavior differ. Most significantly, the level of pollutant no longer reduces to zero in the mitigative region, and the polarized groups never converge to a consensus.

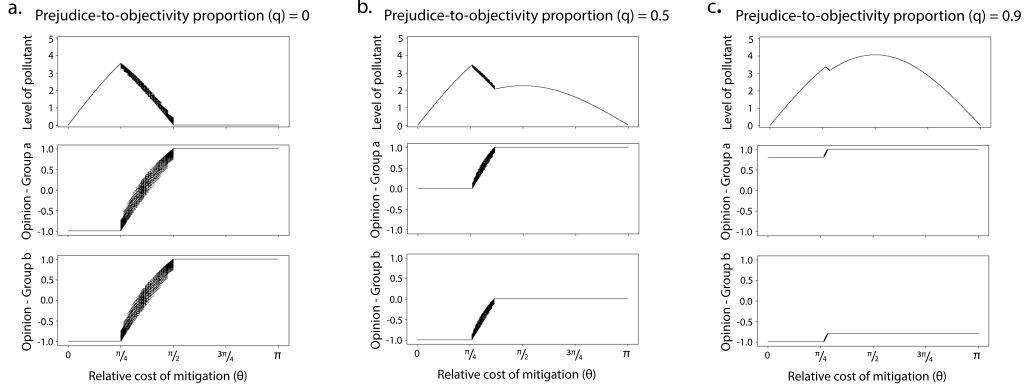


Figure 3: Long-term behavior of the opinion-environmental system over the relative cost of mitigation (θ). Bifurcation diagrams over (θ) visualize the fixed points and behaviors of the level of pollutant and opinion profile in the mitigative, non-mitigative, and cyclic regions in populations with (a) no prejudice ($q = 0$), (b) equal prejudice and objectivity ($q = 0.5$), and (c) high levels of prejudice ($q = 0.9$).

To better understand the origin of these changes, we derived analytical expressions for the fixed points in each region. Full expressions and derivations are provided in Appendix B.2. These expressions match the dynamics observed in the numerical bifurcation diagrams, as shown in Figure D.1.

The analytical expressions for the opinion profile at equilibrium show that an agent's opinion converges to a weighted average of their prejudice and the objective best response. The weight assigned to an agents prejudice is given by the *prejudice-to-objectivity proportion*, denoted $q = \frac{\lambda\omega}{\lambda\omega + (1-\lambda)}$. When $q > 0.5$, the population is more prejudiced than objective and when $q < 0.5$ the population is more objective than prejudiced. The objective best response is weighted by $1 - q = \frac{1-\lambda}{\lambda\omega + (1-\lambda)}$. This structure leads agents in both polarized groups to adopt more mitigative opinions in the mitigative region than in the non-mitigative region, as shown in the lower panels of Figures 3b and 3c. The shift toward the objective best response is more pronounced in Figure 3b, where the prejudice-to-objectivity proportion is lower. Additionally, each polarized group remains closer to their prejudices (with $a(0) = 1, b(0) = -1$ in Figures 2 and 3) when the prejudice-to-objectivity proportion is higher, as illustrated in Figure 3c.

The analytical expression for the environmental fixed point shows that the level of pollutant at equilibrium depends linearly on the prejudice-to-objectivity proportion q . In the mitigative region, the level of pollutant increases with q (as the population becomes more prejudiced and less ob-

jective) and it decreases with q in the non-mitigative region. Notably, the population's social susceptibility, given by $\lambda(1 - \omega)$, has no effect on the long-term environmental outcome. We confirm this numerically by plotting a heatmap of the limiting environmental state over the influence coefficients simplex (Figure D.2). This heatmap demonstrates that the environmental outcome remains constant along lines originating from the social susceptibility vertex, along which q remains constant.

In the cyclic region, cycles still occur in prejudiced populations (Figure 3b, 3c) but their structure is altered. The environmental state continues to oscillate around the cost of mitigation $e(t) = \beta$ and the average opinion oscillates around the same mean as in the non-prejudiced case, $c(t) = 1 - \beta \frac{2\gamma}{\alpha}$. However, unlike the non-prejudiced case, the two polarized groups do not converge to this average. Instead, each group is symmetrically pulled toward its prejudice.

3.4. *Pathways to lower environmental impact in polarized and prejudiced populations*

In this section, we show that reducing the level of pollutant by lowering the cost of mitigation is less effective in populations that are prejudiced and polarized.

The fixed-point expressions for the environment reveal thresholds that determine whether changes in the emission rate (α) or cost of mitigation (β) reduce the level of pollutant. Increasing the emission rate raises the level of pollutant when (β, α) are in the non-mitigative or mitigative region and has no impact in the cyclic region. Conversely, increasing the cost of mitigation increases the level of pollutant when (β, α) are in the cyclic region and has no impact in the mitigative or non-mitigative region. These regions of sensitivity and stagnation are visualized in the numerical bifurcation diagrams of β and α (Figure 4). Overall, the effect of a parameter change is context-dependent. For instance, if the cost of mitigation is higher than the maximum environmental damage ($\beta > \frac{\alpha}{\gamma}$) then reducing the cost has no effect on level of pollutant unless it passes a threshold that shifts the system into the cyclic region.

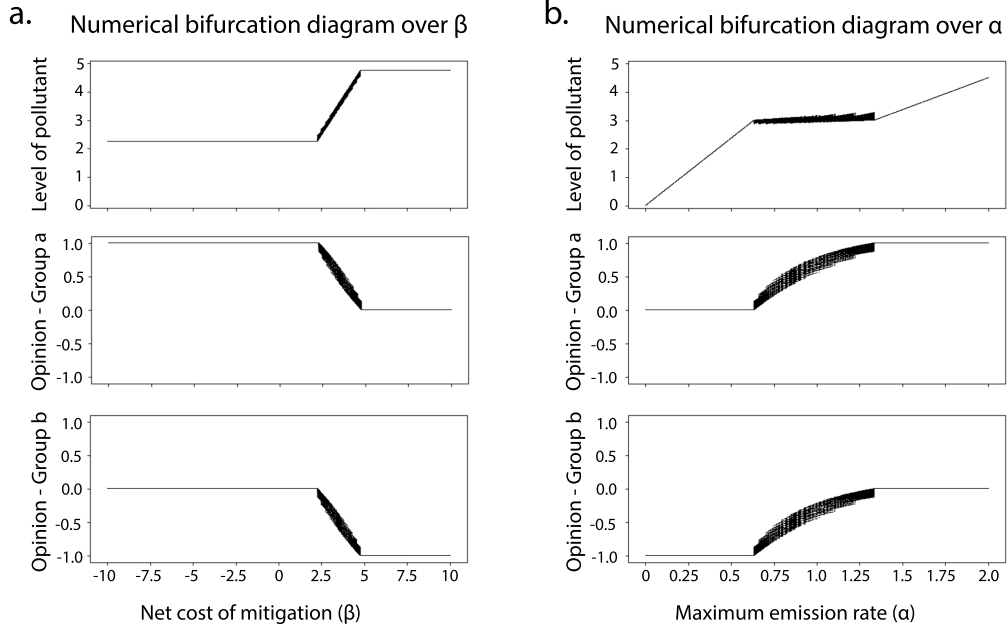


Figure 4: Regions of sensitivity and stagnation in the opinion-environmental system over the cost-emissions parameter space. Bifurcation diagrams visualize fixed points and behaviors of the level of pollutant and opinion profile over different values of the net cost of mitigation (β) and the maximum emission rate (α).

These thresholds correspond to the boundaries of the cyclic region, which lie between the mitigative and non-mitigative regions (defined by $h(e(t)) = 1$ and $h(e(t)) = -1$, respectively). Using these definitions, the lower and upper boundaries of the cyclic region are given by:

$$\theta_L = \tan^{-1}\left(\frac{2}{2 - q(1 + pa(0) + (1 - p)b(0))}\right)$$

$$\theta_H = \tan^{-1}\left(\frac{2}{q(1 - pa(0) - (1 - p)b(0))}\right)$$

These expressions show that as the prejudice-to-objectivity proportion (q) increases, the lower boundary θ_L increases and the upper boundary θ_H decreases. As a result, the cyclic region narrows, which is visible in Figures 2 and 3.

Since only the cyclic region is sensitive to changes in the cost of mitigation, we conclude that as prejudice increases relative to objectivity, lowering the

cost of mitigation has less of an effect on the level of pollutant. This is confirmed in the heatmaps in Figure 2. For low values of the prejudice-to-objectivity proportion, the level of pollutant varies along both the cost and emissions axes (Figure 2a). As prejudice increases, the heatmap shows a vertical gradient, with contours aligned horizontally (Figure 2c). This indicates that the level of pollutant becomes invariant with respect to the cost of mitigation and depend only on the emission rate.

3.5. Convergence to the objective best-response

In this section, we show that the more prejudiced a population is, the further their opinion profile is from the objective best response. In the absence of prejudice, the population converges to the best response.

To quantify the deviation of an opinion profile $o(t)$ from the objective best response $T \in [-1, 1]$ we define the distance function $d_T(o(t)) = \frac{1}{N} \sum_{i=1}^N |o_i(t) - T|$. In a polarized population, the function reduces to $d_T(o(t)) = p|a(t) - T| + (1-p)|b(t) - T|$. Note that the best response $T = h(e(t))$ is only well-defined in the mitigative and non-mitigative regions, where it remains constant at 1 or -1, respectively. In the cyclic region, the environmental state oscillates above and below the cost of mitigation, causing the best response to switch between favoring mitigation and non-mitigation. Therefore, we focus our analysis on the mitigative and non-mitigative regions.

By substituting in the fixed-point expressions from Section 3.3, we derive analytical formulas for the distance to the objective best response in each region (for more details see Appendix C). In both the mitigative and non-mitigative regions, the distance grows linearly with the prejudice-to-objectivity proportion (q). More specifically, when the polarized groups are farther than the confidence bound at equilibrium (i.e. $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} > r$) the distance from the objective best response increases linearly with q . When the groups fall within each other's confidence bound (i.e. $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} \leq r$) the same linear relationship with q holds, provided that the social susceptibility ($\lambda(1-\omega)$) remains fixed. In both cases, a polarized population converges to the objective best response if and only if the prejudice-to-objectivity proportion is 0 ($q = 0$).

4. Discussion

In this study, we developed an analytical framework for understanding opinion-environmental dynamics, focusing on the impact of social suscepti-

bility, prejudice, and objectivity. We showed that polarization persists as long as individuals remain prejudiced. We then characterized the long-term behaviour of the opinion-environmental system, finding that the outcomes depend on the cost of mitigation versus environmental damage. Additionally, individual opinions converge to a weighted average of their personal prejudice and the objective best response. We also show that reducing the cost of mitigation becomes less effective as a tool to reduce the level of pollutant in prejudiced and polarized populations, and that these populations are further from the objective best response. Together, these results highlight how important the factors governing opinion formation are in environmental dynamics.

One of the key results from our study is the strong link between prejudice and persistent polarization. In the absence of prejudice, the individuals in the population converge to a consensus, even when objectivity is arbitrarily small. By contrast, the presence of even minimal prejudice in a population results in indefinite polarization, no matter how objective individuals are. Furthermore, prejudice distances an individual’s opinion from the objective best response, no matter what their prejudiced opinion is. These points suggest that lowering the role of personal prejudices in opinion formation is crucial for consensus on the optimal strategy.

We found that incentivizing mitigation has less of an impact on the environment in prejudiced and polarized populations. This finding adds nuance to previous results which emphasize cost reduction as a pathway to environmental action (Bury et al., 2019; Kumar et al., 2025). In our model, agents who rely more on their prejudice than the objective best response are less dependent on the associated cost of their opinion. As a result, policies aimed at reducing environmental impact by lowering costs may have difficulties if the opinions of individuals are strongly prejudiced. More broadly, our analysis shows that the effectiveness of reducing costs or the maximum emissions rate is dependent on the current costs and emissions as well as the social and psychological factors underlying opinion formation.

Future extensions of opinion-environmental dynamics could incorporate additional mechanisms to capture more realistic opinion-environment dynamics and explore new research areas. For instance, to study how subjective perception of the environment affects opinion formation, the cost function could depend on the perceived, rather than actual, environmental state. To examine the influence of media or political actors, the prejudice term could be used to represent the fixed opinion of a stubborn agent representing a news

source or political party. Further simulation-based approaches would allow for more flexibility in the population, and a relaxing of this study’s simplifying assumptions. Heterogeneity in opinion formation could be incorporated by varying the strength of social-susceptibility, prejudice, and objectivity by individual. Furthermore, initial opinions could be drawn from a bimodal distribution rather than assuming strictly polarized groups.

5. Funding Sources

This work was supported by the Natural Sciences and Engineering Research Council (USRA - 604233 - 2025 and NSERC Discovery Grant to C.T. Bauch. RGPIN-2025-04274) and the Ontario Graduate Scholarship.

References

- Bailey, R.M., Carrella, E., Axtell, R., Burgess, M.G., Cabral, R.B., Drexler, M., Dorsett, C., Madsen, J.K., Merkl, A., Saul, S., 2019. A computational approach to managing coupled human–environmental systems: the poseidon model of ocean fisheries. *Sustainability Science* 14, 259–275. doi:<https://doi.org/10.1007/s11625-018-0579-9>.
- Beckage, B., Gross, L.J., Lacasse, K., Carr, E., Metcalf, S.S., Winter, J.M., Howe, P.D., Fefferman, N., Franck, T., Zia, A., Kinzig, A., Hoffman, F.M., 2018. Linking models of human behaviour and climate alters projected climate change. *Nature Climate Change* 8, 79–84. doi:<https://doi.org/10.1038/s41558-017-0031-7>.
- Bergquist, M., Nilsson, A., Schultz, P.W., 2019. Experiencing a severe weather event increases concern about climate change. *Frontiers in psychology* 10, 220. doi:<https://doi.org/10.3389/fpsyg.2019.00220>.
- Bliuc, A.M., McGarty, C., Thomas, E.F., Lala, G., Berndsen, M., Misajon, R., 2015. Public division about climate change rooted in conflicting socio-political identities. *Nature Climate Change* 5, 226–229. doi:<https://doi.org/10.1038/nclimate2507>.
- Brügger, A., Gubler, M., Steentjes, K., Capstick, S.B., 2020. Social identity and risk perception explain participation in the swiss youth climate strikes. *Sustainability* 12, 10605. doi:<https://doi.org/10.3390/su122410605>.

- Burstein, P., 2003. The impact of public opinion on public policy: A review and an agenda. *Political Research Quarterly* 56, 29–40. doi:<https://doi.org/10.2307/3219881>.
- Bury, T.M., Bauch, C.T., Anand, M., 2019. Charting pathways to climate change mitigation in a coupled socio-climate model. *PLoS Computational Biology* 15, e1007000. doi:<https://doi.org/10.1371/journal.pcbi.1007000>.
- Chiapella, A.M., Nielsen-Pincus, M., Strecker, A.L., 2018. Public perceptions of mountain lake fisheries management in national parks. *Journal of Environmental Management* 226, 169–179. doi:<https://doi.org/10.1016/j.jenvman.2018.08.040>.
- Chinn, S., Hart, P.S., Soroka, S., 2020. Politicization and polarization in climate change news content, 1985–2017. *Science Communication* 42, 112–129. doi:<https://doi.org/10.1177/1075547019900290>.
- Couthures, A., Bizyaeva, A., Varma, V.S., Franci, A., Morarescu, I.C., 2025a. Bifurcation analysis of an opinion dynamics model coupled with an environmental dynamics. *arXiv preprint* doi:<https://doi.org/10.48550/arXiv.2504.03419>.
- Couthures, A., Varma, V.S., Lasaulce, S., Morarescu, I.C., 2025b. Global synchronization of multi-agent systems with nonlinear interactions. *IEEE Control Systems Letters* 9, 354–359. doi:<https://doi.org/10.1109/LCSYS.2025.3573563>.
- Denny, R.C., Marchese, J., Fischer, A.P., 2022. Severe weather experience and climate change belief among small woodland owners: A study of reciprocal effects. *Weather, Climate, and Society* 14, 1065–1082. doi:<https://doi.org/10.1175/WCAS-D-21-0176.1>.
- Friedkin, N.E., Johnsen, E.C., 1990. Social influence and opinions. *Journal of Mathematical Sociology* 15, 193–206. doi:<https://doi.org/10.1080/0022250X.1990.9990069>.
- Funk, C., Tyson, A., Kennedy, B., Johnson, C., 2020. Concern over climate and the environment predominates among these publics. *Pew Research Center* URL: <https://www.pewresearch.org/science/2020/09/29/>

concern-over-climate-and-the-environment-predominates-among-these-publics/.
last accessed: 2025-09-23.

- Guo, T., Gill, D., Johengen, T.H., Cardinale, B.L., 2019. What determines the public's support for water quality regulations to mitigate agricultural runoff? *Environmental Science & Policy* 101, 323–330. doi:<https://doi.org/10.1016/j.envsci.2019.09.008>.
- Hegselmann, R., Krause, U., 2006. Truth and cognitive division of labor: First steps towards a computer aided social epistemology. *Journal of Artificial Societies and Social Simulation* 9, 10. URL: <https://www.jasss.org/9/3/10/10.pdf>.
- Hoffmann, R., Muttarak, R., Peisker, J., Stanig, P., 2022. Climate change experiences raise environmental concerns and promote green voting. *Nature Climate Change* 12, 148–155. doi:<https://doi.org/10.1038/s41558-021-01263-8>.
- Howe, P.D., Leiserowitz, A., 2013. Who remembers a hot summer or a cold winter? the asymmetric effect of beliefs about global warming on perceptions of local climate conditions in the us. *Global environmental change* 23, 1488–1500. doi:<https://doi.org/10.1016/j.gloenvcha.2013.09.014>.
- Irene, J.O., Daniels, C., Irene, B.N.O., Kelly, M., Frank, R., 2024. A social identity approach to understanding sustainability and environmental behaviours in south africa. *Local Environment* 0, 1–19. doi:<https://doi.org/10.1080/13549839.2024.2376546>.
- Iyengar, S., Sood, G., Lelkes, Y., 2012. Affect, not ideology: A social identity perspective on polarization. *Public opinion quarterly* 76, 405–431. doi:<https://doi.org/10.1093/poq/nfs038>.
- Kumar, A.S., Josić, K., Bauch, C.T., Anand, M., 2025. From opinion polarization to climate action: A social-climate model of the opinion spectrum. *arXiv preprint* doi:<https://doi.org/10.48550/arXiv.2503.04689>.
- Leiserowitz, A., Verner, M., Wood, E., Carman, J., Ordaz Reynoso, N., Thulin, E., Rosenthal, S., Marlon, J., Buttermore, N., 2023. *International public opinion on climate change, 2023*. New Haven, CT: Yale Program on Climate Change communication and Data for Good at

Meta URL: <https://climatecommunication.yale.edu/publications/international-public-opinion-on-climate-change-2023/>. last accessed: 2025-09-23.

- Moore, F.C., Lacasse, K., Mach, K.J., Shin, Y.A., Gross, L.J., Beckage, B., 2022. Determinants of emissions pathways in the coupled climate-social system. *Nature* 603, 103–111. doi:<https://doi.org/10.1038/s41586-022-04423-8>.
- Myers, T.A., Maibach, E.W., Roser-Renouf, C., Akerlof, K., Leiserowitz, A.A., 2013. The relationship between personal experience and belief in the reality of global warming. *Nature climate change* 3, 343–347. doi:<https://doi.org/10.1038/nclimate1754>.
- Noorazar, H., Vixie, K.R., Talebanpour, A., Hu, Y., 2020. From classical to modern opinion dynamics. *International Journal of Modern Physics C* 31, 2050101. doi:<https://doi.org/10.1142/S0129183120501016>.
- Papp, Z., 2022. Environmental attitudes, environmental problems and party choice. a large-n comparative study. *Political Geography* 97, 102652. doi:<https://doi.org/10.1016/j.polgeo.2022.102652>.
- Satake, A., Leslie, H.M., Iwasa, Y., Levin, S.A., 2007. Coupled ecological–social dynamics in a forested landscape: Spatial interactions and information flow. *Journal of Theoretical Biology* 246, 695–707. doi:<https://doi.org/10.1016/j.jtbi.2007.01.014>.
- Sigdel, R., Anand, M., Bauch, C.T., 2017. Competition between injunctive social norms and conservation priorities gives rise to complex dynamics in a model of forest growth and opinion dynamics. *Journal of Theoretical Biology* 432, 132–140. doi:<https://doi.org/10.1016/j.jtbi.2017.07.029>.
- Sloggy, M.R., Suter, J.F., Rad, M.R., Manning, D.T., Goemans, C., 2021. Changing opinions on a changing climate: the effects of natural disasters on public perceptions of climate change. *Climatic Change* 168, 25. doi:<https://doi.org/10.1007/s10584-021-03242-6>.
- Smith, E.K., Bognar, M.J., Mayer, A.P., 2024. Polarisation of climate and environmental attitudes in the united states, 1973-2022. *NPJ Climate Action* 3, 2. doi:<https://doi.org/10.1038/s44168-023-00074-1>.

Tilman, A.R., Plotkin, J.B., Akçay, E., 2020. Evolutionary games with environmental feedbacks. *Nature Communications* 11, 915. doi:<https://doi.org/10.1038/s41467-020-14531-6>.

Tilman, A.R., Watson, J.R., Levin, S., 2017. Maintaining cooperation in social-ecological systems: effective bottom-up management often requires sub-optimal resource use. *Theoretical Ecology* 10, 155–165. doi:<https://doi.org/10.1007/s12080-016-0318-8>.

Weitz, J.S., Eksin, C., Paarporn, K., Brown, S.P., Ratcliff, W.C., 2016. An oscillating tragedy of the commons in replicator dynamics with game-environment feedback. *Proceedings of the National Academy of Sciences* 113, E7518–E7525. doi:<https://doi.org/10.1073/pnas.1604096113>.

Appendix A.

Appendix A.1. Proof of Lemma

We want to prove that if $a(0) > b(0)$ then $a(t) \geq b(t) \forall t$. We will proceed by induction. The base case ($t = 0$) is given. Assume that for some $t \geq 0$ that $a(t) \geq b(t)$. If $|a(t) - b(t)| > r$:

$$\begin{aligned} a(t+1) &= \frac{\tau_o - 1 + \lambda(1 - \omega)}{\tau_o} a(t) + \frac{\lambda\omega}{\tau_o} a(0) + \frac{1 - \lambda}{\tau_o} h(e(t)) \\ b(t+1) &= \frac{\tau_o - 1 + \lambda(1 - \omega)}{\tau_o} b(t) + \frac{\lambda\omega}{\tau_o} b(0) + \frac{1 - \lambda}{\tau_o} h(e(t)) \\ \implies a(t+1) - b(t+1) &= \frac{\tau_o - 1 + \lambda(1 - \omega)}{\tau_o} (a(t) - b(t)) + \frac{\lambda\omega}{\tau_o} (a(0) - b(0)) \geq 0 \\ \implies a(t+1) &\geq b(t+1) \end{aligned}$$

If $|a(t) - b(t)| \leq r$:

$$\begin{aligned} a(t+1) &= \frac{\tau_o - 1}{\tau_o} a(t) + \frac{\lambda(1 - \omega)}{\tau_o} (pa(t) + (1 - p)b(t)) + \frac{\lambda\omega}{\tau_o} a(0) + \frac{1 - \lambda}{\tau_o} h(e(t)) \\ b(t+1) &= \frac{\tau_o - 1}{\tau_o} b(t) + \frac{\lambda(1 - \omega)}{\tau_o} (pa(t) + (1 - p)b(t)) + \frac{\lambda\omega}{\tau_o} b(0) + \frac{1 - \lambda}{\tau_o} h(e(t)) \\ \implies a(t+1) - b(t+1) &= \frac{\tau_o - 1}{\tau_o} (a(t) - b(t)) + \frac{\lambda\omega}{\tau_o} (a(0) - b(0)) \geq 0 \\ \implies a(t+1) &\geq b(t+1) \end{aligned}$$

Appendix A.2. Proof of Proposition 2

Let $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} < r$. First we want to show that $P(t)$ eventually is lower than r and that once this occurs it is lower than r for all future t . From our assumption we can state that $\exists \epsilon > 0$ such that $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} + \epsilon < r$. If $\exists$ no t such that $P(t) \leq r$ then $\forall t$ we can write:

$$\begin{aligned} P(t) &= \left(\frac{\tau_o - 1 + \lambda(1-\omega)}{\tau_o}\right)^t P(0) + \frac{\lambda\omega P(0)}{\tau_o} \sum_{n=0}^{t-1} \left(\frac{\tau_o - 1 + \lambda(1-\omega)}{\tau_o}\right)^n \\ &= \left(\frac{\tau_o - 1 + \lambda(1-\omega)}{\tau_o}\right)^t \left(P(0) - \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)}\right) + \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} \end{aligned}$$

Since $\left|\frac{\tau_o - 1 + \lambda(1-\omega)}{\tau_o}\right| < 1$ therefore $\lim_{t \rightarrow \infty} P(t) = \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)}$. Therefore for ϵ defined above $\exists T$ such that $P(T) - \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} < \epsilon$ which would imply $P(T) < \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} + \epsilon < r$ which is a contradiction. Therefore $\exists T$ such that $P(T) \leq r$. Therefore $\forall t$:

$$\begin{aligned} p(t+T) &= \left(1 - \frac{1}{\tau_o}\right)^t p(T) + \frac{\lambda\omega P(0)}{\tau_o} \sum_{n=0}^{t-1} \left(1 - \frac{1}{\tau_o}\right)^n \\ &= \left(1 - \frac{1}{\tau_o}\right)^t p(T) + \lambda\omega P(0) \left(1 - \left(1 - \frac{1}{\tau_o}\right)^t\right) \\ &< \left(1 - \frac{1}{\tau_o}\right)^t r + r \left(1 - \left(1 - \frac{1}{\tau_o}\right)^t\right) = r \end{aligned}$$

Therefore $\lim_{t \rightarrow \infty} P(t) = \lim_{t \rightarrow \infty} \left(1 - \frac{1}{\tau_o}\right)^t P(T) + \lambda\omega P(0) \left(1 - \left(1 - \frac{1}{\tau_o}\right)^t\right) = \lambda\omega P(0)$. Let $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} > r$. We want to show that $P(t)$ is always above r . For contradiction assume $\exists T$ such that $P(T) \leq r$. Let T be the smallest such T . This implies that:

$$\begin{aligned} P(T) &= \left(\frac{\tau_o - 1 + \lambda(1-\omega)}{\tau_o}\right)^t \left(P(0) - \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)}\right) + \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} \\ &> \left(\frac{\tau_o - 1 + \lambda(1-\omega)}{\tau_o}\right)^t \left(P(0) - \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)}\right) + r > r \end{aligned}$$

Which is a contradiction. Therefore $P(t) > r \forall t$, which implies that $\lim_{t \rightarrow \infty} P(t) = \lim_{t \rightarrow \infty} \left(\frac{\tau_o - 1 + \lambda(1-\omega)}{\tau_o}\right)^t \left(P(0) - \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)}\right) + \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} = \frac{\lambda\omega P(0)}{1-\lambda(1-\omega)}$.

Appendix A.3. Convergence to consensus in non-prejudiced populations

In non-prejudiced populations, if $P(0) > r$ then $P(t+1) = a(t+1) - b(t+1) = (1 - \frac{1}{\tau_o} + \frac{\lambda}{\tau_o})(a(t) - b(t)) = \frac{\tau_o - 1 + \lambda}{\tau_o} P(t)$. This implies that $P(t) = (\frac{\tau_o - 1 + \lambda}{\tau_o})^t P(0)$. Note that since $\tau_o > 1$ and $\lambda < 1$, $|\frac{\tau_o - 1 + \lambda}{\tau_o}| < 1$. Therefore $P(t) \leq r \iff (\frac{\tau_o - 1 + \lambda}{\tau_o})^t P(0) \leq r \iff t \geq \frac{\ln(r/p(0))}{\ln(\frac{\tau_o - 1 + \lambda}{\tau_o})}$. This means that no matter the initial polarization, the two polarized groups are within the confidence bound of each other at $T = \lfloor \frac{\ln(r/p(0))}{\ln(\frac{\tau_o - 1 + \lambda}{\tau_o})} + 1 \rfloor$.

Appendix B.

Appendix B.1. Proof of Proposition 3

If $c(t) = c(t+1)$ then $c(t) = \frac{\tau_o - 1 + \lambda}{\tau_o} c(t) + \frac{1 - \lambda}{\tau_o} \text{sign}(-\beta + e(t)) \implies c(t) = \text{sign}(-\beta + e(t))$. If $e(t) = e(t+1)$ then $e(t) = (1 - \gamma)e(t) + u(c(t)) \implies e(t) = \frac{1}{\gamma} u(c(t))$. Therefore, $c(t) = \text{sign}(-\beta + \frac{\alpha}{2\gamma}(1 - c(t)))$. Since the sign function can only take on values of -1, 0, or 1 there are only 3 possible fixed points.

If $c(t) = 1$ then $1 = \text{sign}(-\beta)$ which implies that $\beta < 0$. When $c(t) = 1$, then $e(t) = \frac{\alpha}{2\gamma}(1 - 1) = 0$. The dynamics when $\beta < 0$ reduce to $c(t+1) = \frac{\tau_o - 1 + \lambda}{\tau_o} c(t) + \frac{1 - \lambda}{\tau_o}$, $e(t+1) = (1 - \gamma)e(t) + \frac{1}{2}\alpha(1 - c(t))$ whose Jacobian at $(c, e) = (1, 0)$ is: $A = \begin{pmatrix} \frac{\tau_o - 1 + \lambda}{\tau_o} & 0 \\ -\frac{1}{2}\alpha & 1 - \gamma \end{pmatrix}$. The eigenvalues of A are $\frac{\tau_o - 1 + \lambda}{\tau_o}$ and $1 - \gamma$ which are both within the unit circle, implying that $(c, e) = (1, 0)$ is a stable fixed point when $\beta < 0$.

If $c(t) = -1$ then $-1 = \text{sign}(-\beta + \frac{\alpha}{\gamma}) \implies \beta > \frac{\alpha}{\gamma}$. When $c(t) = -1$, then $e(t) = \frac{\alpha}{2\gamma}(1 - (-1)) = \frac{\alpha}{\gamma}$. The Jacobian at $(c, e) = (-1, \frac{\alpha}{\gamma})$ is the same as the case above, $A = \begin{pmatrix} \frac{\tau_o - 1 + \lambda}{\tau_o} & 0 \\ -\frac{1}{2}\alpha & 1 - \gamma \end{pmatrix}$. Therefore, $(c, e) = (-1, \frac{\alpha}{\gamma})$ is a stable fixed point when $\beta > \frac{\alpha}{\gamma}$.

Appendix B.2. Fixed points in prejudiced populations

When $\frac{\lambda\omega P(0)}{1 - \lambda(1 - \omega)} > r$, the steady-state conditions $a(t+1) = a(t)$, $b(t+1) = b(t)$ can be solved since the polarized groups are independent at equilibrium. When $\frac{\lambda\omega P(0)}{1 - \lambda(1 - \omega)} \leq r$, we are able to separate the dynamics of the two groups by assuming that the groups are separated by $\lambda\omega P(0)$, based on the results of Proposition 2. The expressions were simplified by setting $h(e(t)) = 1$ in the mitigative region and $h(e(t)) = -1$ in the non-mitigative regions.

Similarly, the expressions for the environment at equilibrium were evaluated by solving the steady-state condition $e(t+1) = e(t)$ and substituting in the fixed point opinion profile. The oscillation centers were estimated by substituting $e(t) = \beta$ into the expression for the environment and then solving for $a(t)$ and $b(t)$.

When $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} > r$, the long-term behavior in the mitigative and non-mitigative region is:

$$\begin{aligned} a_m(t) &= qa(0) + (1-q) & a_n(t) &= qa(0) - (1-q) \\ b_m(t) &= qb(0) + (1-q) & b_n(t) &= qb(0) - (1-q) \\ e_m(t) &= \frac{\alpha}{2\gamma}q(1 - pa(0) - (1-p)b(0)) \\ e_n(t) &= \frac{\alpha}{2\gamma}(2 - q(1 + pa(0) + (1-p)b(0))) \end{aligned}$$

The environment and opinions in the cyclic region will oscillate around:

$$\begin{aligned} a_c(t) &\approx 1 - \beta \frac{2\gamma}{\alpha} + (1-p)qP(0) \\ b_c(t) &\approx 1 - \beta \frac{2\gamma}{\alpha} - pqP(0) \\ e_c(t) &\approx \beta \end{aligned}$$

When $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} \leq r$, the long-term behavior in the mitigative and non-mitigative region is:

$$\begin{aligned} a_m(t) &= qa(0) + (1-q) - \lambda(1-\omega)(1-p)qP(0) \\ b_m(t) &= qb(0) + (1-q) + \lambda(1-\omega)pqP(0) \\ e_m(t) &= \frac{\alpha}{2\gamma}q(1 - pa(0) - (1-p)b(0)) \\ a_n(t) &= qa(0) - (1-q) - \lambda(1-\omega)(1-p)qP(0) \\ b_n(t) &= qb(0) - (1-q) + \lambda(1-\omega)pqP(0) \\ e_n(t) &= \frac{\alpha}{2\gamma}(2 - q(1 + pa(0) + (1-p)b(0))) \end{aligned}$$

The environment and opinions in the cyclic region will oscillate around:

$$\begin{aligned} a_c(t) &\approx 1 - \beta \frac{2\gamma}{\alpha} + (1 - p)\lambda\omega P(0) \\ b_c(t) &\approx 1 - \beta \frac{2\gamma}{\alpha} - p\lambda\omega P(0) \\ e_c(t) &\approx \beta \end{aligned}$$

Appendix C.

Appendix C.1. Distance from the objective best response

To analyze how far the population is from the 'truth' we can define the distance from the truth T as $d_T = \frac{1}{N} \sum_{i=1}^N |o_i(t) - T|$. In our polarized model this function reduces to $d_T = p|a(t) - T| + (1 - p)|b(t) - T|$. We can only investigate the truth when in the non-mitigative and mitigative regions as the truth is well-defined in those regions. In the cyclic region, the truth is constantly oscillating between mitigation and non-mitigation since the environmental cost continually oscillates above and below the cost of mitigating. We must also separate our analysis dependent on whether $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} > r$.

In the mitigative region $T = 1$. When $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} > r$, we have that $a_m(t) = qa(0) + (1 - q)$ and $b_m(t) = qb(0) + (1 - q)$. Substituting these into d_T we find that: $d_T = q(p|a(0) - 1| + (1 - p)|b(0) - 1|)$. In the non-mitigative region $T = -1$, $a_n(t) = qa(0) - (1 - q)$ and $b_n(t) = qb(0) - (1 - q)$, which implies that $d_T = q(p|a(0) + 1| + (1 - p)|b(0) + 1|)$. In both cases, the distance from the truth increases linearly with the proportion of prejudice, no matter what those prejudice's are. Additionally, a polarized population converges onto the truth if and only if there is no prejudice.

In the case that $\frac{\lambda\omega P(0)}{1-\lambda(1-\omega)} \leq r$, the expressions for the long-term behavior of $a(t)$ and $b(t)$ change. In the mitigative region $T = 1$, $a_m(t) = qa(0) + (1 - q) - \lambda(1 - \omega)(1 - p)qP(0)$, $b_m(t) = qb(0) + (1 - q) + \lambda(1 - \omega)pqP(0)$. This implies $d_T = q(p|a(0) - \lambda(1 - \omega)(1 - p)P(0) - 1| + (1 - p)|b(0) + \lambda(1 - \omega)pP(0) - 1|)$. In the non-mitigative region $d_T = q(p|a(0) - \lambda(1 - \omega)(1 - p)P(0) + 1| + (1 - p)|b(0) + \lambda(1 - \omega)pP(0) + 1|)$. In both cases, if the amount of social susceptibility is kept as a constant, the distance to the truth increases linearly with the proportion of prejudice, no matter what those prejudice's are. As in the previous case, a polarized population converges onto the truth if and only if there is no prejudice.

Appendix D.

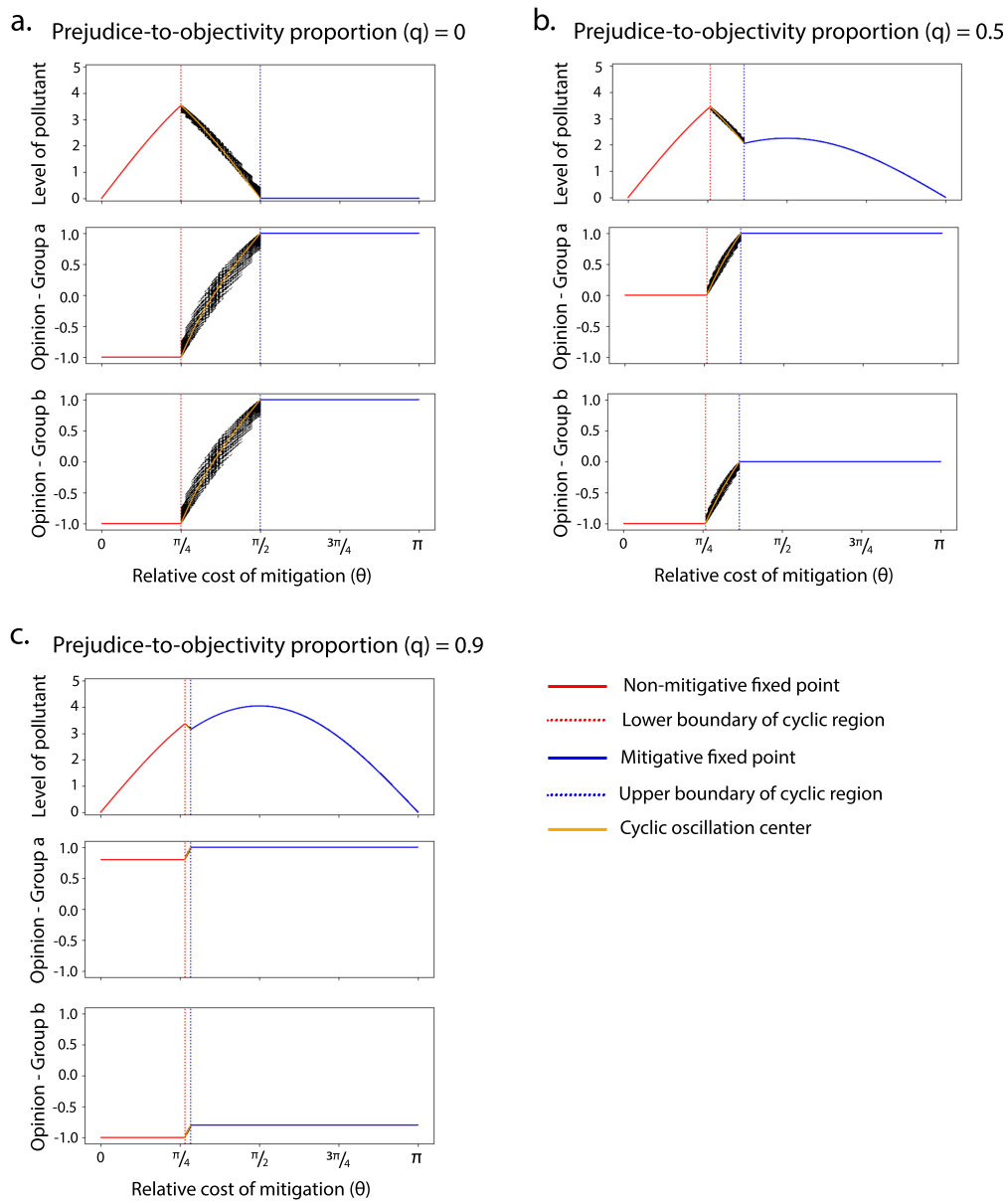


Figure D.1: Validation of fixed point expressions. Overlay of the derived expressions for fixed points and long-term behavior with numerical bifurcation diagrams over the relative cost of mitigation (θ). Expressions show alignment in populations with (a) no prejudice ($q = 0$), (b) equal prejudice and objectivity ($q = 0.5$), and (c) high levels of prejudice ($q = 0.9$).

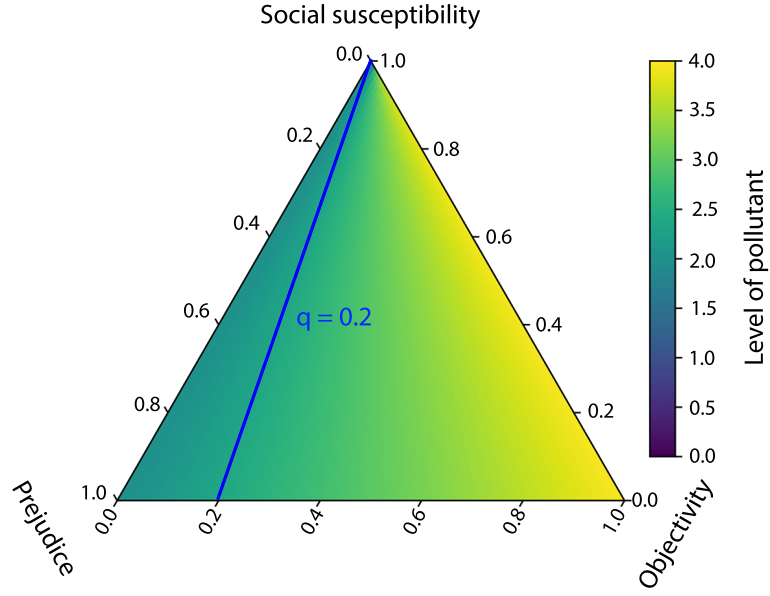


Figure D.2: Effect of influence coefficients on the level of pollutant. Heatmap of the long-term level of pollutant over varying strengths of social susceptibility, prejudice, and objectivity of the population. Simulations were performed at $\alpha = 0.8, \beta = 5, \gamma = 0.2$.