Can Decentralization Improve Social Welfare?

Yexuan Li*

2025-07-15

Abstract

This is the abstract of the paper. Briefly summarize your model, the intuition, and the key findings here.

Keywords:

1 Introduction

2 Basic Model Setup

2.1 Static Discrete Graphical Model

We begin by defining the structure of the model as a graph.

$$\mathcal{G} = (\mathcal{P}, \mathcal{L}, \mathcal{F}, u, c) \tag{1}$$

This graph consists of a set of players $\mathcal{P}$ (you can view them as states now, and later on we can expand our model, so they can also be any participants in a game), links set $\mathcal{L}$, preference distribution set $\mathcal{F}$, some kind of utility function u, and cost function c.

First we consider a discrete model with only two candidates. That is to say, each state has two possible policy, we denote them as 1 and 2. So the set of all possible policy profiles is given by:

$$POL = \{(pol_1, pol_2, \ldots, pol_n) \mid pol_i \in \{0, 1\}\}. \tag{2}$$

We specify the player set as:

$$\mathcal{P} = \{P_1, P_2, P_3, \ldots, P_n\}, \tag{3}$$

in which we have n states in all, and denote $N = \{1, 2, \ldots, n\}$.

The links between players are denoted:

$$\mathcal{L} = \{\overline{P_i P_j} \mid i, j \in \{1, 2, \ldots, n\}, i \neq j\} \tag{4}$$

Each player has a preference drawn from the set:

$$\mathcal{F} = \{(x, q - x) \mid q \in \mathbb{R}^+,\, 0 \leq x \leq q\}. \tag{5}$$

*Tsinghua University, liyexuan23@mails.tsinghua.edu.cn

Here, we denote q as the weight of a state in the decision-making system, which typically correlates with population or economic strength. To simplify the model, we assume all states have equal weight, that is, $q \equiv 1$.

Next we want to assume simple forms of utility and cost function.

Each state's utility depends on its policy choice and preference $f_i = (f_{i,1}, f_{i,2}) \in \mathcal{F}$ for all i:

$$u_i = \begin{cases} f_{i,1} & \text{if } pol_i = 1; \\ f_{i,2} & \text{if } pol_i = 0. \end{cases} \tag{6}$$

The cost between players i and j is defined as:

$$c_{ij} = \begin{cases} 0 & \text{if policy}_i = \text{policy}_j; \\ f_{i,2} + f_{j,1} & \text{if policy}_i = 1,\ \text{policy}_j = 0; \\ f_{i,1} + f_{j,2} & \text{if policy}_i = 0,\ \text{policy}_j = 1. \end{cases} \tag{7}$$

This cost function is designed to simulate the real-world maintenance costs. For example, if cannabis is legal in one state but illegal in its neighboring states, border checkpoints may be established to prevent cannabis flow, or efforts made to appease opposition within the state. Here, we assume that the intensity of the black market and opposition forces is proportional to the state's preference of the opposite policy.

The total social welfare τ is defined as the sum of individual utilities minus the overall costs:

$$\tau = \sum_{i \in N} \left(u_i - \frac{1}{2} {\sum_j}' c_{ij} \right), \tag{8}$$

Here, ${\sum_j}'$ means the summation over all j adjacent to i.

Finally, we define the optimal policy profile as the one that maximizes total social welfare:

$$pol = \arg \max_{pol \in \text{POL}} \tau \tag{9}$$

2.2 Example: Three-State Model

Consider three states with unit weights. Their preferences are:

$$f_1 = (a, 1-a), \quad f_2 = (b, 1-b), \quad f_3 = (c, 1-c), \quad a, b, c \in [0, 1].$$

Assume a complete graph where each pair of states is adjacent.

For each policy profile $pol = (pol_1, pol_2, pol_3) \in \{0, 1\}^3$, the social welfare $\tau_{pol_1 pol_2 pol_3}$ is computed as the sum of utilities minus the sum of costs on mismatched edges.

Case 1: $pol = (1, 1, 1)$

$$u_1 = a, \quad u_2 = b, \quad u_3 = c,$$

$$\text{cost} = 0,$$

$$\tau_{111} = a + b + c. \tag{10}$$

Case 2: $pol = (1, 1, 0)$

$$u_1 = a, \quad u_2 = b, \quad u_3 = 1 - c,$$

Edges with mismatched policies: (1,3) and (2,3),

$$c_{13} = (1 - a) + c, \quad c_{23} = (1 - b) + c,$$

$$\text{cost} = c_{13} + c_{23} = 2 - a - b + 2c,$$

$$\tau_{110} = a + b + (1 - c) - (2 - a - b + 2c) = 2(a + b) - 3c - 1. \tag{11}$$

Case 3: $pol = (1, 0, 1)$

$$u_1 = a, \quad u_2 = 1 - b, \quad u_3 = c,$$

Mismatched edges: (1,2) and (2,3),

$$c_{12} = (1 - a) + b, \quad c_{23} = b + (1 - c),$$

$$\text{cost} = c_{12} + c_{23} = 2b + 2 - a - c,$$

$$\tau_{101} = a + (1 - b) + c - (2b + 2 - a - c) = 2a - 3b + 2c - 1. \tag{12}$$

Case 4: $pol = (0, 1, 1)$

$$u_1 = 1 - a, \quad u_2 = b, \quad u_3 = c,$$

Mismatched edges: (1,2) and (1,3),

$$c_{12} = a + (1 - b), \quad c_{13} = a + (1 - c),$$

$$\text{cost} = c_{12} + c_{13} = 2a + 2 - b - c,$$

$$\tau_{011} = (1 - a) + b + c - (2a + 2 - b - c) = -3a + 2b + 2c - 1. \tag{13}$$

Case 5: $pol = (1, 0, 0)$

$$u_1 = a, \quad u_2 = 1 - b, \quad u_3 = 1 - c,$$

Mismatched edges: (1,2) and (1,3),

$$c_{12} = (1 - a) + b, \quad c_{13} = (1 - a) + c,$$

$$\text{cost} = c_{12} + c_{13} = 2 - 2a + b + c,$$

$$\tau_{100} = a + (1 - b) + (1 - c) - (2 - 2a + b + c) = 3a - 2b - 2c. \tag{14}$$

Case 6: $pol = (0, 1, 0)$

$$u_1 = 1 - a, \quad u_2 = b, \quad u_3 = 1 - c,$$

Mismatched edges: (1,2) and (2,3),

$$c_{12} = a + (1 - b), \quad c_{23} = (1 - b) + c,$$

$$\text{Cost} = c_{12} + c_{23} = 2 - 2b + a + c,$$

$$\tau_{010} = (1 - a) + b + (1 - c) - (2 - 2b + a + c) = -2a + 3b - 2c. \tag{15}$$

Case 7: $pol = (0, 0, 1)$

$$u_1 = 1 - a, \quad u_2 = 1 - b, \quad u_3 = c,$$

Mismatched edges: (1,3) and (2,3),

$$c_{13} = a + (1 - c), \quad c_{23} = b + (1 - c),$$

$$\text{Cost} = c_{13} + c_{23} = a + b + 2 - 2c,$$

$$\tau_{001} = (1 - a) + (1 - b) + c - (a + b + 2 - 2c) = -2a - 2b + 3c. \tag{16}$$

Case 8: $pol = (0, 0, 0)$

$$u_1 = 1 - a, \quad u_2 = 1 - b, \quad u_3 = 1 - c,$$

$$\text{Cost} = 0,$$

$$\tau_{000} = 3 - a - b - c. \tag{17}$$

Preference Conditions for Optimal Policies:

Our goal is to determine, for each policy profile *pol*, the conditions on the preferences a, b, c under which this policy yields the highest total social welfare among all eight possible policies. Specifically, we identify the inequalities that characterize when a given policy is socially optimal.

As an example, consider the policy $pol = (1, 1, 1)$. It is optimal if and only if its social welfare τ_{111} exceeds that of every other policy. Comparing τ_{111} with all others gives the following system of inequalities:

$$\begin{cases} a + b + c > 2(a + b) - 3c - 1, \\ a + b + c > 2a - 3b + 2c - 1, \\ a + b + c > -3a + 2b + 2c - 1, \\ a + b + c > 3a - 2b - 2c, \\ a + b + c > -2a + 3b - 2c, \\ a + b + c > -2a - 2b + 3c, \\ a + b + c > 3 - a - b - c. \end{cases}$$

Solving these yields the preference region where policy $(1,1,1)$ is optimal:

$$\begin{cases} a+b+2c>1, \\ b+c>1, \\ a+c>1, \\ b<2a+c, \\ a<b+c, \\ a+b<c, \\ a+b+c>\frac{3}{2}. \end{cases}$$

Based on the derived conditions for $pol = (1,1,1)$ to be optimal, here are three example preference triples (a,b,c) that satisfy the inequalities:

$$\begin{cases} a=0.6, \quad b=0.8, \quad c=0.5; \\ a=0.7, \quad b=0.9, \quad c=0.4; \\ a=0.8, \quad b=1.0, \quad c=0.3. \end{cases}$$

2.3 Model Expansion

To clearly illustrate the preference requirements under which each policy configuration is socially optimal, we visualize the eight possible policies in different colors within the three-dimensional space $(a,b,c) \in [0,1]^3$.

From Figure 1, we observe that when a, b, and c are all nonzero, only two policies can be optimal: $pol = (1,1,1)$ or $pol = (0,0,0)$. Moreover, these two regions are separated by a clear boundary defined by $a+b+c = 1.5$.

This phenomenon arises because the cost function in our model dominates the utility structure. Therefore, we need to consider the following modifications to the model.

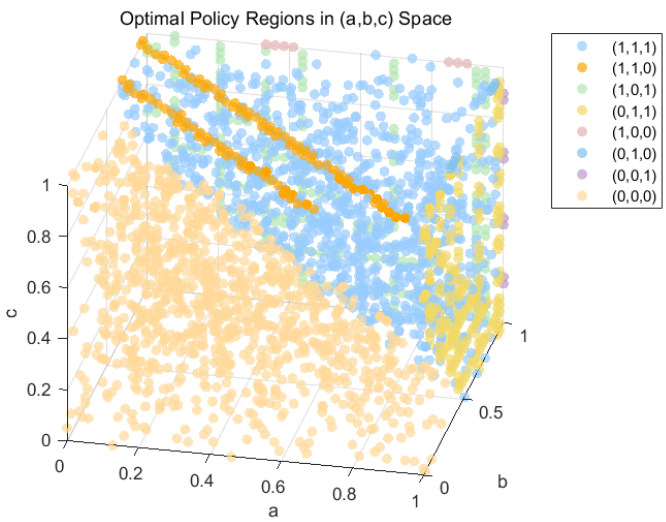

Figure 1: Regions Corresponding to Different Optimal Policy Profiles

To reflect varying levels of stabilization cost across states, we introduce an influence coefficient $b_{ij} \in [0,1]$ for each edge (i,j) in the cost function. Here, we suppose $b_{ij} = b_{ji}$. Specifically, we modify the original cost function c_{ij} by multiplying it with b_{ij}:

$$c'_{ij} = b_{ij} \cdot c_{ij}, \tag{18}$$

where $b_{ij} = 0$ indicates that there is no stabilization cost between state i and state j (i.e., no incentive to align their policies and they are independent of each other), and $b_{ij} = 1$ recovers the original model with full stabilization cost applied.

The new formulation of total social welfare becomes:

$$\tau = \sum_{i \in N} \left(u_i - \frac{1}{2} {\sum_j}' b_{ij} \cdot c_{ij} \right). \tag{19}$$

This refinement allows us to interpolate between two extremes:

- When $b_{ij} \equiv 0$, the stabilization cost disappears, and the optimal policy is for each state to choose its locally preferred policy (i.e., a fully decentralized system).
- When $b_{ij} \equiv 1$, the original model is recovered, where stabilization costs are significant and may incentivize more centralized or coordinated policies.

In this way, the parameter b_{ij} provides a tunable mechanism to analyze how the strength of inter-state frictions affects the trade-off between local autonomy and centralized policy coordination.

Additionally, we introduce the decentralization benefit μ, which measures the difference in the maximum total social welfare between a decentralized policy and a centralized one under a given initial preference profile.

Mathematically, μ, which depends on a graph, is defined as:

$$\mu(\mathcal{G}) = \max_{pol \in POL, pol \notin \{\mathbf{0}, \mathbf{1}\}} \tau_{pol} - \max_{pol \in POL, pol \in \{\mathbf{0}, \mathbf{1}\}} \tau_{pol}, \tag{20}$$

where $\mathbf{0}$ denotes the policy profile that every regional policy is 0 and $\mathbf{1}$ for 1.

This quantity μ captures the welfare gain (or loss) from decentralization relative to centralization.

We now briefly explain how our model fits into the broader class of network utility-cost frameworks. Our model emphasizes the role of border costs, which arise due to mismatches in local policies between adjacent nodes, and uses them to compute various network-level quantities such as total social welfare and decentralization benefit.

In more general models—such as those used in social network analysis—there is often an additional focus on each point's reaction to the surrounding environment. This is usually captured via local cost terms that depend on the point's alignment with its neighbors or some aggregate signal from the network.

To facilitate potential extensions of our framework, we show below that, under a simple linear cost function that depends on individual preferences, these local response costs differ from our model only by a constant multiplicative factor. Therefore, our model can be seen as a special case of the general framework with linear preference-aligned local cost functions.

Of course, more sophisticated extensions are possible. For example, one may introduce quadratic costs or cross terms. In these cases, additional cost terms can be added directly into the social welfare formula. While the network behavior may shift depending on the cost function, the overall modeling methodology remains the same.

We now introduce a local cost function at each point to capture the player's dissatisfaction with its own policy choice. Specifically, for each point $i \in N$, its local cost is defined as:

$$c_i = q_i - u_i, \tag{21}$$

where $q_i \in \mathbb{R}^+$ is, as defined above, the player's weight.

Now total social welfare should be:

$$\tau = \sum_{i \in N} \left(u_i - \frac{1}{2} \sum_j{}' b_{ij} \cdot c_{ij} - c_i \right). \tag{22}$$

Substituting $c_i = q_i - u_i$, we obtain:

$$\begin{aligned}
\tau &= \sum_{i \in N} \left(u_i - \frac{1}{2} \sum_j{}' b_{ij} \cdot c_{ij} - (q_i - u_i) \right) \\
&= \sum_{i \in N} \left(2u_i - q_i - \frac{1}{2} \sum_j{}' b_{ij} \cdot c_{ij} \right) \\
&= \sum_{i \in N} 2u_i - \sum_{i \in N} q_i - \sum_{i \in N} \frac{1}{2} \sum_j{}' b_{ij} \cdot c_{ij}.
\end{aligned} \tag{23}$$

This shows that the general model differs only by a linear scaling factor 2 before u_i and a constant offset $\sum_{i \in N} q_i$.

2.4 Example: Will decentralization enhance total social welfare?

To illustrate the behavior of the model under partial stabilization cost, we consider a simple setting with three states, whose preference profiles are given by $a = 0.8$, $b = 0.6$, and $c = 0.1$. We assume that all inter-state links are fully connected and the stabilization cost coefficient is uniform across all links: $b_{ij} = 0.5$ for all $i \neq j$.

Under this setting, we exhaustively evaluate the total social welfare τ for all $2^3 = 8$ possible policy combinations. The results indicate that the optimal policy profile is:

$$pol^* = (1, 1, 0),$$

with a corresponding maximum social welfare of

$$\tau_{110} = 1.90.$$

We further compare this outcome with the best centralized policy, i.e., either $(0, 0, 0)$ or $(1, 1, 1)$. In this case, the best centralized welfare is achieved by the profile $(1, 1, 1)$, yielding

$$\tau_{111} = 1.40.$$

Therefore, the decentralization benefit μ is computed as:

$$\mu = \tau_{110} - \tau_{111} = 0.40.$$

This example demonstrates that under partial stabilization cost, allowing states to diverge in their policies can produce a measurable welfare gain over a fully centralized arrangement.

2.5 Analysis of Algorithm

Is this problem NP-hard?

2.6 Dynamic Analysis

We now extend our static model into a dynamic framework in order to analyze evolving public preferences, electoral competition, and strategic policy adjustments over time.

From the previously defined total social welfare function τ, we first assume that the influence coefficients b_{ij} are independent of time t. However, the border cost c_{ij} and the utility function u_i both depend on the preference distribution f_i, which may evolve over time.

Therefore, the source of dynamics in our model lies in the fact that preferences f_i are endogenous and influenced by past policy choices or resource investments. Formally, we consider:

$$\tau(t) = \sum_{i \in N} \left(u_i(t) - \frac{1}{2} \sum_j{}' b_{ij} \cdot c_{ij}(t) \right), \tag{24}$$

where both $u_i(t)$ and $c_{ij}(t)$ are functions of the time-varying preference profile $f_i(t)$. We fix b_{ij} as constants over time, and study how the endogenous evolution of preferences drives the whole graph.

2.6.1 Constant Speed Evolution

We begin our dynamic analysis with the simplest scenario: constant speed evolution of preferences over time. This formulation serves as a benchmark and highlights how local policy choices gradually shape long-term preferences.

First, we clarify the interpretation of preference distributions. Under normalization, the preference $f_i(t) = (f_{i,1}(t), f_{i,2}(t))$ can be viewed as the proportion of population in state i supporting policies 1 and 2, respectively, at time t.

In this setting, we assume that when a state implements policy $pol_i(t) \in \{1, 2\}$, the share of the population supporting this policy increases gradually, while support for the opposing policy decreases by the same amount, maintaining normalization. We define the preference dynamics as:

$$\begin{cases} \dot{f}_{i,1}(t) = (-1)^{pol_i(t)+1} \cdot \eta, \\ \dot{f}_{i,2}(t) = (-1)^{pol_i(t)} \cdot \eta, \end{cases} \tag{25}$$

where $\eta > 0$ is a constant representing the speed of preference updating.

However, this simple model has a very limited scope of applicability. To ensure that each component of the preference vector remains within the normalized range $[0, 1]$, the above evolution rule is only valid over short period. Otherwise, linear accumulation will eventually cause values to exceed the valid range.

This is also intuitively reasonable. Public opinion shifts exhibit diminishing marginal effects over time, and parties will not continue to invest resources in states that are already firmly support them (so-called "safe red" or "safe blue" states).

2.6.2 Diminishing Marginal Speed Evolution

As analyzed above, we now consider the case of diminishing marginal speed. Mathematically, we can incorporate a convergence factor into the previous differential equation. Specifically, we have formula as:

$$\begin{cases} \dot{f}_{i,1}(t) = (-1)^{pol_i(t)+1} \cdot \eta \cdot \left(\frac{1}{2} + (-1)^{pol_i(t)+1}\left(\frac{1}{2} - f_{i,1}(t)\right)\right), \\ \dot{f}_{i,2}(t) = (-1)^{pol_i(t)} \cdot \eta \cdot \left(\frac{1}{2} + (-1)^{pol_i(t)}\left(\frac{1}{2} - f_{i,2}(t)\right)\right). \end{cases} \tag{26}$$

For clarity, we simplify the equations above separately for the cases when $pol_i(t) = 1$ and $pol_i(t) = 2$:

$$\begin{cases} \dot{f}_{i,1}(t) = \eta \cdot (1 - f_{i,1}(t)), \\ \dot{f}_{i,2}(t) = -\eta \cdot f_{i,2}(t), \end{cases} \quad \text{if } pol_i(t) = 1; \tag{27}$$

$$\begin{cases} \dot{f}_{i,1}(t) = -\eta \cdot f_{i,1}(t), \\ \dot{f}_{i,2}(t) = \eta \cdot (1 - f_{i,2}(t)), \end{cases} \quad \text{if } pol_i(t) = 2. \tag{28}$$

2.6.3 Campaign Investment

Finally, we consider the effect of campaign investments. Suppose there are exactly two parties A and B, supporting policy 1 and 2 respectively. The campaign resources invested in state i by parties A and B are denoted by $m_{i,1}$ and $m_{i,2}$. We incorporate these investments as additional terms in the differential equations for preference evolution:

$$\begin{cases} \dot{f}_{i,1}(t) = \cdots + \frac{m_{i,1} - m_{i,2}}{m_{i,1} + m_{i,2}} \cdot \xi_i, \\ \dot{f}_{i,2}(t) = \cdots + \frac{m_{i,2} - m_{i,1}}{m_{i,1} + m_{i,2}} \cdot \xi_i, \end{cases} \tag{29}$$

where ξ_i represents the speed at which campaign investment influences the evolution of preferences in state i.

3 Model Application

3.1 Division of Electoral District and Best-scale Problem

Beyond its mathematical structure, our model offers insights into real-world phenomena such as division of electoral district and formation of tribes. From the perspective of a centralized social planner aiming to maximize the total social welfare τ, one natural strategy is to minimize the border cost that arises from policy misalignments across connected players. When we consider cross-term cost structures of the form $b_{ij} \cdot c_{ij}$, it becomes clear that we can both reduce the influence coefficient b_{ij} and the opposite preference c_{ij}.

We now briefly examine how this framework sheds light on these factors:

- **Geography:** We observe that in many countries, administrative boundaries are often drawn along natural geographic features such as mountain ranges, rivers, or deserts. These features serve as physical barriers that limit interaction and integration between regions.

  In the context of our model, such geographic separation effectively reduces the stabilization cost between regions by lowering the influence coefficient b_{ij}. That is, where travel and communication are difficult due to geographic constraints, there is less need for coordinated policy.Therefore, by placing boundaries along geographic features, central planners can indirectly reduce total border costs without altering preferences or policies.

- **Culture and Languages:** For similar reasons, many administrative or political boundaries are aligned with cultural and linguistic differences. When regions differ significantly in language, religion, customs, or historical identity, maintaining a unified policy may become costly and ineffective.

  In our model, these "pure" tribes have very low opposite preferences, that is $|f_{ij} - f_{ji}| \approx 1$, which in turn decreases c_{ij} if thet have different policy.

  This also effectively lowers b_{ij} across culturally distinct regions, as their interaction would be very difficult.

These examples illustrate how our framework can serve as a theoretical foundation for analyzing institutional design problems about political coordination.

Finally, we would like to explore the problem of optimal scale, namely: in a natural network, what is the appropriate size for organizing subgroups? Here, we introduce a local cost term representing the administrative cost of governance inherent to each region. Our considerations are as follows: when the scale is small, internal preferences tend to be more unified, resulting in lower boundary costs; however, the total cost increases as the number of such groups grows. Conversely, when the scale is large, internal preferences become more complex, leading to higher boundary costs, but the administrative cost per region decreases. Therefore, there exists an optimal scale that balances these trade-offs. Our next step is to simulate or calculate this optimal scale and compare it with actual administrative divisions observed in reality.

3.2 Competition Strategy

As previously discussed, centralized planning may lead to policy distortions at the local level, that is, the final policy in one region may not reflect the majority preference of its residents. Although our earlier analysis used a relatively extreme setting where $b_{ij} \equiv 1$, the distortion phenomenon can arise in a broad range of stabilization cost configurations.

This observation motivates a new question: assuming the presence of multiple parties within each state, what strategic behavior might they adopt under the centralized decision-making framework?

Suppose certain parties forecast that they will inevitably lose in a particular state or are guaranteed to win. Would they then have incentives to transfer some of their supporter or funds to alter situation in other states?

This leads us to a game theory extension of our model. Instead of treating preferences as fixed parameters, we now introduce the possibility of endogenous manipulation: parties may adjust their distribution of preferences by reallocating voters inside or across regions. Such idea raises several questions:

- What is the optimal voter transfer strategy for a given party, conditional on other parties' actions?
- Under what network and cost conditions is voter manipulation feasible?
- Can such strategies increase or decrease overall social welfare τ?

In this part, we will give a definition of strategic opinion transfer in electoral competition. Here, we refer to the ability of two competing parties to manipulate the effective preference distribution by either incentivizing their voters to misreport preferences or by reallocating voters across districts. By altering the preference distribution of a state, one can change both the boundary cost and utility, which in turn may lead to a change in the optimal policy in purpose of maximizing social welfare.

3.2.1 Misreporting

We first illustrate this possibility through a simple example involving a single edge between two states, and then we would give a more mathematical form.

Example 3.1 (Misreporting in Two-state Model)**.** *Let state* 1 *have a preference of* $f_{1,1} = 0.9$ *for policy 1, and state* 2 *have a preference of* $f_{2,1} = 0.4$*. Let* b_{12} *denote the border influence coefficient. We now enumerate the four possible policy profiles for the two states and compute the corresponding social welfare* τ*. Most importantly, here we consider party A misreporting some of its supporters in state* 1*, whose number is* Δ *.*

Case 1: $pol = (1, 1)$

$$\tau_{11} = 0.9 + 0.4 = 1.3, \tag{30}$$

$$\tau'_{11} = 0.9 - \Delta + 0.4 = 1.3 - \Delta. \tag{31}$$

Case 2: $pol = (0, 0)$

$$\tau_{00} = 0.1 + 0.6 = 0.7, \tag{32}$$

$$\tau_{00} = 0.1 + \Delta + 0.6 = 0.7 + \Delta. \tag{33}$$

Case 3: $pol = (1, 0)$

$$\tau_{10} = 0.9 + 0.6 - b \cdot (0.1 + 0.4) = 1.5 - 0.5b, \tag{34}$$

$$\tau'_{10} = 0.9 - \Delta + 0.6 - b \cdot (0.1 + \Delta + 0.4) = 1.5 - 0.5b - (1 + b) \cdot \Delta. \tag{35}$$

Case 4: $pol = (0, 1)$

$$\tau_{01} = 0.1 + 0.4 - b \cdot (0.9 + 0.6) = 0.5 - 1.5b, \tag{36}$$

$$\tau'_{01} = 0.1 + \Delta + 0.4 - b \cdot (0.9 - \Delta + 0.6) = 0.5 - 1.5b + (1 + b) \cdot \Delta. \tag{37}$$

We wonder when the optimal policy profile would shift from $(1, 0)$ *to* $(1, 1)$*. We need to calculate inequalities below:*

$$
\begin{cases}
1.5 - 0.5b > 1.3, \\
1.5 - 0.5b > 0.7, \\
1.5 - 0.5b > 0.5 - 1.5b, \\
\quad 1.3 - \Delta > 0.7 + \Delta, \\
\quad 1.3 - \Delta > 1.5 - 0.5b - (1 + b)\Delta, \\
\quad 1.3 - \Delta > 0.5 - 1.5b + (1 + b)\Delta.
\end{cases}
\tag{38}
$$

The results are:

$$
\begin{cases}
b < 0.4, \\
\Delta < 0.3, \\
0.5b + b\Delta > 0.2, \\
0.8 + 1.5b > (2 + b)\Delta.
\end{cases}
\tag{39}
$$

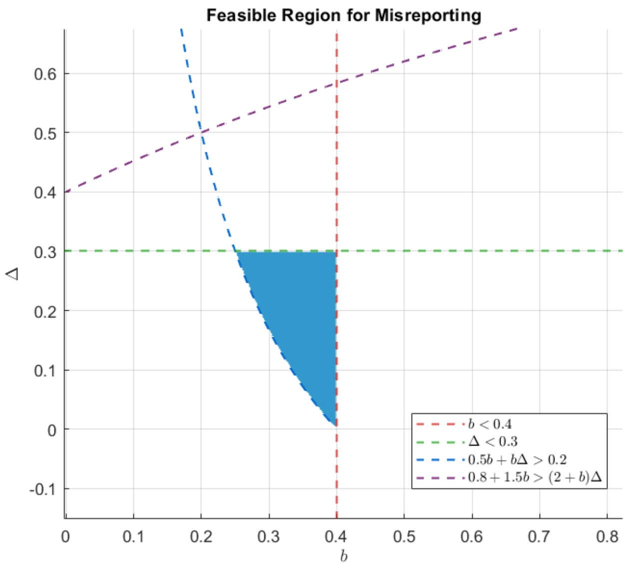

Figure 2: Feasible Region for Misreporting in $\Delta - b$ space

From Figure 2, we can observe that under the current parameter setting, there indeed exists a region where Party A can strategically misreport part of its supporters' preferences to influence the border cost and consequently change the optimal policy in a neighboring state.

3.2.2 Transformation between states

We now consider the possibility of strategic manipulation through voter transformation between states. Specifically, we define how such manipulation alters local preferences and how it may affect the overall social welfare τ under different policies.

We still use two-state model like before. Let $f_{1,1}$ denote the fraction of voters in state 1 who prefer Policy 1, and $f_{2,1}$ denote the same in state 2. If a mass Δ of policy 1 supporters are transferred from state 1 to state 2, the adjusted preferences are defined as follows:

$$\begin{cases} f'_{1,1} = \dfrac{f_{1,1}-\Delta}{f_{1,1}-\Delta+f_{1,2}} = \dfrac{f_{1,1}-\Delta}{1-\Delta}, \\ f'_{1,2} = \dfrac{f_{1,2}}{f_{1,1}-\Delta+f_{1,2}} = \dfrac{f_{1,2}}{1-\Delta}, \\ f'_{2,1} = \dfrac{f_{2,1}+\Delta}{f_{2,1}+\Delta+f_{2,2}} = \dfrac{f_{2,1}+\Delta}{1+\Delta}, \\ f'_{2,2} = \dfrac{f_{2,2}}{f_{2,1}+\Delta+f_{2,2}} = \dfrac{f_{2,2}}{1+\Delta}. \end{cases} \tag{40}$$

Example 3.2 (Voter Transformation in Two-state Model)**.** *Suppose initially $f_{1,1} = 0.9$ and $f_{2,1} = 0.4$, and let $\Delta > 0$ be the amount of voters moved from state 1 to state 2. We now calculate new preference distributions and total social welfare τ under four policy profiles:*

$$\begin{cases} f'_{1,1} = \dfrac{0.9-\Delta}{1-\Delta}, \\ f'_{1,2} = \dfrac{0.1}{1-\Delta}, \\ f'_{2,1} = \dfrac{0.4+\Delta}{1+\Delta}, \\ f'_{2,2} = \dfrac{0.6}{1+\Delta}. \end{cases} \tag{41}$$

Case 1: *pol* $= (1, 1)$

$$\tau_{11} = 0.9 + 0.4 = 1.3, \tag{42}$$

$$\tau'_{11} = \frac{0.9-\Delta}{1-\Delta} + \frac{0.4+\Delta}{1+\Delta}. \tag{43}$$

Case 2: *pol* $= (0, 0)$

$$\tau_{00} = 0.1 + 0.6 = 0.7, \tag{44}$$

$$\tau'_{00} = \frac{0.1}{1-\Delta} + \frac{0.6}{1+\Delta}. \tag{45}$$

Case 3: *pol* $= (1, 0)$

$$\tau_{10} = 0.9 + 0.6 - b \cdot (0.1 + 0.4) = 1.5 - 0.5b, \tag{46}$$

$$\tau'_{10} = \frac{0.9-\Delta}{1-\Delta} + \frac{0.6}{1+\Delta} - b \cdot \left(\frac{0.1}{1-\Delta} + \frac{0.4+\Delta}{1+\Delta}\right). \tag{47}$$

Case 4: *pol* $= (0, 1)$

$$\tau_{01} = 0.1 + 0.4 - b \cdot (0.9 + 0.6) = 0.5 - 1.5b, \tag{48}$$

$$\tau'_{01} = \frac{0.1}{1-\Delta} + \frac{0.4+\Delta}{1+\Delta} - b \cdot \left(\frac{0.9-\Delta}{1-\Delta} + \frac{0.6}{1+\Delta}\right). \tag{49}$$

We also wonder when the optimal policy profile would shift from $(1,0)$ *to* $(1,1)$. *We need to calculate inequalities below:*

$$\begin{cases} 1.5 - 0.5b > 1.3, \\ 1.5 - 0.5b > 0.7, \\ 1.5 - 0.5b > 0.5 - 1.5b, \\ \dfrac{0.9-\Delta}{1-\Delta} + \dfrac{0.4+\Delta}{1+\Delta} > \dfrac{0.1}{1-\Delta} + \dfrac{0.6}{1+\Delta}, \\ \dfrac{0.9-\Delta}{1-\Delta} + \dfrac{0.4+\Delta}{1+\Delta} > \dfrac{0.9-\Delta}{1-\Delta} + \dfrac{0.6}{1+\Delta} - b \cdot \left(\dfrac{0.1}{1-\Delta} + \dfrac{0.4+\Delta}{1+\Delta} \right), \\ \dfrac{0.9-\Delta}{1-\Delta} + \dfrac{0.4+\Delta}{1+\Delta} > \dfrac{0.1}{1-\Delta} + \dfrac{0.4+\Delta}{1+\Delta} - b \cdot \left(\dfrac{0.9-\Delta}{1-\Delta} + \dfrac{0.6}{1+\Delta} \right). \end{cases} \tag{50}$$

The results are:

$$\begin{cases} b < 0.4, \\ \dfrac{0.8-\Delta}{1-\Delta} + \dfrac{-0.2+\Delta}{1+\Delta} > 0, \\ (1+b) \cdot \dfrac{0.4+\Delta}{1+\Delta} + b \cdot \dfrac{0.1}{1-\Delta} > \dfrac{0.6}{1+\Delta}, \\ (1+b) \cdot \dfrac{0.9-\Delta}{1-\Delta} + b \cdot \dfrac{0.6}{1+\Delta} > \dfrac{0.1}{1-\Delta}. \end{cases} \tag{51}$$

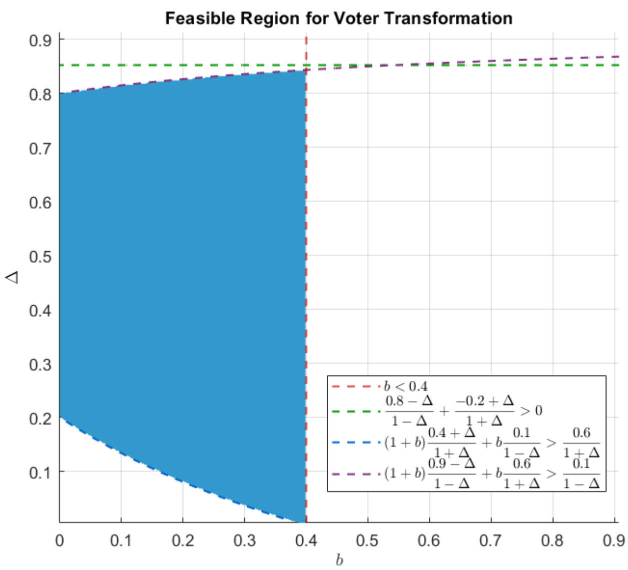

Figure 3: Feasible Region for Voter Transformation in $\Delta - b$ space

From Figure 3, we can observe that under the current parameter setting, there indeed exists a region where Party A can strategically transfer part of its supporters' preferences to influence the border cost and consequently change the optimal policy in a neighboring state.

3.2.3 Equilibrium under Strategic Preference Manipulation

We now provide a unified game theory framework for strategic competition by parties through preference misreporting or voter transformation.

Consider a finite set of states $\mathcal{P} = \{P_1, \dots, P_n\}$, and m parties $Party = \{p_1, \dots, p_m\}$ competing among m policies, that is, $POL = \{(pol_1, \dots, pol_n) \mid pol_i \in \{1, 2, \dots, m\}\}$. Each party $p_j \in Party$ has some voters with fixed total population mass $n_j = \sum_{i \in N} f_{i,j}$. Within each region $i \in N = (1, 2, \dots, n)$ and $j \in M = (1, 2, \dots, m)$, a fraction of voters $f_{i,j} \in [0, 1]$ prefer policy j, so $\sum_{j \in M} f_{i,j} = q_i$.

Definition 3.1 (Electoral Strategy). *Each party can redistribute its voters either by:*

- ***Misreporting:*** *let voters in region i report other preferences in state i;*
- ***Transformation:*** *moving a mass $\Delta_{ij}^p > 0$ of voters from state i to j supporting for policy $p \in M$.*

Remember that we have defined a preference distribution $f_i = (f_{i,1}, f_{i,2}, \dots, f_{i,m})$. In another view, every component of this vector, for example f_{ij}, is its strategy in state i of party p_j. So we let $s^j \in \mathcal{S} = \{(f_{1,j}, f_{2,j}, \dots, f_{n,j}) \mid f_{i,j} \geq 0 \quad \text{for all} \quad i \in N\}$ denotes the preference choice of party j. Given strategies $s = (s^1, s^2, \dots, s^m)$, the central planner selects a policy vector $pol = (pol_1, \dots, pol_n) \in \{1, 2, \dots, m\}^n$ to maximize social welfare τ.

Definition 3.2 (Party's Utility). *We define the **utility of party** p_j as:*

$$U^j(s^j, s^{-j}) := \sum_{i \in N} g_i^j \cdot \mathbb{1}_{\{pol_i = j\}}, \tag{52}$$

where

- *$\mathbb{1}_{\{pol_i = j\}}$ is the indicator function that equals 1 if the ith component of optimal policy profile equals j, and 0 otherwise*
- *$g_i^j \in [0, 1]$ is a weight representing the importance of wining state i for party j*
- *$s^j \in \mathcal{S}$ is the strategy chosen by party j, and s^{-j} denotes the collection of strategies of all other parties*

Definition 3.3 (Nash Equilibrium). *We define a **Nash equilibrium under strategy competition** as a profile of strategies $(s^{1*}, s^{2*}, \dots, s^{m*}) \in \underbrace{\mathcal{S} \times \mathcal{S} \times \cdots \times \mathcal{S}}_{m \text{ times}} = \mathcal{S}^m$ such that:*

$$U^j(s^{j*}, s^{-j*}) \geq U^j(s^j, s^{-j*}), \quad \forall s^j \in \mathcal{S}, \quad \forall j \in M. \tag{53}$$

That is, no party has an incentive to change its strategy, given other parties' choices. This equilibrium reflects the stable manipulation pattern where no further misreporting or transformation would shift the central planner's optimal policy profile.

next check this equil about τ is optimal competition.

Strategic agents may cause inefficiencies.

3.3 Other Applications

4 Discussions