The Interaction of Communication, Social Preferences, and Inequality: Model and Experiment

Kirill Zhazhin*

September 21, 2024

Work in progress
(Click here for the latest version.)

Abstract

In the absence of information transmission, how does communication affect voter behavior? We propose a model where agents can persuade others by shaping their social preferences. We incorporate the competing notions of inequality aversion, efficiency, and fairness into agents' distributional preferences. In a public goods-like game, we demonstrate that when messaging is costly, and initial wealth distribution is highly unequal, the inequality is perpetuated by efficiency-promoting messages from wealthier agents. We design a laboratory experiment to test how such limited access to communication shapes participant's social preferences, conditioned on the perceived fairness of the initial distribution (luck or effort), the varying cost of messages (free or costly), and the presence of identity cues (revealed or hidden).

*University of California, Santa Cruz (email: kzhazhin@ucsc.edu). Adviser: Kristian Lopez Vargas (email: klopezva@ucsc.edu). While this draft is solely authored by Kirill Zhazhin, it draws upon many contributions to a larger joint project of Kristian Lopez Vargas and Louis Putterman (email: louis_putterman@brown.edu). I acknowledge the draft's place as part of this work by using pronouns *We* and *our* instead of *I* and *mine* throughout the paper. Mistakes, however, are mine and mine alone.

1 Introduction

Even a polarized society can be united around some values. In the U.S., a large majority of Democrats and Republicans believe in individual freedoms, feel sympathetic toward people experiencing poverty, and think that the government spends too little on infrastructure.[1] Yet, once the issue becomes more granular, be it supporting a new infrastructure bill, approving welfare spending increases, or voting for redistribution programs, the groups diverge dramatically. Given that people report their fundamental values truthfully, why does the same set of values fail to prevent the divergence in opinions and behavior?

Access to information has been a prominent explanation, both in theoretical (Kamenica 2019) and applied (Mengel & Weidenholzer 2023) literature. The arguments are straightforward and believable. Senders, endowed with some private information and represented by media (e.g., Enikolopov et al. 2011), politicians (e.g., Roemer 1998), or fellow voters (e.g., Iaryczower et al. 2018), can induce a beneficial shift in receiver's beliefs by selectively revealing their information. Conversely, voters can be biased against a policy that predisposes them to manipulation (Jeong 2019) and biased updating (Coutts 2019). These arguments, however, exclude political messages that do not contain private information and instead aim to persuade voters differently – by manipulating the way the fundamental values are incorporated into their decisions.

This paper studies how voters use communication to induce a particular social preference structure in the utility of others without transmitting any information about the outcomes. The context for our study is a society with a highly unequal initial distribution of wealth deciding on a tax to fund public good creation and some redistribution program. We use this setting to develop a model that includes three competing notions of social preferences: *inequality aversion*, *efficiency preference*, and *fairness*. The model also allows voters to engage in an open-forum discussion by sending messages, which could be costly, and receiving others' messages at no cost. We then design a laboratory experiment to identify the impact of communication on the role social preferences play in voters' decision-making and subsequent distributional outcomes for society as a whole.

We position our model and design in the context of two prominent features of the modern political landscape. First, there is a clear abundance of political messages that aim to reshape voters' priorities. Take, for example, ad campaigns against several recent

[1] The statement draws on a variety of results from the 2021 American Values Survey (Jones et al. 2021), 2023 AP-NORC poll (AP-NORC 2023), and 2019 Cato Welfare Work Wealth Survey Report (Ekins 2019).

sugar-sweetened beverage taxes in referendums in the U.S. Anti-tax messengers recognized that voters have a general preference for social good but no certainty on how it should be incorporated into their opinions about the specific tax. Many ads highlighted statements such as "The beverage tax (...) does nothing to solve our real challenges, like creating jobs, reducing crime and improving schools" (Marriott III & Dillard 2021). This message is clearly uninformative [2] and constructed to convince voters that social preferences should only factor into their decision if the policy would impact some specific dimensions. It induces voters to integrate the societal benefits of a policy into their thinking in a way that completely disregards some dimensions and prioritizes others while maintaining the claim of caring about a general idea of social good. We reflect this in our model and experimental design by allowing a prominent message to induce either a preference for *efficiency*, *fairness*, or *inequality aversion*. This aspect relies on a crucial assumption that every voter desires to care about the social, and specifically distributional, aspects of the after-tax outcome but can be easily swayed in how exactly this preference manifests in their utility.[3]

The second feature of the modern political landscape we highlight is unequal access to communication driven by economic inequality. On a primitive level, the cost of actively engaging might be too high for poor voters. The cost of buying ads, making contributions, and engaging in advocacy represents a more significant portion of their income. Moreover, decreasing the marginal utility of every dollar implies that they lose more utility for the same absolute cost. Prior literature has shown that wealth concentration could lead to media capture (Corneo 2006) and how imposing even a tiny message cost could dramatically reduce message use (Kriss et al. 2016). Our model and experimental design feature communication in an open forum, which is a meaningfully different medium as compared to the media and targeted messaging commonly used in the literature. In fact, we postulate that our setup is closer to some settings where voters engage in political communication with each other.

A salient example of the role of wealth in open forum communication is the prioritization of paid users' posts and comments on social platforms like Twitter [4] (Vincent

[2] It is hard to imagine any voter believing that a beverage tax could reduce crime and getting persuaded otherwise by this ad. However, we admit that such messages could be aimed to take advantage of receivers' limited attention (Maćkowiak et al. 2023). Over-saturating the information set with such messages could prevent informative signals from reaching the voters. This is plausible but doesn't explain why, out of all uninformative messages, these exact ones are persistently chosen by advocates and identified as persuasive by stakeholders (Jou et al. 2014).

[3] it is possible to somewhat relax this assumption by allowing for heterogeneity in voter types and starting positions and adding some rigidity in the amount of communication required to shift from one preference structure to another. We plan to explore this in the future.

[4] Also referred to as X in an ongoing re-branding attempt I do not wish to participate in.

2022). For instance, replies made by free users can be demoted and sometimes even hidden, which effectively limits the ability to communicate. The subscribers, in contrast, often enjoy greater reach and algorithmic priority. This context makes an open forum setting attractive and prime for exploration. Therefore, we reflect its core characteristics in our experimental design.

Our main theoretical finding is that wealth inequality can perpetuate itself through communication, even if it is available to everyone. This occurs because wealthier agents face relatively lower cost while expecting higher gain, thus it makes them more likely to engage in communication and try to influence the social preferences of others. They can shift the collective decision-making in a direction that favors efficiency, thereby preserving inequality in the final outcome. The poorer agents, facing higher relative costs for communication, are less likely to participate, which further tilts the balance of influence towards the wealthy. This dynamic creates a feedback loop where inequality in wealth leads to inequality in communication, which in turn leads to further entrenchment of wealth inequality.

Our main experimental contribution is the design of an experiment that would allow us to systematically test how communication, wealth inequality, and social preferences interact to shape voting behavior and policy outcomes. We implement three treatment arms: (1) the randomness of the initial wealth distribution, (2) the cost of communication, and (3) the transparency of messenger identity. This would allow us to observe how these factors influence wealthier agents' ability to shape others' social preferences. Our experiment also disentangles the effects of different social preference structures—such as *inequality aversion*, *efficiency concerns*, and *fairness*—on collective decision-making in the taxation game. Furthermore, by varying the messaging cost, we can identify the conditions under which wealth inequality is likely to perpetuate through strategic communication. Our experimental design provides a controlled environment to observe the dynamics predicted by our model. OUR data will offer insights into how policy interventions, such as equal access to political communication or altering the visibility of a messenger's identity, could mitigate the entrenchment of inequality in real-world settings.

The rest of the paper is structured as follows. We wrap up the introduction section with a discussion of our contributions to the literature on political persuasion, social preferences, and interaction between inequality, communication, and politics. In the next section, we describe the model, explain its predictions, and mention potential extensions. Section 2 details our experimental design and connects it to the model by mirroring its predictions in formulating testable hypotheses. We then outline the esti-

mation strategy in Section 4. Since this project is a **work in progress**, we conclude with a description of further steps, the timeline for the experiment, and potential challenges.

1.1. Contributions and Related Literature

This paper contributes to several broad streams of literature. First, it relates to research on political persuasion and the mechanisms behind it. In most economic literature, regardless of it being theoretical (Kamenica & Gentzkow 2011, Alonso & Câmara 2016), applied (Iyengar & Simon 2000, DellaVigna & Gentzkow 2010), or experimental (Baysan 2022, Djourelova 2023), persuasion is thought of as influencing behavior via the provision of information. This definition is limited by construction. On the one hand, such focus on information transmission allowed research to make great strides in revealing how free (Chakraborty & Harbaugh 2010), costly (Kriss et al. 2016), and acquired (Martinelli 2006) messages can transmit information to a possibly biased receiver (Galperti 2019) and be beneficial for an observably biased sender (Jeong 2019). On the other hand, this focus prevented the literature from expanding the definition of persuasion to a broader process driven by priming receivers' preferences and outcome framing, which are recognized as essential elements of political persuasion in the field of Political Science (Cwalina et al. 2014, Druckman 2022).

In fact, this gap between fields is often acknowledged but then dismissed. For example, communication can change preferences if agents make threats or promises, thus changing expected outcomes and shifting preferences (Kamenica 2019). We consider this and similar information-driven explanations incomplete as they fail to explain many experimental findings, such as varying impacts of medium, timing, and informativeness of communication on agents' behavior. We bridge this gap by proposing a complementary mechanism that incorporates elements from both fields. Our mechanism, reflected in the model and experimental design, is unique in enabling uninformative messages to change how fundamental social preferences manifest themselves in receivers' utility.

This paper, therefore, also contributes to the literature on social preferences and their interaction with communication. The topic of social (or other-regarding) preferences has received attention from a wide variety of perspectives – some complementary (e.g., Alesina et al. 2018, and Gärtner et al. 2017) and other conflicting (e.g., Engelmann & Strobel 2004, and Fehr et al. 2006, Bolton & Ockenfels 2006). Many attempts have been made to synthesize different theoretical notions of social preferences (Alesina & Giuliano 2011) and estimate them, both in the lab (Ackert et al. 2007, Bicchieri & Xiao 2009, Durante et al. 2014) and field (Gualtieri et al. 2019, Kuhn 2019). As far as we

know, our model is the first to introduce strategic communication to a setting with the three most prominent (Tucker & Xu 2023) components of social preferences – *efficiency preference* (Charness & Rabin 2002), *inequality aversion* (Fehr & Schmidt 1999), and *fairness* (Alesina & Angeletos 2005).

In the lab, evidence consistently shows that communication increases the weight of social preferences in decision-making (e.g., Xiao & Houser 2007).Our experimental design is unique in enabling us to (i) separate the effects of three components mentioned above in a non-distortionary[5] tax game with redistribution, (ii) public good provision, (iii) introduce communication, and (iv) test how inequality, communication, and voting decisions interact.[6] We draw upon the literature closest to each of the four elements in the following way. For (i) and (ii), we expand the taxation setup that captures social preferences of (Durante et al. 2014) by incorporating the public good creation mechanism seen in (Kriss et al. 2016). For (iii) and (iv), we add communication in a setting close to (Gantner et al. 2019) and explore the impact of inequality on socially-oriented behavior as studied in (Tavoni et al. 2011).

Consequently, our paper contributes to the broad literature on how economic inequality impacts political outcomes. In part of the literature, experiencing inequality alone reduces preferences for redistribution (Roth & Wohlfart 2018). In contrast, some research found that individual perceptions of wage inequality might have the opposite effect (Kuhn 2019) and higher levels of inequality lead to higher tax rates (Agranov & Palfrey 2015) in an experimental setting. The impact of information on demand for redistribution was also found to be heavily dependent on the content of the message and setting of the study. For example, the same information about wealth inequality and inter-generational mobility had the opposite effects between the U.S. and Western Europe (Hoy & Mager 2021). It has been firmly established, however, that high wealth inequality could enable the rich to manipulate information published in the media (Besley & Prat 2006, Petrova 2008). In turn, manipulation, and exclusion of information and media sources could significantly alter voter decisions (Enikolopov et al. 2011). Our work is unique in studying the direct causal chain between wealth inequality that puts

[5]We see the decision to make the tax non-distortionary as equivalent to making initial endowments exogenous to our model and separating their assignment from the voting game in our experiment. This essentially eliminates labor-supply distortions commonly seen in papers that use the conventional Meltzer–Richard model (e.g., Agranov & Palfrey 2015).

[6]We are also aware of a comprehensive analysis of social preferences performed in (Krawczyk & Le Lec 2021). While informative, their analysis only allowed an **individual-level** consistency comparison between model predictions and choices. Our experimental design does not have this limitation. Some experimental work also builds upon revealed preference ideas and tests the rationalizability of choices under several models of social preferences (Andreoni & Miller 2002, Fisman et al. 2007). We plan to explore this approach as an extension or a complementary addition to the project.

limits on communication and voting outcomes beneficial to the rich.

The paper also relates to a growing body of studies that use online platforms for conducting real-time group experiments (Huber et al. 2023). Inspired by (Flecke & Bachler 2024), we plan to contribute our observations on data quality, consistency with findings from on-the-ground labs, and challenges in implementation. Additionally, we make a small contribution to the experimental literature that uses real-effort tasks for measuring effort provision. Our experimental design is different in its uncommon setting (Carpenter & Huet-Vaughn 2019) and combining a diverse set of tasks (Charness et al. 2018) to achieve a random distribution of abilities, thus ensuring that initial wealth assignment is exogenous.

2 Model

2.1. Setting, Payoffs, and Voting

Setting.—A finite group of $2n+1 \geq 1$ voters must choose one tax level τ from a set of policy alternatives T. Each voter $i \in \mathcal{N} = \{1, ..., 2n+1\}$, either randomly or based on their ability, is endowed with initial wealth ω_i, drawn without replacement from a predetermined, highly unequal distribution Ω.[7] Chosen tax τ is flat, i.e. uniformly imposed on all endowments, and serves a dual purpose. First, it funds some public good creation through technology $g(\omega, \tau)$, $g : \Omega \times \mathrm{T} \to \mathbb{R}$. Second, it always equally redistributes the initial tax revenue $\tau\bar{\omega}$. Note that the optimal levels of public good and redistribution are implied to be voter-specific, a fact that we take advantage of in the construction of the payoff function.

Payoffs.— A voter's payoff $\pi(\omega, \tau), \pi : \Omega \times \mathrm{T} \to \mathbb{R}$ is the total monetary compensation that player i receives given some tax τ and average wealth $\bar{\omega}$.

$$\pi_i(\omega_i, \tau) = (1-\tau)\omega_i + \tau\bar{\omega} + g(\omega_i, \tau) \tag{1}$$

The first two elements are pretty straightforward, while $g(\omega_i, \tau)$ is admittedly not. On its face, it might seem to go against the conventional logic of taxes creating distortions instead of additional wealth. In our model, however, we abstract from decisions

[7] Since the number of voters is discrete, Ω is characterized by a step cdf. This, however, is not necessary for our setup to work as an extension to a continuum of voters, and a continuum of endowments is expected to produce similar results

in markets that could be distorted by taxes (e.g., labor supply) and focus more on how government spending on public goods could be an incredibly lucrative investment and not only return the full government budget G but also generate some additional wealth.[8] However, we do reflect the fact that public good could be over-funded or funded beyond its capacity, such that it only returns money necessary for redistributing $\tau\bar{\omega}$ and nothing more $g = 0$.[9]

We construct π and g to be have the following properties:

$$\pi(\omega_i, \tau) \in C^3 \iff g(\omega_i, \tau) \in C^3 \tag{A}$$

$$\begin{cases} \frac{\partial \pi(\tau,\omega_i)}{\partial \tau} \geq 0, \forall \tau \in [0, \hat{\tau}(\omega_i)] \\ \frac{\partial \pi(\tau,\omega_i)}{\partial \tau} < 0, \forall \tau \in (\hat{\tau}(\omega_i), 1] \\ \frac{\partial \pi(\tau,\omega_i)}{\partial \omega_i} > 0, \forall \tau \in [0, 1] \end{cases} \iff \begin{cases} \frac{\partial g(\tau,\omega_i)}{\partial \tau} \geq \omega_i - \bar{\omega}, \forall \tau \in [0, \hat{\tau}(\omega_i)] \\ \frac{\partial g(\tau,\omega_i)}{\partial \tau} < \omega_i - \bar{\omega}, \forall \tau \in (\hat{\tau}(\omega_i), 1] \\ \frac{\partial g(\tau,\omega_i)}{\partial \omega_i} > 0, \forall \tau \in [0, 1] \end{cases} \tag{B}$$

$$\begin{cases} \frac{\partial \hat{\tau}(\omega_i)}{\partial \omega_i} < 0 \\ \frac{\partial^2 \hat{\tau}(\omega_i)}{\partial \omega_i^2} < 0 \end{cases} \impliedby \begin{cases} \frac{\partial^2 g(\hat{\tau}(\omega_i),\omega_i)}{\partial \tau \partial \omega_i} < 0 \\ \frac{\partial^3 g(\hat{\tau}(\omega_i),\omega_i)}{\partial \tau^3} > 0 \end{cases} \tag{C}$$

where $\hat{\tau}(\omega_i) = \arg\max_\tau \{\pi_i(\tau, \omega_i)\}$. Property (A) is a fairly standard differentiability assumption. Set of Properties (B) ensures the (unique) **single-peakedness** of the payoffs for each wealth level ω_i and that total wealth does not decrease as a result of taxation. The final set (C) configures payoffs in such a way that optimal tax decreases with wealth and that differences in optimal tax are proportional to the relative change in wealth. In general, all three allow us to achieve a payoff function shape that is illustrated for our experimental setup in Appendix Figure 3.

While many polynomial functions easily achieve (A) and (B), ensuring (C) is not a trivial task even with only a sufficiency condition of g's properties for π's. Notice, however, that everything does not have to hold on the entire $[0, 1]$ line! In fact, almost all conditions [10] are only needed from the lowest optimal tax to the highest that directly follow from our constructed wealth distribution. The reason for that is that starting from a selfish standpoint, no agent would want to induce a tax that's lower than $\underline{\tau} =$

[8] A very salient example is a public good of tax collection and enforcement by the Internal Revenue Service in the U.S. It is estimated that every dollar of funding brings in \$5 to \$9 dollars in tax revenue (Swagel 2021).

[9] This allows us to avoid the effects that the perception of "wasteful" government spending could have on social preferences, as seen in (Sheremeta & Uler 2021).

[10] The exception is a line of conditions for π and g in (B) with the form $\frac{\partial f(\tau,\omega_i)}{\partial \omega_i} > 0$, but that is easy to ensure.

$\hat{\tau}(\max\{\omega_i\})$ and higher than $\bar{\tau} = \hat{\tau}(\min\{\omega_i\})$. Therefore, we restrict (A), (B), and (C) to $[\underline{\tau}, \bar{\tau}]$, which allows us to maintain the desired shape and maintain helpful properties for our analysis. Regardless of this simplification, we do contend that functions that follow all three properties $\mathrm{T} = [0, 1]$ and $\omega_i \geq 2$ exist.[11] From this point on, we abstract from any particular functional form and treat g as a standalone element in our model.

Voting Rule.—At the end of the game ($t = 2$), each voter submits their preferred tax τ_i. In the current iteration of our model, we implement a random dictator procedure where each voter is equally likely to be pivotal.

$$P(\tau^{Final} = \tau_i | \tau_{j \neq i}) = P(\tau^{Final} = \tau_i) = \frac{1}{2n+1}, \forall i \in \mathcal{N} \tag{2}$$

This allows us to model voter behavior as if they are pivotal and abstract from integrating beliefs about the likelihood of swaying the outcome. Prior to this setup, we considered multiple options, including median voter and plurality setups. When it comes to plurality, it can be characterized by the complex strategic behavior of coalition building around certain income cutoffs.[12] Since such behavior is outside the scope of our study, we do not pursue this avenue. The other option, median voter, was rejected to preserve the incentives of agents to influence everyone, not just the voter at the middle of the distribution Ω. We will come back to median voter when discussion the possible extensions and future steps.

2.2. Social Preferences in Utility

Utility.— We start with a simple conceptual model of social preferences:

$$U_i(\pi_i, \tau_i, \pi_{j \in \mathcal{N}}) = \underbrace{u_i(\pi_i(\tau_i))}_{\text{Individual Preference}} + \underbrace{\lambda_i v(\pi_{j \in \mathcal{N}}(\tau_i))}_{\text{Social Preference}} \tag{3}$$

This particular setup makes three crucial assumptions. First, every voter cares about their personal outcome π_i and the social outcome $\pi_{j \in \mathcal{N}}$. Second, both portions of the utility are separable into u_i and v respectively, and v is weighed by some parameter λ.

[11] The polynomial below achieves all outlined properties $\forall \tau \in [0, 1]$ and $\omega_i \in [2, 25]$ and underpins our experimental design when scaled up.

$$g(\tau^M, \omega_i) = 8.5(\tau + \frac{\omega_i}{\max\{\omega_i\}} - 1)^3 + b(\tau - 0.05 - \frac{1}{\omega_i^2})^2 + 22.5\tau + 10\ln(\omega_i - 1)$$

[12] Although, some experimental evidence shows that coalitions do not manifest even if private communication is allowed (Gantner et al. 2019).

Third, we assume that these two functions can be reduced to payoffs and apply it to u_i in Equation 4 and v later on.

$$U_i(\tau_i) = \pi_i(\tau_i) + \lambda_i v(\pi_{j\in\mathcal{N}}(\tau_i)) \tag{4}$$

Proposition 1. *A purely selfish ($\lambda_i = 0$) agent i will always vote for a tax that maximizes her payoff π_i, that is $\tau_i^* = \arg\max_\tau\{U_i(\tau)\} = \hat{\tau}(\omega_i)$.*

This proposition is the direct consequence of the chosen voting rule and utility construction. Given the positive probability of being pivotal, choosing $\tau_i^* = \hat{\tau}(\omega_i)$ is the dominant strategy of every player.

Efficiency.— We incorporate the preference for efficiency in a straightforward manner similar to (Charness & Rabin 2002). Each agent just averages the sum of all payoffs and incorporates it into the utility.

$$U_i^E(\tau_i) = (1-\lambda_i)\pi_i(\tau_i) + \lambda_i^E \frac{1}{2n+1}\sum_{j\in\mathcal{N}} \pi_j(\tau_i) \tag{5}$$

Proposition 2a outlines a very intuitive logic in how agents deviate from purely selfish choice $\hat{\tau}$ to an alternative τ^* when their social preferences are efficiency-oriented. We frame the deviations as movements closer to or further away from $\hat{\tau}$ of some other agent. This approach significantly simplifies the analysis of strategic behavior in the presence of messages.

Proposition 2a. *Efficiency-oriented ($\lambda_i^E > 0$) agent i would deviate from $\tau_i^* = \hat{\tau}(\omega_i)$ to*

1. *A payoff maximizing tax $\hat{\tau}(\omega_j) \in \mathrm{T} = [0,1]$ of agent j iff*

$$\lambda_i^E = \frac{(2n+1)(\omega_i - \bar{\omega} - g_i'(\hat{\tau}))}{\sum_{k\in\mathcal{N}}(-\omega_k + \bar{\omega} + g_k'(\hat{\tau}))} \tag{2a.1}$$

2. *A unique $\tau_i^E \in \mathrm{T} = [0,1]$ if $\hat{\tau}_i \neq \arg\max_\tau\{\sum_{j\in\mathcal{N}}\pi_j(\tau)\}$ and λ_i^E violates (2a.1) $\forall j \neq i$. This deviation improves the payoff of agent j iff*

$$\underbrace{(\tau_i^E - \hat{\tau}_i)(\bar{\omega} - \omega_j)}_{\textit{Distributional Change}} + \underbrace{g_j(\tau_i^E) - g_j(\hat{\tau}_i)}_{\textit{Public Good Change}} > 0 \tag{2a.2}$$

3. *Some $\tau_i^E \in \mathrm{T}$, where T is discrete and $\tau_i^E \neq \arg\max_{\tau\in\mathrm{T}}\{\pi_i(\tau)\}$, if*

$$\lambda_i^E > \frac{(2n+1)\overbrace{[(\tau_s - \tau_i^E)(\bar{\omega} - \omega_i) + g_i(\tau_i^E) - g_i(\tau_s)]}^{\text{Personal Change: } \underset{\hat{\tau}_i \to \tau_i^E}{\Delta} \pi_i}}{\underbrace{\sum_{k \in \mathcal{N}} [(\tau_s - \tau_i^E)(\bar{\omega} - \omega_k) + g_k(\tau_i^E) - g_k(\tau_s)]}_{\text{Social Change: } 1/(2n+1) \underset{\hat{\tau}_i \to \tau_i^E}{\Delta} \sum_{j \in \mathcal{N}} (\pi_j)}}, \quad \forall \tau_s \in \mathrm{T} : \tau_s \neq \tau_i^E \tag{2a.3}$$

This deviation improves the payoff of agent j iff (2a.2) holds for τ^* and $\hat{\tau}_i$.

Proposition 2a is straightforward by construction. Agent i deviates to j's purely selfish tax $\hat{\tau}_j$ only if her efficiency parameter λ_i^E equals to the ratio between personal and total social payoff portions of the first order condition from $\max_\tau\{U_i^E(\tau)\}$, i.e., condition 2a.1 holds. Since this is unlikely for any discrete society[13], the deviation is going to be unique and beneficial to j only if the sum of distributional change and public good change in j's payoff is positive, i.e., condition 2a.2 holds. In cases when the choice set T is discrete, the deviation happens if and only if the preference parameter λ_i^E is larger than the ratio between $\underset{\hat{\tau}_i \to \tau_i^E}{\Delta} \pi_i$ and $\frac{1}{2n+1} \underset{\hat{\tau}_i \to \tau_i^E}{\Delta} \sum_{j \in \mathcal{N}} (\pi_j)$.

Inequality Aversion.—The concept of inequality aversion was introduced in the seminal paper by (Fehr & Schmidt 1999). We draw heavily on it but also make some adjustments. First, we do not differentiate between envy ($\pi_i < \pi_j$) and guilt ($\pi_i > \pi_j$) parameters, a departure from the common setup in the literature (e.g., He & Wu 2016). Second, we avoid a quadratic specification seen in (Alesina & Giuliano 2011) for the convenience it provides in deriving the deviation conditions.

$$U_i^I(\tau_i) = \pi_i(\tau_i) + \lambda_i^I \frac{1}{2n+1} \sum_{j \neq i} |(\pi_j(\tau_i) - \pi_i(\tau_i))| \tag{6}$$

Proposition 2b. *Inequality-averse ($\lambda_i^I > 0$) agent i would deviate from $\tau_i^* = \hat{\tau}(\omega_i)$ to*

1. *A payoff maximizing tax $\hat{\tau}(\omega_j) \in \mathrm{T} = [0, 1]$ of agent j iff*

$$\lambda_i^I = \frac{(2n+1)(\omega_i - \bar{\omega} - g_i'(\hat{\tau}))}{\sum_{k \neq i} |(\omega_k - \omega_i + g_k'(\hat{\tau}) - g_i'(\hat{\tau}))|} \tag{2b.1}$$

[13] We conjecture that it is impossible for any society that is not a continuum of agents because of the way π is constructed. We leave the construction of the proof for the future as the idea is not essential for our model or experimental design.

2. *A unique* $\tau_i^I \in \mathrm{T} = [0,1]$ *if* $\hat{\tau}_i \neq \arg\max_\tau\{\sum_{j\in\mathcal{N}} \pi_j(\tau)\}$ *and* λ_i^I *violates (2b.1)* $\forall j \neq i$. *This deviation improves the payoff of agent* j *iff*

$$(\tau_i^I - \hat{\tau}_i)(\bar{\omega} - \omega_j) + g_j(\tau_i^I) - g_j(\hat{\tau}_i) > 0 \tag{2b.2}$$

3. *Some* $\tau_i^I \in \mathrm{T}$, *where* T *is discrete and* $\tau_i^I \neq \arg\max_{\tau\in\mathrm{T}}\{\pi_i(\tau)\}$, *if* $\forall \tau_s \in \mathrm{T} : \tau_s \neq \tau_i^I$

$$\lambda_i^I > \frac{(2n+1)[(\tau_s - \tau_i^I)(\bar{\omega} - \omega_i) + g_i(\tau_i^I) - g_i(\tau_s)]}{\sum\limits_{k\neq i}(|\tau_i^I(\omega_k - \omega_i) + g_k(\tau_i^I) - g_i(\tau_i^I)| - |\tau_s(\omega_k - \omega_i) + g_k(\tau_s) - g_i(\tau_s)|)} \tag{2b.3}$$

This deviation improves the payoff of agent j iff (2b.2) holds for τ_i^I *and* $\hat{\tau}_i$.

Fairness.—Preferences for a *fair* or just distribution are more difficult to capture. The key question is what constitutes a *fair* payoff in a context where the initial endowment is exogenous. There are many views on this question ranging from weighting the distance of actual payoffs to average payoff in the efficiency-maximizing setup to prioritizing complete equality (Alesina & Giuliano 2011). We assume that agents share the same *fair* benchmark. They have the same opinion about where in the income distribution each person i should have been put (ω_i^F) regardless of realized distribution (ω_i). This logic is similar to knowing the objective outcome in the presence of noise (Alesina & Angeletos 2005).

$$U_i^F(\tau_i) = \pi_i(\tau_i) + \lambda_i^F \frac{1}{2n+1} \sum_{j\in\mathcal{N}} |(\pi_j(\tau_i, \omega_j) - \pi_j(\tau_i, \omega_j^F))| \tag{7}$$

Proposition 2c. *Fairness-oriented* ($\lambda_i^F > 0$) *agent* i *would deviate from* $\tau_i^* = \hat{\tau}(\omega_i)$ *to*

1. *A payoff maximizing tax* $\hat{\tau}(\omega_j) \in \mathrm{T} = [0,1]$ *of agent* j *iff*

$$\lambda_i^F = \frac{(2n+1)(\omega_i - \bar{\omega} - g_i'(\hat{\tau}))}{\sum_{j\in\mathcal{N}} |(\omega_j^F - \omega_j + g_j'(\hat{\tau}) - g'(\hat{\tau}, \omega_j^F))|} \tag{2c.1}$$

2. *A unique* $\tau_i^F \in \mathrm{T} = [0,1]$ *if* $\hat{\tau}_i \neq \arg\max_\tau\{\sum_{j\in\mathcal{N}} \pi_j(\tau)\}$ *and* λ_i^F *violates (2c.1)* $\forall j \neq i$. *This deviation improves the payoff of agent* j *iff*

$$(\tau_i^F - \hat{\tau}_i)(\bar{\omega} - \omega_j) + g_j(\tau_i^F) - g_j(\hat{\tau}_i) > 0 \tag{2c.2}$$

3. *Some* $\tau_i^F \in \mathrm{T}$, *where* T *is discrete and* $\tau_i^F \neq \arg\max_{\tau\in\mathrm{T}}\{\pi_i(\tau)\}$, *if* $\forall \tau_s \in \mathrm{T} : \tau_s \neq \tau_i^F$

$$\lambda_i^F > \frac{(2n+1)[(\tau_s - \tau_i^F)(\bar{\omega} - \omega_i) + g_i(\tau_i^F) - g_i(\tau_s)]}{\sum_{j \in \mathcal{N}} (|(1-\tau_i^F)(\omega_j - \omega_j^F) + g_j(\tau_i^F) - g(\tau_i^F, \omega_j^F)| - |(1-\tau_s)(\omega_j - \omega_j^F) + g_j(\tau_s) - g(\tau_s, \omega_j^F)|)} \tag{2c.3}$$

This deviation improves the payoff of agent j iff (2c.2) holds for τ_i^F and $\hat{\tau}_i$.

2.3. (Costly) Messages that Shift Preferences

Messages.—Denote message space $\mathcal{M} = \{\emptyset, m^F, m^E, m^I, m^U\}$, where m^F is a message that promotes fairness, m^E promotes efficiency, m^I promotes equity, and m^U does not contain any information in regard to social preferences. The message could be free ($c = 0$), extremely prohibitive ($c > \max\{\omega_i\}_{i \in \mathcal{N}}$), or anything in between ($0 < c < \max\{\omega_i\}_{i \in \mathcal{N}}$). Therefore, when $\min\{\omega_i\}_{i \in \mathcal{N}} < c < \max\{\omega_i\}_{i \in \mathcal{N}}$, there is a limit on communication for poor voters. There is no cost of receiving messages, and all messages are public.

Agents are strategic in how they choose messages. We assume that at time $t = 1$ everyone starts with a purely selfish utility, which is just their payoff, and chooses a message such that it maximizes this utility:

$$m_i^* = \arg\max_{m \in \mathcal{M}} \{\pi_i(m)\} \tag{8}$$

In order for messages to matter for our payoff, we modify the utility function by assuming that the prevalent message type $M = maj\{m_i\}$ dictates the overall shape of social preferences. This is true, however, only for people who did not send the message of the same type $m_i \neq M$. Selfish agents who were able to induce a particular social norm for integrating distribution into the utility do not need to conform to it. If there is no consensus, everyone keeps their initial utility form. We theorize that this exact mechanism is behind the phenomenon of uninformative messages still having persuasive power.

Proposition 3. *A selfish agent i would never send a message*

1. *promoting efficiency, m^E, if they are the poorest ($\omega_i = \min\{\omega_j\}_{j \in \mathcal{N}}$)*
2. *promoting inequality aversion, m^I, if they are the richest ($\omega_i = \max\{\omega_j\}_{j \in \mathcal{N}}$)*

3. *that is unrelated to social preferences,* m^U*, if* $c > 0$ *regardless of their place in the distribution*

Proposition 3 follows from simple dynamics of how messages might affect voters in different relative positions to i's wealth. For the poor, it never makes sense to promote efficiency as it could only shift the expected tax away from their optimal decision. Same logic applies for the rich, as any equity-promoting message could only increase the expected tax. The proposition holds true, however, only for highly unequal distributions Ω and somewhat homogeneous λ_i. We leave the exploration of these dynamics for the future.

Proposition 4. *Given some small (*$c < \min\{\omega_i\}$*) but significant enough (*$c \gg 0$*) cost* c*, there exists Subgame Perfect Nash Equilibrium* (m_i^*, τ_i^*) *such that:*

1. *Relatively poor agents* (P) *do not send any messages, i.e.* $m_p^* = \{\emptyset\}, \forall p \in P$.
2. *Relatively rich agents* (R) *send efficiency-promoting messages, i.e.* $m_r^* = m^E, \forall r \in R$*, and form efficiency-oriented social norm.*
3. *All* $p \in P$ *shift their strategy* τ_p^* *from* $\hat{\tau}_p$ *to* τ_p^E*, while all* $r \in R$ *continue playing* $\hat{\tau}_r$.
4. *The final payoff distribution is more unequal than the one achieved with no communication.*

Proposition 4 contains perhaps the most important prediction of our model. In a situation where communication is available to everyone, but the expected gain from it is less than the cost for a low-wealth portion of the population, rich voters can take advantage of it. They can induce a particular social preference in others while not having to adhere to it themselves. This lets initial inequality, be it fair or unfair, perpetuate itself in the outcome through communication. We focus on this idea and explore it further in our experiment.

2.4. Extensions: Expressive Voting and Identity Cues

Expressive vs. Instrumental.—The first potential extension that we could make is incorporating the dual nature of the vote. In our current model, we assume that voters make decisions purely based on instrumental considerations—choosing policies that maximize their individual utility given the expected outcomes. However, voting can also be expressive, where individuals cast their votes to express their values and receive

utility from the social aspect regardless of the outcome (Eguia & Hu 2024). Expressive voting, in this sense, could significantly alter the predictions. For instance, if voters are motivated by expressive concerns and the number of agents is high enough, even a relatively small λ could lead to significant deviations from purely selfish behavior. The idea of expressive voting is especially relevant in the median-voter types of setups. There, consistent deviations from payoff-maximizing tax can be explained by voters perceiving the probability of being pivotal as almost non-existent. Thus, they prioritize their social preferences and vote exclusively in adherence to the social norm of how distribution should be incorporated into the utility.

Identity.— An interesting idea that we plan to test in our experiment is the role of the messenger's identity in the persuasiveness of communication. Driven by some interesting results from the literature (Dewatripont & Tirole 2005, Sances 2013), we hypothesize that revealing the identity of the sender, i.e., their initial endowment ω, could attenuate the message effects in dictating the exact form of social preferences. This idea follows a similar logic to in-group vs. out-group dynamics seen in other settings. At this stage of the project, we are yet to attempt to conceptualize this effect in our model.

3 Experimental Design

3.1. Setting

Experimental sessions will be conducted online, with each session including 9 participants and lasting 60 minutes. Participants will be recruited from Prolific.ac, an online platform explicitly designed for academic research. Prolific is chosen due to its ability to provide a diverse and relatively naive participant pool, which has been shown to yield data quality comparable to other platforms like Amazon Mechanical Turk (Palan & Schitter 2018) and in-person labs (Douglas et al. 2023). The experiment is programmed using oTree (Chen et al. 2016). It allows for real-time interaction among participants, a key element for capturing the dynamics of communication and decision-making under conditions of inequality.

3.2. Protocol

Sessions will begin with instructions that appear on participants' computer screens and are simultaneously read aloud by the experimenter. Subjects will be informed they will receive a show-up fee plus an additional payoff that will depend on the experiment's

outcome. The core of the experiment will consist of three parts, which we refer to as "Part 1", "Part 2", and "Part 3". Near the end of the experiment, one of the rounds will be randomly selected for payment. The timing of each session is illustrated in Figure 1.

We start Part 1 by telling subjects that each will be assigned one of 9 possible wealth levels, ranging from 9 to 125 experimental currency units (ECUs). Figure 4, shown on subjects' computer screens, illustrates the distribution of the initial wealth. We then will explain that provisional earnings are assigned to subjects in two possible ways: (1) randomly or (2) according to performance in a computer-based real-effort task (e.g., Tetris). Which method would actually be used to assign payoffs to subjects would be determined by a random draw at the end of Part 1. We tell subjects they will be able to alter the initial distribution by taxing earnings and redistributing the modified proceeds equally among all; in particular, they will be asked to choose a proportional tax rate ranging from 0% to 100% in increments of 10%. We will also explain to the subjects how the decision rule is implemented in Figure 5. We will illustrate the effect of taxation on earnings graphically and through a table. The table is produced in Figure 6.

We continue the experiment in Part 2, where the income distribution rule is finalized, and each individual observes both their wealth ω_i and ranking in real effort task. Then, we will reveal the results of a random draw for the cost of the message (free, costly, or unaffordable) and if each message will include information about the sender. Then, participants will be allowed to send one message to others. The communication occurs in an open forum/broadcast manner, and everyone can see all messages that are sent. When everyone finishes writing and sending their messages, they can read all communication that occured.

Part 3 includes participants reading the messages and then moving on to a screen where they can vote for the tax rate that should be imposed on the group. The median of all chosen tax rates will be the one implemented for the whole group. The interface for making this choice can be seen in Figure 7.

3.3. Treatments

Table 1 provides a detailed description of the treatment branches used in the experimental study. We outline three main variables: the *Income determination method* (i), *Message cost* (ii), and *Sender identity* (iii). All four variables vary between subjects, allowing us to test specific hypotheses derived from the model.

The *Income determination method* refers to how individuals are assigned pretax payoffs, with the method being either random or based on effort. We speculate that this

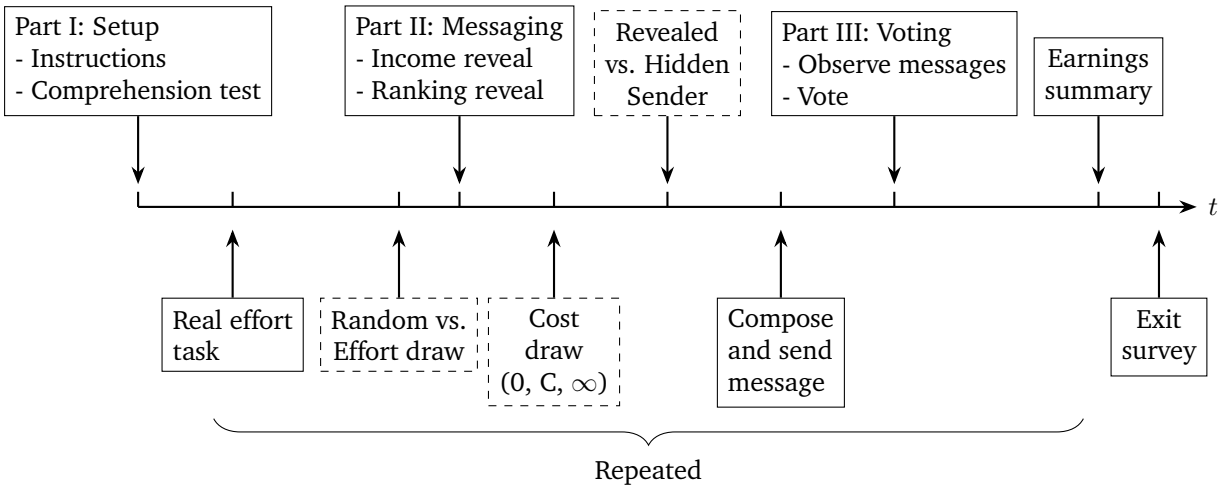

Figure 1: Session Structure and Timing

variable can play a role in determining the fair endowment benchmark (ω^F) as defined in the model. The *Message cost* variable decreases the expected return to a message within the experimental setup. The costs can be set at 0, C, or ∞, corresponding to free communication, a significant but non-prohibitive cost ($0 < c < \omega_{min}$), or a prohibitive cost ($c > \omega_{max}$), respectively. The *Sender identity* variable concerns whether the initial wealth of the message sender is revealed or hidden during the experiment. We speculate that the weight of each message (m_i) in the formation of the collective norm (M) could depend on the sender's identity. When identity is revealed, it may attenuate the persuasive power of messages from wealthier individuals, as recipients might discount these messages, perceiving them as self-serving.

3.4. Hypotheses

Each Proposition in our model can be directly tested with our experimental setup. We focus our attention on replicating previous research on the existence of social preferences, finding the communication effects (or lack thereof) on voting behavior, and analyzing the outcomes in the context of message content and senders identity.

Hypothesis 1. [Primary] *Messages are persuasive and can change how social preferences are integrated into utility.*

We plan to test this by exploring different message costs in treatment branch (ii). If results for free communication or unequal communication are consistently different from no communication, it implies that agents can indeed be swayed by messages in how they integrate social preferences in their utility.

Treatment variable	Description	Values	Model	Hypotheses
Income determination method	Method used to assign individuals to pretax payoffs	Random, Effort	Determines fair endowment ω^F	Influences demand for fairness
Message cost	Regulates who has the opportunity to send messages	0, C, ∞	Free $c = 0$, significant $0 < c < \omega_{min}$, or prohibitive $c > \omega_{max}$	Unequal access to messages ensures a better outcome for the wealthier group
Sender identity	Initial wealth of the message sender	Revealed, Hidden	Weight of m_i in formation of norm M	Inclusion of identity attenuates persuasion

Table 1: Description of treatment variables.

Hypothesis 1a. [Secondary] *Purely selfish utility predicts the behavior of voters in a no-communication, effort driven setting.*

This hypothesis reflects a simplifying assumptions that agents start of from the purely selfish perspective. The prediction for behavior would be straightforward – in no-communication branches, voters follow Proposition 1. The falsification of this hypothesis would be somewhat expected, however. As discussed before, many studies show that even in environments with no interactions, agents still display some preferences for re-distribution and fairness. Propositions 2a-2c provide a great set of benchmarks for such behavior. If we find that social preferences manifest even without communication, we would be able to distinguish between them.

Hypothesis 1b. [Secondary] *Messages with different content induce different social preference structure.*

Since we distinguish between message types in our model, we hypothesize that messages with, e.g., fairness-related content would induce a preference for fairness U^F. This idea connects to Proposition 3 in that some types of messages and behavior should not be rationalizable, e.g., equity messages from the rich. By varying message cost and conducting enough sessions with different communication outcomes M, we could compare the decisions under varying M to the model's benchmark predictions U^M.

Hypothesis 2. [Primary] *Initial inequality could perpetuate itself through costly communication even if everyone can afford to send a message.*

The difference between costly and no communication is exactly the influence the rich have secured through messaging. While communication is available to everyone, the expected gain from it could be less than the cost for a low-wealth portion of the population, as seen in Proposition 4. In turn, we predict the voting outcome will be consistently more favorable toward the group with an out-sized presence in messaging. This would allow initial inequality, be it fair or unfair, to perpetuate itself in the outcome through communication.

Hypothesis 2a. [Secondary] *Disclosure can mitigate the effects of unequal access to messaging.*

A logical follow-up to the primary hypothesis is considering how the problem can be remedied. One apparent way to do it is to lower messaging costs for everyone, which is explored in the Treatment Branch (ii). However, this is a somewhat radical solution that does not appear realistic. A less demanding alternative is to require disclosure of the income level of each sender, Treatment (iii). Since even in an unequal environment, messaging is accessible to a portion of the distribution, revealing the sender's identity exactly identifies their payoff preferences. This information can then either be used in a simple manner, e.g., for out-group discrimination, or in a more sophisticated analysis of the message content. Both could influence how persuasive each message is.

3.5. Ethics Approval

The experimental design has received the Human Research Subjects approval. Office of Research Compliance Administration, UC Santa Cruz, determined the exempt status of the project from the IRB review. Full approval details can be found in the IRB# HS-FY2021-16 protocol.

4 Data and Estimation

4.1. Data

The data collected at the end of each round will include the initial and final wealth, real task effort ranking, messaging decision, and final vote. In addition, at the end of

each session, the exit survey will collect demographic and socioeconomic information of participants. The scope of uses other than the direct purpose of the study for this data is limited. However, the variety of treatments and covariates collected can be used to establish additional patterns in behavior and can later serve as a justification for additional experiments.

4.2. Estimation Approach

Our estimation approach is driven by two outcomes – individual vote (τ_i) and societal (τ_s) outcome. With our experimental design as a foundation, we plan to analyze the interaction between communication, social preferences, and inequality in three distinct steps. First, we will test Hypothesis 1a by estimating the role of social preferences in the voting choices of our participants. Second, we will explore Hypothesis 1, 1b, and **??** by estimating the interaction between communication and taxation choice in different treatment branches. Third, we will investigate the content of the messages and, if some patterns emerge, classify them. This will enable us to investigate Hypothesis 2, i.e., the potential heterogeneous effects of different message types, and provide insight into how people use communication in our setup.

Social Preferences.— To establish the presence of social preferences, we will initially abstract in our regression from exact functional structures discussed in the model. This leads us to the following semi-parametric partially linear setup.

$$\begin{aligned} \tau_i &= \alpha\pi_i + \beta T_i + g(\Pi_{j\neq i}, X_i') + T_i \times g(\Pi_{j\neq i}, X_i') + \varepsilon_i \\ \tau_s &= \beta T_s + g(\Pi_s, X_s') + T_s \times g(\Pi_s, X_s') + \varepsilon_s \end{aligned} \tag{9}$$

where π_i is the individual payoff, T is the vector of individual (T_i) and session (T_s) treatments and their interactions, $g(\Pi, X')$ is a function that depends on the payoff distribution either of others $\Pi_{j\neq i}$ or its general shape Π_s, and a vector of demographic characteristics of the individual participant (X_i) or aggregated across all session participants X_s'. The last portion $T \times g(\Pi, X')$ is the interaction between all treatments and the distributional preference. Here and in all the following specifications, we plan to cluster standard errors on the individual level for τ_i and session level for τ_s as we run multiple rounds in each session and individuals make multiple unrelated voting decisions.

Communication.— While the previous step draws on a conservative null of other agents' payoffs having no effect on the voting decisions. We anticipate, however, the rejection of the hypothesis. Assuming that this prediction will be confirmed, we plan

to test the specifications of preferences for fairness, efficiency, and inequality aversion we outlined in our model. On the individual level, the estimation procedure would transform to

$$\tau_i = \alpha\pi_i + \beta T_i + \lambda G(\Pi_{j\neq i}) + \delta T_i \times G(\Pi_{j\neq i}) + \varepsilon_i$$
$$\text{where } G = \frac{1}{2n+1}\begin{pmatrix} \sum_{j\neq i}\pi_j \\ \sum_{j\neq i}|\pi_j - \pi_i| \\ \sum_j |\pi_j - \pi_j^F| \end{pmatrix} \tag{10}$$

This is a preliminary specification that includes a vector of social preference structures for *efficiency*, *inequality aversion*, and *fairness*. We directly connect it to our model by estimating individual coefficients in a vector λ. For the fairness benchmark π^F, we use the ranking of the real effort task that precedes our experiment. In the random wealth assignment treatment, we reveal it to participants. We anticipate that this would predispose them to benchmark the fair outcome to the imagined outcome in which the wealth was assigned based on this ranking.

Message content.—We plan to analyze the content of messages (m^k)to classify them by type K as uninformative (m^U) or promoting either efficiency (m^E), equity (m^I), or fairness (m^F). The classification would either be done manually or, more likely, using recent advances in supervised learning algorithms. We envision this process to start with the manual classification of a portion of messages. The next step would include tokenizing, preprocessing, and converting them into numeric data using the bag of words method. Then, we would train, cross-validate, and compare regularized ML estimators (e.g., ridge, lasso).

Even though the process outlined above is reasonable, it has its problems. First, the process of manual classification is time-consuming and prone to errors. Second, it is unclear if the quality of the prediction can be trustworthy. Given recent developments in the area of natural language models, a more promising alternative that we are considering is to use one of them in conjunction with both the message content and the context of our study.

In fact, several deep learning-based models have surpassed the classical approaches (Minaee et al. 2021). Among them, BERT [14] model seems to be the most applicable as it can be trained on plain text for language comprehension and classification (Devlin et al. 2018). The primary benefit of using BERT is that it allows for fine-tuning of parameters for a target task (Sun et al. 2019). We could choose them in such a way as to maxi-

[14]BERT– Bidirectional Encoder Representations from Transformers

mize the accuracy of a simple prediction of the message's type in regard to our specific definitions of social preferences. We could also use this approach to determine the existence and type k of the consensus M^k, thus avoiding researcher bias in constructing the variable.

Therefore, the specification that accounts for message content and the impact of the consensus formed during communication on the structure of social preferences takes the following structure.

$$\begin{aligned}\tau_i = \alpha\pi_i + \beta T_i + \gamma_0 m_i^k + \lambda G(\Pi_{j\neq i}) \\ + \delta T_i \times G(\Pi_{j\neq i}) + \gamma_1 m_i^k \times M^k \times G(\Pi_{j\neq i}) \\ + \gamma_2 M^k \times T_i \times G(\Pi_{j\neq i}) \\ + \gamma_3 m_i^k \times M^k \times T_i \times G(\Pi_{j\neq i}) + \varepsilon_i\end{aligned} \tag{11}$$

We plan to construct m^k and M^k in our data as matrices of vectors such that.

$$m^k = \begin{pmatrix} m^E \\ m^I \\ m^F \\ m^F \end{pmatrix}, \text{ where } m^i = \begin{cases} 1, \text{ if classified as } k\text{-promoting} \\ 0, \text{ otherwise} \end{cases} \tag{12}$$

4.3. Power Analysis

In anticipation of running the experiment, we conduct power analysis following guidelines outlined in (Vasilaky & Brock 2020). The statistical power of our experiment primarily depends on the anticipated features of the voting behavior. The payoff function is constructed so that each wealth level has its own unique tax t_i^* that maximizes their after-tax income. As outlined above, $v^* = t_i^*$ is a weakly dominant voting strategy. Thus, we assume that in equilibrium under the null, the final votes are uniformly distributed over $T^* = \{t_i^*\}_1^9$. This gives us the following population parameters: $sd = 0.26$, $\sigma^2 = 0.067$, $mean = 0.5$. Another important consideration is the expected Average Treatment Effect (ATE) for each branch. We implement a conservative approach that sets the expected ATE at the level of the lowest possible deviation from optimal tax t_i^* that is observable for each individual. In our case, this value is -0.1 given the logic of political persuasion toward lower tax when the access to messages is unequal. The approach is similar for other branches. Notably, the experiment includes several treatments. Therefore, we adjust for multiple hypothesis testing when conducting power analysis. Figure 2 reports the results. The power of 0.8 is achievable by 9 players in 11 sessions consisting

of 3 game rounds.

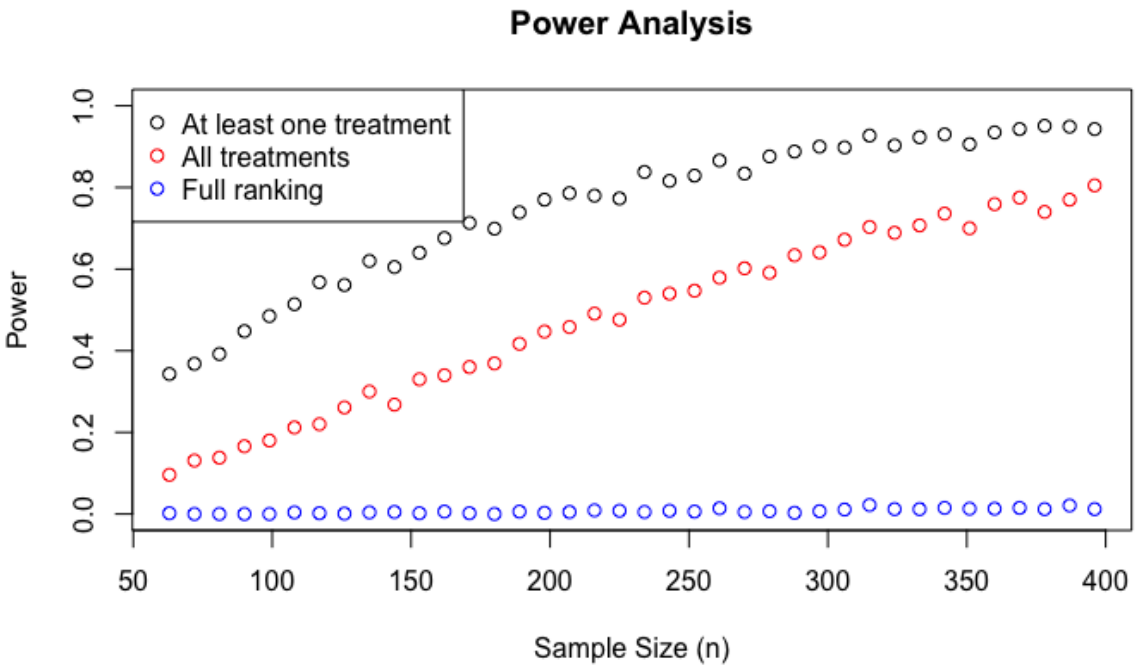

Figure 2: Power Analysis for Different Sample Sizes. Assumes that the true effect is -0.1, i.e., one jump from the preferred tax rate toward a lower one given there is unequal messaging. Population parameters are as follows: $sd = 0.26$, $\sigma^2 = 0.067$, $mean = 0.5$.

5 Discussion and Further Steps

In this paper, we developed a model and proposed an experimental design with a goal to explore the complex interaction between communication, social preferences, and wealth inequality. The current iteration of the model highlights how communication can be a powerful tool for influencing voter behavior, particularly in settings with significant initial wealth disparities. By allowing agents to send messages that shape social preferences, such as inequality aversion, efficiency, and fairness, we have shown that wealthier individuals are expected to exploit communication to perpetuate inequality. They do so by promoting efficiency at the expense of equity. This finding underscores the importance of considering both the content and accessibility of political messages in understanding how social preferences are formed and manifested in voting behavior.

Our experimental design is near completion in development and almost ready for implementation. It will allow us to test the predictions of our model by varying the

cost of communication, the method of income determination, and the transparency of sender identity. Once the experimental setup is finalized and programming concludes, we intend to promptly take advantage of the already secured funding. We believe that our design will provide us with a robust dataset that will offer valuable insights into the mechanisms driving voter behavior in unequal societies. The insights from our model and experiment will contribute to the broader literature on political persuasion, social preferences, and the intersection of inequality and communication.

5.1. Timeline

The timeline for the experimental implementation is outlined in Figure 9. From August to October 2024, we plan to finalize the development of the oTree web platform and run several pilots to prepare for the full-scale implementation. The primary stage of the project is planned to take place from September to December 2024, when we anticipate collecting most of the data. From November 2024 to February 2025, we plan to analyze data and prepare some preliminary results that could be presented at UCSC's Brown Bag seminar and the Behavioral, Experimental, and Theoretical Economics workshop. This would give us necessary feedback that would later be incorporated into the final draft, which we plan to complete by May 2025.

There are two sets of outputs for this project. Primary outputs are a pilot testing report, collected data, an initial draft of findings, and the final paper. Secondary outputs are exploratory data analysis unrelated to the research question, intermediary drafts, and workshop presentations. Once the paper is finalized through soliciting feedback from colleagues, circulating the preprint through SSRN/arXiv, and presenting it, we plan to submit the paper to a peer-reviewed journal.

5.2. Budget

The current experimental budget is $5,000 in the form of Dissertation Research Grant from UCSC Department of Economics. The budget will be used to conduct the experiment online via Prolific, as discussed in Section 3.1. The experiment will require payments to approximately 300 participants, with an average compensation of $15 per participant, totaling $4,500. The remaining $500 were allocated to finalizing the experimental software interface programmed using oTree, ensuring that all elements function seamlessly during the experiment.

Additionally, we are in the process of applying for further funding through the Hu-

mane Studies Fellowship offered by the Institute for Humane Studies. [15] If successful, this additional funding will allow us to expand the scope of the experiment, possibly increasing the number of participants or exploring additional treatment branches. We made the initial expenditures in July 2024 and expect to make primary expenses during the period from September to November 2024.

[15] For more information, see IHS website.

References

Ackert, L. F., Martinez-Vazquez, J. & Rider, M. (2007), 'Social preferences and tax policy design: some experimental evidence', *Economic Inquiry* **45**(3), 487–501.

Agranov, M. & Palfrey, T. R. (2015), 'Equilibrium tax rates and income redistribution: A laboratory study', *Journal of Public Economics* **130**, 45–58.

Alesina, A. & Angeletos, G.-M. (2005), 'Fairness and redistribution', *American economic review* **95**(4), 960–980.

Alesina, A. & Giuliano, P. (2011), Preferences for redistribution, *in* 'Handbook of social economics', Vol. 1, Elsevier, pp. 93–131.

Alesina, A., Stantcheva, S. & Teso, E. (2018), 'Intergenerational mobility and preferences for redistribution', *American Economic Review* **108**(2), 521–554.

Alonso, R. & Câmara, O. (2016), 'Persuading voters', *American Economic Review* **106**(11), 3590–3605.

Andreoni, J. & Miller, J. (2002), 'Giving according to garp: An experimental test of the consistency of preferences for altruism', *Econometrica* **70**(2), 737–753.

AP-NORC (2023), 'Many dissatisfied with the government's spending priorities.', AP-NORC Press Release.
URL: *https://apnorc.org/projects/many-dissatisfied-with-the-governments-spending-priorities/ [Accessed 15-08-2024]*

Baysan, C. (2022), 'Persistent polarizing effects of persuasion: Experimental evidence from turkey', *American Economic Review* **112**(11), 3528–3546.

Besley, T. & Prat, A. (2006), 'Handcuffs for the grabbing hand? media capture and government accountability', *American economic review* **96**(3), 720–736.

Bicchieri, C. & Xiao, E. (2009), 'Do the right thing: but only if others do so', *Journal of Behavioral Decision Making* **22**(2), 191–208.

Bolton, G. E. & Ockenfels, A. (2006), 'Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment', *American Economic Review* **96**(5), 1906–1911.

Carpenter, J. & Huet-Vaughn, E. (2019), Real-effort tasks, *in* 'Handbook of research methods and applications in experimental economics', Edward Elgar Publishing, pp. 368–383.

Chakraborty, A. & Harbaugh, R. (2010), 'Persuasion by cheap talk', *American Economic*

Review **100**(5), 2361–2382.

Charness, G., Gneezy, U. & Henderson, A. (2018), 'Experimental methods: Measuring effort in economics experiments', *Journal of Economic Behavior & Organization* **149**, 74–87.

Charness, G. & Rabin, M. (2002), 'Understanding social preferences with simple tests', *The quarterly journal of economics* **117**(3), 817–869.

Chen, D. L., Schonger, M. & Wickens, C. (2016), 'otree—an open-source platform for laboratory, online, and field experiments', *Journal of Behavioral and Experimental Finance* **9**, 88–97.

Corneo, G. (2006), 'Media capture in a democracy: The role of wealth concentration', *Journal of Public Economics* **90**(1-2), 37–58.

Coutts, A. (2019), 'Good news and bad news are still news: Experimental evidence on belief updating', *Experimental Economics* **22**(2), 369–395.

Cwalina, W., Falkowski, A. & Newman, B. (2014), 'Persuasion in the political context: opportunities and threats', *The handbook of persuasion and social marketing* pp. 61–128.

DellaVigna, S. & Gentzkow, M. (2010), 'Persuasion: empirical evidence', *Annu. Rev. Econ.* **2**(1), 643–669.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018), 'Bert: Pre-training of deep bidirectional transformers for language understanding', *arXiv preprint arXiv:1810.04805* .

Dewatripont, M. & Tirole, J. (2005), 'Modes of communication', *Journal of political economy* **113**(6), 1217–1238.

Djourelova, M. (2023), 'Persuasion through slanted language: Evidence from the media coverage of immigration', *American economic review* **113**(3), 800–835.

Douglas, B. D., Ewell, P. J. & Brauer, M. (2023), 'Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona', *Plos one* **18**(3), e0279720.

Druckman, J. N. (2022), 'A framework for the study of persuasion', *Annual Review of Political Science* **25**(1), 65–88.

Durante, R., Putterman, L. & Van der Weele, J. (2014), 'Preferences for redistribution and perception of fairness: An experimental study', *Journal of the European Economic Association* **12**(4), 1059–1086.

Eguia, J. & Hu, T.-W. (2024), Voter polarization and extremism, Technical report.

Ekins, E. E. (2019), 'What americans think about poverty, wealth, and work', *The Cato Institute* .

Engelmann, D. & Strobel, M. (2004), 'Inequality aversion, efficiency, and maximin preferences in simple distribution experiments', *American economic review* **94**(4), 857–869.

Enikolopov, R., Petrova, M. & Zhuravskaya, E. (2011), 'Media and political persuasion: Evidence from russia', *American economic review* **101**(7), 3253–3285.

Fehr, E., Naef, M. & Schmidt, K. M. (2006), 'Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment', *American Economic Review* **96**(5), 1912–1917.

Fehr, E. & Schmidt, K. M. (1999), 'A theory of fairness, competition, and cooperation', *The quarterly journal of economics* **114**(3), 817–868.

Fisman, R., Kariv, S. & Markovits, D. (2007), 'Individual preferences for giving', *American Economic Review* **97**(5), 1858–1876.

Flecke, S. L. & Bachler, S. (2024), 'Conducting real-time interactive experiments on prolific: A guide for researchers', *Available at SSRN 4814645* .

Galperti, S. (2019), 'Persuasion: The art of changing worldviews', *American Economic Review* **109**(3), 996–1031.

Gantner, A., Horn, K. & Kerschbamer, R. (2019), 'The role of communication in fair division with subjective claims', *Journal of Economic Behavior & Organization* **167**, 72–89.

Gärtner, M., Mollerstrom, J. & Seim, D. (2017), 'Individual risk preferences and the demand for redistribution', *Journal of Public Economics* **153**, 49–55.

Gualtieri, G., Nicolini, M. & Sabatini, F. (2019), 'Repeated shocks and preferences for redistribution', *Journal of Economic Behavior & Organization* **167**, 53–71.

He, H. & Wu, K. (2016), 'Choice set, relative income, and inequity aversion: an experimental investigation', *Journal of Economic Psychology* **54**, 177–193.

Hoy, C. & Mager, F. (2021), 'American exceptionalism? differences in the elasticity of preferences for redistribution between the united states and western europe', *Journal of Economic Behavior & Organization* **192**, 518–540.

Huber, C., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Weitzel, U., Abellán, M., Adayeva, X., Ay, F. C., Barron, K. et al. (2023), 'Competition and moral behavior:

A meta-analysis of forty-five crowd-sourced experimental designs', *Proceedings of the National Academy of Sciences* **120**(23), e2215572120.

Iaryczower, M., Shi, X. & Shum, M. (2018), 'Can words get in the way? the effect of deliberation in collective decision making', *Journal of Political Economy* **126**(2), 688–734.

Iyengar, S. & Simon, A. F. (2000), 'New perspectives and evidence on political communication and campaign effects', *Annual review of psychology* **51**(1), 149–169.

Jeong, D. (2019), 'Using cheap talk to polarize or unify a group of decision makers', *Journal of Economic Theory* **180**, 50–80.

Jones, R. P., Jackson, N., Orcés, D., Huff, I. & Holcomb, T. (2021), *Competing Visions of America: An Evolving Identity Or a Culture Under Attack?: Findings from the 2021 American Values Survey*, PRRI.

Jou, J., Niederdeppe, J., Barry, C. L. & Gollust, S. E. (2014), 'Strategic messaging to promote taxation of sugar-sweetened beverages: lessons from recent political campaigns', *American journal of public health* **104**(5), 847–853.

Kamenica, E. (2019), 'Bayesian persuasion and information design', *Annual Review of Economics* **11**(1), 249–272.

Kamenica, E. & Gentzkow, M. (2011), 'Bayesian persuasion', *American Economic Review* **101**(6), 2590–2615.

Krawczyk, M. & Le Lec, F. (2021), 'How to elicit distributional preferences: A stress-test of the equality equivalence test', *Journal of Economic Behavior & Organization* **182**, 13–28.

Kriss, P. H., Blume, A. & Weber, R. A. (2016), 'Coordination with decentralized costly communication', *Journal of Economic Behavior & Organization* **130**, 225–241.

Kuhn, A. (2019), 'The subversive nature of inequality: Subjective inequality perceptions and attitudes to social inequality', *European Journal of Political Economy* **59**, 331–344.

Maćkowiak, B., Matějka, F. & Wiederholt, M. (2023), 'Rational inattention: A review', *Journal of Economic Literature* **61**(1), 226–273.

Marriott III, R. W. & Dillard, J. P. (2021), 'Sweet talk for voters: a survey of persuasive messaging in ten us sugar-sweetened beverage tax referendums', *Critical Public Health* **31**(4), 477–486.

Martinelli, C. (2006), 'Would rational voters acquire costly information?', *Journal of Economic Theory* **129**(1), 225–251.

Mengel, F. & Weidenholzer, E. (2023), 'Preferences for redistribution', *Journal of Economic Surveys* **37**(5), 1660–1677.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M. & Gao, J. (2021), 'Deep learning–based text classification: a comprehensive review', *ACM computing surveys (CSUR)* **54**(3), 1–40.

Palan, S. & Schitter, C. (2018), 'Prolific. ac—a subject pool for online experiments', *Journal of behavioral and experimental finance* **17**, 22–27.

Petrova, M. (2008), 'Inequality and media capture', *Journal of public Economics* **92**(1-2), 183–212.

Roemer, J. E. (1998), 'Why the poor do not expropriate the rich: an old argument in new garb', *Journal of Public Economics* **70**(3), 399–424.

Roth, C. & Wohlfart, J. (2018), 'Experienced inequality and preferences for redistribution', *Journal of Public Economics* **167**, 251–262.

Sances, M. W. (2013), 'Is money in politics harming trust in government? evidence from two survey experiments', *Election Law Journal* **12**(1), 53–73.

Sheremeta, R. M. & Uler, N. (2021), 'The impact of taxes and wasteful government spending on giving', *Experimental Economics* **24**(2), 355–386.

Sun, C., Qiu, X., Xu, Y. & Huang, X. (2019), How to fine-tune bert for text classification?, *in* 'Chinese computational linguistics: 18th China national conference, CCL 2019, Kunming, China, October 18–20, 2019, proceedings 18', Springer, pp. 194–206.

Swagel, P. (2021), 'The Effects of Increased Funding for the IRS', Congressional Budget Office Blog.
URL: *https://www.cbo.gov/publication/57444 [Accessed 18-08-2024]*

Tavoni, A., Dannenberg, A., Kallis, G. & Löschel, A. (2011), 'Inequality, communication, and the avoidance of disastrous climate change in a public goods game', *Proceedings of the National Academy of Sciences* **108**(29), 11825–11829.

Tucker, S. & Xu, Y. (2023), 'Fairness,(perception of) inequality, and redistribution preferences', *Journal of Economic Surveys* **37**(5), 1529–1533.

Vasilaky, K. N. & Brock, J. M. (2020), 'Power (ful) guidelines for experimental economists', *Journal of the Economic Science Association* **6**(2), 189–212.

Vincent, J. (2022), 'Twitter says paying Blue subscribers now get 'prioritized rankings in conversations".
URL: *https://www.theverge.com/2022/12/23/23523845/twitter-blue-*

paying-priority-replies-conversations *[Accessed 21-08-2024]*

Xiao, E. & Houser, D. (2007), Emotion expression and fairness in economic exchange, Technical report, George Mason University, Interdisciplinary Center for Economic Science.

6 Appendix

6.1. Interface

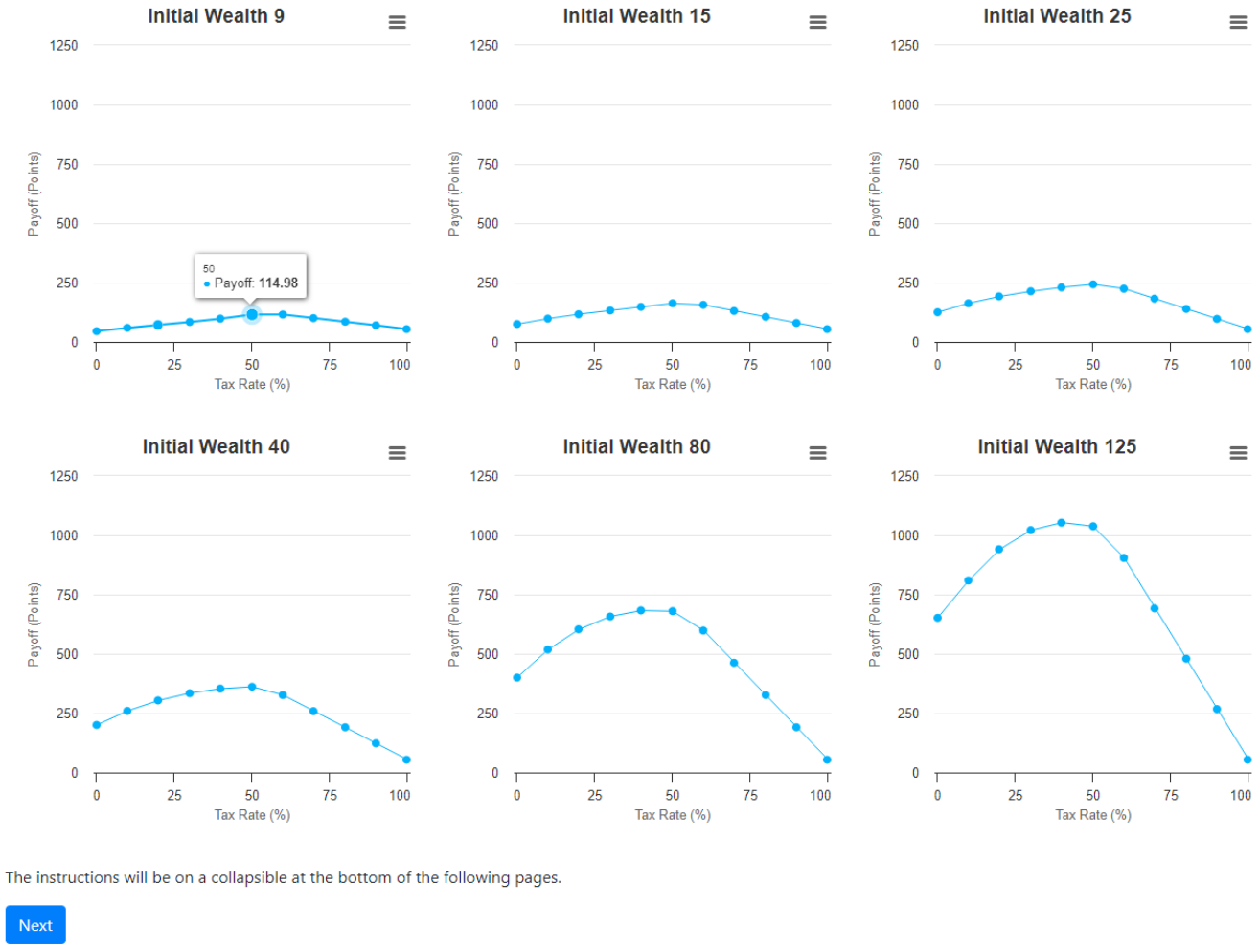

Figure 3: Payoff Shape

Introduction

Instructions

General Idea

This experiment consists of 2 rounds of interaction. At the beginning of each round, you will be randomly assigned to a group of nine citizens. All rounds will have a similar structure and rules: You will first perform a task. Your initial wealth will be determined by your performance in the task or by luck. Then all the citizens will choose a tax level that will apply to the whole group.

Before choosing the tax level, you may receive the chance to send a message to other citizens. We detail the rules below.

Generating Initial Wealth

In every round, every citizen will perform a task. After the task, the computer will generate a ranking of performance in the task for all nine citizens. Your position in the ranking may or may not determine your initial wealth following a probabilistic rule.

There will be six wealth levels: 9, 15, 25, 40, 80, and 125. If your wealth is generated by your ranking the following table will apply:

Ranking in task	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
Initial Wealth (Points)	125 points	80 points	40 points	25 points	25 points	15 points	15 points	15 points	9 points

Notice there are nine citizens, but only six different wealth levels. Two citizens, ranks 4 and 5, both have wealth level 25, and three citizens, ranks 6, 7 and 8, each have wealth level 15.

After the effort task, your ranking will be determined, and the the computer will flip a virtual coin. If it comes up heads, all citizens will have their initial wealth determined by their ranking in the task as shown in the table above. If it comes up tails, every citizen's wealth will be randomly shuffled and intial wealth will be based completely on luck.

Figure 4: Initial Wealth

Selecting a Tax Rate for the Group

citizens' initial wealth levels help to determine their earnings in the experiment by serving as a potential resource that they may invest in an economic activity. Exactly how much each citizen earns is also determined by (1) the amounts of public goods (analogous to roads, bridges, public transit, and other government services in the real world) that are funded by taxes paid by the citizens, and (2) the amount of tax that each individual pays. In general, you'll earn more if there are more public goods being funded by the tax payments of the group as a whole, and you'll earn more if the amount of tax that you personally pay is lower, since paying more tax leaves you less to put into your own economic activity.

Once all citizens have been informed of their initial wealth levels, the group will collectively decide the tax level (from 0-100%) that will be binding for all citizens. Each person will vote for a tax rate that they would like to be applied to all citizens, themselves included. Once the votes are submitted, the median of all submitted preferred tax rates will be determined and selected as the tax rate applicable to the group.

How is the median determined?

In the context of nine values, the median M is defined as the value such that there are five numbers that are smaller than or equal to M; and there are five numbers that are greater than or equal to M. For example, suppose you and your friends must all choose an amount to contribute to pay for the food at a party. You select $5, and the other eight citizens of your group have chosen: $1, $4, $8, $9, $10, $7, $8, $2. To find the median number, you can sort from low to high as in the table below, then find the number in the middle position (position 5). That is the median.

Number Order	1	2	3	4	**5**	6	7	8	9
Contribution	$2	$3	$3	$6	**$6**	$9	$10	$12	$16

The median in the example would be $6, which is the middle amount in a sorted list. When you and other citizens select a tax rate, the median amount will be computed in this fashion.

Figure 5: Voting procedure explanation

Payoffs

The payoff takes into consideration the resources that have been paid to the government. In particular, we consider the direct benefit from public goods (ex.: the benefit of social security, or unemployment insurance) and the indirect benefit via positive effects of public goods on our own private productivity (ex.: the benefit that companies receive from having roads and private property rights).

To better understand how these payoffs work, please see the table below that shows a citizen's potential payoff based on their initial wealth level and tax rate.

Initial Wealth (Points)	Tax Rates (%)										
	0	10	20	30	40	50	60	70	80	90	100
9	45	60	72	84	98	115	116	100	85	70	54
15	75	98	117	132	147	163	156	131	105	80	54
25	125	163	192	213	230	242	224	182	139	97	54
40	200	260	304	334	353	362	326	258	190	122	54
80	400	518	604	658	683	680	598	462	326	190	54
125	625	809	941	1021	1053	1038	904	692	479	267	54

Similarly, the following graphs show the payoffs for every tax rate. Each wealth level has its own graph where you can see the relationship between the round payoff and the group-chosen tax rate.

Figure 6: Page with payoff table

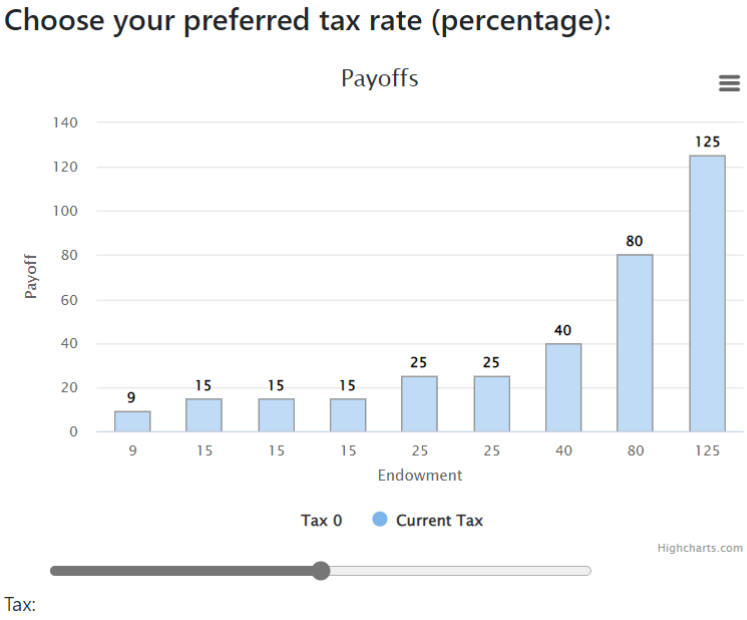

Figure 7: Tax rate system

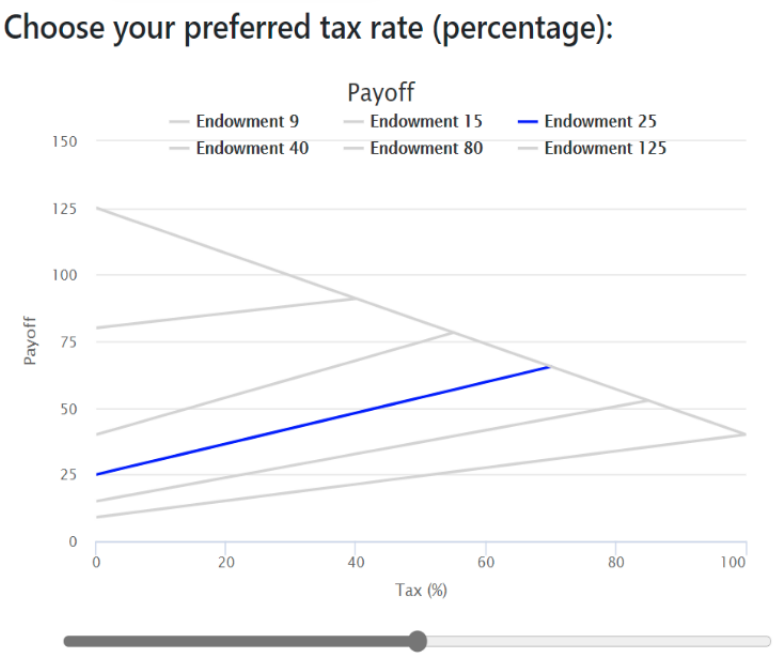

Figure 8: Full distribution of wealth for each tax level

6.2. *Timeline for the Experiment*

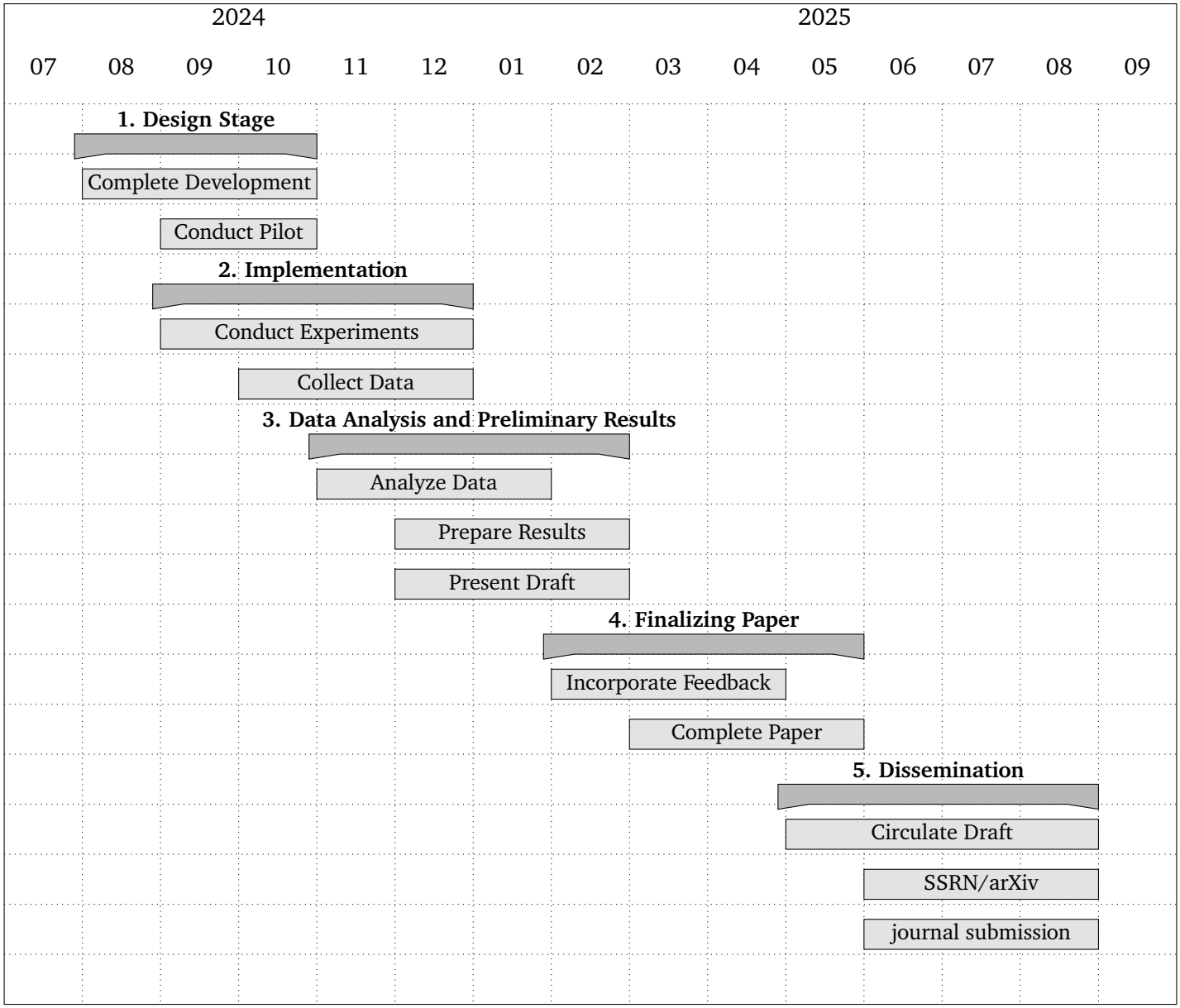

Figure 9: Experiment timeline