Polarization in Group Interactions*

Ganesh Iyer
University of California, Berkeley

Hema Yoganarasimhan
University of Washington

March 30, 2017

*We thank Przemek Jeziorski, Yuichiro Kamada, Sridhar Moorthy, Jiwoong Shin, Jidong Zhou and seminar participants at Carnegie-Mellon University, Columbia University, Harvard University, MIT, University of Toronto and Yale University. Authors contributed equally and are listed alphabetically. Email: giyer@haas.berkeley.edu, hemay@uw.edu.

Polarization in Group Interactions

ABSTRACT

We study the phenomenon of strategic polarization in group interactions. Agents with private preferences choose a public action (e.g., voice opinions), and the mean of their actions represents the groups outcome. They face a trade-off between influencing the group outcome and truth-telling. In equilibrium, agents strategically shade their actions towards the extreme leading to polarization. The group outcome is also more extreme than the mean preference. Compared to a simultaneous actions game, randomized or exogenous sequential actions lowers polarization when agents' preferences are relatively similar. Endogenizing the order of moves always increases polarization, though it is also welfare enhancing.

I Introduction

Society consists of many formal and informal decision-making groups that shape its social, political, and cultural landscape. These might include community groups, senate committees, juries, political action committees, education boards, faculty bodies, and discussion forums. Socio-political opinions of individuals are often subject to influence by others in their social group, and in turn individuals' actions may be designed to influence the behavior of others.

A natural premise of group deliberations going back to Rawls (1971) is that allowing for discussions among members can result in the exchange of differing views thereby promoting greater alignment of opinions. Nevertheless it is common for individuals to enter a group interaction and become more divergent in their opinions. Indeed, a body of experimental evidence in social psychology shows that instead of enabling greater alignment, deliberations often lead groups to become more *polarized* (Isenberg, 1986). As an actual example, after the Newtown shooting in December 2012 citizens in various community and local groups across the country came together to deliberate and exchange opinions on various aspects of gun ownership (Kaufman, 2013). And in a number of instances these group deliberations exacerbated the divide in the opinions of the participants.

This paper examines the phenomenon of polarization of the actions of agents who interact on political, social and even moral issues: i.e., when members of a group take actions (e.g., voice opinions) that are more extreme in the direction of their pre-deliberation preferences. This characterization of polarization of the observed actions of individuals in group interactions is consistent with the idea of polarization of decisions described in the existing literature (Sunstein, 2002). Consider the following examples of group deliberations which illustrate relevant aspects of the phenomenon:

- In 2015, the state of Texas passed the *campus carry law*, formally known as Senate Bill 11 (SB 11) (Aguilar, 2016). Following this, the University of Texas assembled a nineteen member working group to discuss how to implement this law into practice. The working group members with disparate views deliberated on implementation dimensions such as where to allow guns on campus, age limits, and the manner in which guns may be carried on campus. The deliberations led to some unexpected outcomes such as the conclusion that the committee would not recommend for a ban on guns in the classroom (Campus Carry Policy, 2015).[1]

- In 2005, the Kansas education board discussed and approved changes to the way evolution is taught in public schools. The board required science textbooks to mention *intelligent design* and qualify that evolution is simply a theory and not a proven fact (Slevin, 2005; Kansas State Education Board, 2006).

These examples highlight some themes which are focus of our analysis: First, they represent contexts in which agents have deeply held preferences about social, political, or moral issues. Many socio-cultural debates that dominate our public policy discussions also fall under this category – the abortion debate, how much immigration to allow, should there be separation of church, and state and the size of government. In these cases, individuals participating in a group deliberation may find themselves facing a natural trade-off: While an individual would find it costly to deviate from her deeply held preference, she may still want to move the overall group outcome towards her preference.

Second, the action is not necessarily binary; rather it is a choice of the extent or magnitude of an implementable outcome. In the campus carry example, the group's choice on what to recommend was not a simple "should guns be allowed on campus or not?." Rather it was a nuanced decision on where it would be allowed (classrooms), not allowed (child-care units), and allowed with discretion (single-user offices), where it can be stored (in-person or locked vehicle), who is it allowed for (over 21 years, with license). Finally, the group outcomes after deliberations may end up looking more extreme than one would expect *a priori*. Even gun rights advocates may balk at the idea of allowing guns in classrooms, and many religiously inclined individuals may not advocate teaching intelligent design in science classes. In both the examples above, there were members who believed the group's eventual stance to be unduly extreme (Campus Carry Policy, 2015; Slevin, 2005). This pattern of polarized group outcomes is not uncommon. A cursory reading of current news would also suggest that many socio-political issues show evidence of polarization despite

[1] See Huitlin (2015) and Armed Campuses (2016) for comprehensive discussions of campus carry laws across different states and the recent developments in this area.

the presence of moderating influences (Cohn, 2014).

We develop a theory of strategic polarization of group outcomes and the analysis has two related objectives: First we connect the emergence of polarized decision outcomes to strategic motives of individuals in group interactions. When deliberations within a heterogenous group are about deeply held political, social or religious convictions, then individuals can display a natural motivation to move the ultimate group outcome in the direction of their preference. But while taking actions to influence the group outcome, an individual might also incur costs of deviating from her preference. The model connects this trade-off between the desire for group influence and truth-telling to polarization at the individual and the group level, i.e., where individuals take actions that are more extreme than their true preferences and the aggregate group outcome is more extreme than the mean of the group's true preferences.

Second we consider how the timing of actions of the agents affects polarization. In group deliberations members may voice opinions simultaneously without observing each other's opinions. Alternatively, they may voice their opinions sequentially in which case those who speak/act later will be able to observe the opinions of those who spoke before.[2] We ask how the timing of actions affects the degree of polarization. Would the degree of polarization of expressed opinions be greater when members speak simultaneously or in sequence? If individuals were to speak sequentially who has the greater incentive to speak first, those who are more extreme or those who are more moderate?

We develop a model of group interactions where agents have heterogeneous preferences over an issue. The basic analysis considers a group of two agents. Each agent's utility function has two components: First, an agent incurs disutility if she chooses an action (voices an opinion) that is different from her true preference. This is represented as a convex cost which is increasing in the extent of the misalignment between her action and

[2] An example is the deliberations within the Federal Open Market Committee which sets the short term federal funds rates (Lopez-Moctezuma, 2016). Members of the committee with heterogenous preferences (due to differing views on inflation, unemployment, or output growth) express their preferred policy position on the target rate sequentially and in an order that varies across meetings. The committee chairperson then summarizes these positions into an overall group directive.

her true preference. This can be seen as a reputational (or even a psychological) cost of misreporting her true preference. Second, an agent cares about the distance of the group's outcome to her true preference. This represents the group influence motive – individuals would like to move the group's outcome towards their true preference. The game consists of each agent privately observing her true preference and choosing a publicly observable action (opinion). The mean of their public actions represents the group's outcome. Within this structure we investigate how the strategic behavior of the agents and the timing of actions (simultaneous vs. sequential) affect the degree of polarization.

In the simultaneous game agents indulge in *strategic shading* represented by the distance between their true preferences and their observed actions. Shading towards the extreme or polarization occurs at both individual and group levels. Further, agents with extreme preferences shade more than moderates, because they anticipate that the group outcome is likely to be farther away from their preference. Thus while all groups tend towards extremity, groups whose average preferences are more extreme tend to shade even more. Importantly, an agent's incentive to shade in the simultaneous game is not a function of the polarization of the true preferences, i.e, the extent of shading is independent of the preference distribution. In other words, the model is designed to show that the polarization of observed actions does not necessarily stem from polarization of preferences, but rather from the strategic motivations of the agents.

The analysis of the timing of actions and the comparison of the simultaneous and sequential choice games establishes some of the important results of the paper. With sequential actions polarization occurs whenever the agent who moves later is relatively more extreme compared to the first agent, whereas moderation occurs if the agent who moves later has less extreme preferences. The second agent's motivation looms larger on the joint outcome because she can condition her action on the observed action of the first agent and pull the outcome closer to her preference. Related to this, the analysis also highlights the rationale for why agents have the incentive to wait to react to the actions of others rather than to move first and get the benefit of setting the agenda. By moving later an agent can observe earlier actions and compensate for them accordingly. This comes with the benefit that if the previous action was not too extreme, then the second agent need not incur the cost of exaggerating her actions. In contrast, if an agent were to move first, then in anticipation of the second period, she would have to become more extreme which reduces her utility. Given group influence motive, waiting to react to the actions of others is more attractive than attempting to set the agenda. The comparison of the different timings shows that if the preferences of the agents are relatively similar then the mean outcome is more moderate in the sequential actions game, whereas if the agents have dissimilar preferences it is the simultaneous actions game that results in more moderate group outcomes.

We also endogenize the timing of actions by allowing each agent to participate in a first-price sealed bid auction for the right to determine the sequence in which the agents will choose actions (i.e., the speaking order). Agents with more extreme preferences bid more for the right to determine the speaking order, and upon winning all agents regardless of their preferences prefer to wait. More importantly, because the more extreme agents bid more, the group outcome in the endogenous sequential game is always polarized. But the extent of polarization can be higher or lower as compared to the simultaneous game: when the players preferences are relatively similar endogenizing the speaking order leads to less polarization as compared to the case when the agents take simultaneous actions. The implication is that when the players are similarly inclined, allowing for endogenous timing helps to reduce group polarization.

II Related Research

A stream of research in social psychology starting from Lord et al. (1979) shows that groups of individuals who hold differing opinions about social or political issues use available information in a biased manner, by incorporating confirming evidence more readily than disconfirming evidence. This it is argued may make individuals to move further apart in their beliefs after viewing the same evidence.[3] Consistent with these experimental findings Rabin and Schrag (1999) propose a model of confirmatory bias where agents ignore signals

[3]Similar effects have been argued for through experiments in various contexts ranging from social class stereotyping (Darley and Gross, 1983), opinions on the nuclear deterrence (Plous, 1991), and the evaluation of the state of the economy (Kinder and Mebane, 1983)

which do not confirm with their initial impression, and update in the direction of their current beliefs generating polarization.

Several recent papers have provided different economic explanations for belief polarization. Dixit and Weibull (2007) analyze model of Bayesian updating by agents with heterogeneous normally distributed priors about a true (policy) state and a common noise. In this set-up while the mean beliefs of the group may diverge after observing the common signal, the individual beliefs are not farther apart according to stochastic dominance and so individual level polarization does not occur under Bayesian updating. Nevertheless, polarization of individual beliefs is shown to occur by Baliga et al. (2013) due to ambiguity aversion of individuals who observe the common signal. A paper by Acemoglu et al. (2009) considers a Bayesian learning problem for agents with different priors who are uncertain about the conditional distribution of signals and show that even a tiny amount of signal uncertainty can lead to significant disagreement in asymptotic beliefs.[4] Finally, Benoit and Dubra (2016) consider a model in which there is a main issue on which agents update their beliefs, but also private information on an unrelated ancillary issue which affects the interpretation of additional information. Polarization may arise even from rational Bayesian updating due the way different agents select themselves in interpreting the additional information. In contrast to this literature our analysis is about the polarization of observed actions which arises from the strategic incentives of agents in group interactions. Further, the trade-off between the incentive to influence the group outcome and truth-telling can be seen as a general rationalization of observed polarization which is independent of the different contexts that motivate the studies in psychology.

There is a substantial literature on the measurement of polarization and its linkage to social and political conflicts. An important paper by Esteban and Ray (1994) provides an axiomatic characterization of polarization as a measure of within group homogeneity of individuals on some variable (such as income or wealth) but across group heterogeneity. The idea is that the polarization measure is more closely linked to conflict in society than

[4] Other papers in this area include Kondor (2012) who shows that belief polarization can be generated when agents see different private signals that are correlated with a common public signal. A similar idea is present in Andreoni and Mylovanov (2012).

common measures of diversity such as inequality or fractionalization.[5] The subsequent empirical work on conflict has investigated its relationship to ethnic and social diversity and shows that ethnolinguistic fractionalization does not have a significant effect on the probability of civil wars; see for e.g., Fearon and Laitin (2003). However, Montalvo and Reynal-Querol (2005) go on to show that polarization has a significant link to conflict and the incidence of wars. Obviously the focus of our analysis is not the measurement of polarization, but rather on how polarization of actions arises in strategic interactions of agents who care about group influence and truth-telling.

Polarization has also been analyzed in the large literature on political and electoral competition and much of it is concerned with how voter preference distributions affect candidate behavior. While the classic Hotelling-Downs framework shows the logic for why candidates may converge to the median voter preferences, subsequent research has examined rationales for the prevalence of policy divergence in observed electoral competitions. For example, Alesina (1988) shows that if candidates care not only about being elected, but also about which policy is ultimately implemented, they may announce convergent platforms before elections and then implement their preferred platform ex-post. Roemer (1994) shows that uncertainty about voter preferences can lead candidates to diverge, while Glaeser et al. (2005) show that political parties may strategically choose more extreme platforms to target their core market that is more influenced by the candidate's information. Kamada and Kojima (2014) consider a Hotelling-Downs setup in which voters have convex preferences and show that this can lead candidate policies to diverge in equilibrium. While these papers analyze a game between strategic candidates dealing with voters, we consider a game where agents are themselves voters but who can take strategic actions to influence other agents.[6]

[5] It may be noted that the Esteban and Ray (1994) polarization measure is consistent with the identification-alienation framework: i.e., individuals care about identification with others of similar income, and are alienated from those who are dissimilar; see Duclos et al. (2004).

[6] Polarization has also been studied in the context of firms' incentives to slant media. Mullainathan and Shleifer (2005) show that media firms slant news when consumers have a taste for seeing their beliefs confirmed while Gentzkow and Shapiro (2010) provide a reputation based rationale for media bias even with Bayesian consumers.

III Model Preliminaries

We first present the basic model of group interactions, where the mean of actions of the agents is seen as the group outcome. Consider a group of two agents i and j, where each agent's preference (denoted as x_i and x_j) is independently drawn from a distribution $g(x)$, which is symmetric around zero and with support over the real line $\mathbb{R}$. The cumulative density of the distribution is given by $G(x) = \int_{-\infty}^{x} g(t)dt$ and $G(\infty) = 1$.[7]

Agent i's true preference or type x_i is her private information and in the introductory examples this would represent the agent's true conviction on religious or political issues such as gun rights, or the nature of science education in schools, or the size of government. Both agents simultaneously choose a publicly observable action, $\{a_i, a_j\} \in \mathbb{R}$. As already indicated before, in a group deliberation an agent's action can be interpreted as her voiced preference or opinion. After both agents have spoken, assume that a neutral third-party implements the mean of their voiced preferences or actions as the group outcome.

The utility of agent i is given by the following convex loss function:

$$u(x_i, a_i, a_j) = -r(x_i - a_i)^2 - (1-r)(x_i - \bar{a})^2 \tag{1}$$

where $\bar{a} = \frac{a_i + a_j}{2}$, and $r \in (0, 1)$ represents the relative weights that the agents places on the different components of their utility. Agents obtain dis-utility from two sources. First, their utility is decreasing in the distance between their action and their preference, *i.e.*, they prefer to voice opinions close to their true preference. This could stem from a disinclination to misreport their preferences or from potential reputational concerns. Second, their utility is decreasing in the distance between the group outcome ($\bar{a}$) and their true preference. In other words, agents' have a taste for influencing the group's outcome. In the case of the University of Texas working group on the campus carry issue, the chairman of the committee would have to produce a recommendation for the University administrators which is a summary of the average group opinions. Therefore, in the deliberations members of the committee would have the incentive to take actions that would sway the recommendations towards their preference on the issue. A greater value of r represents issues for which agents have stronger

[7]In the analysis, to illustrate some of the results we use as an example preferences that are independently drawn from $U[-1, 1]$ (and actions $\{a_i, a_j\} \in \mathbb{R}$).

relative preference for voicing opinions that are consistent with their true preferences.

We consider a game in which nature first draws the preferences x_i and x_j for the agents based on which they choose their publicly observable actions. The actions a_i and a_j may be chosen simultaneously in which case each agent's choice is contingent only on the private information about her preference. Alternatively, the agents may move sequentially in which case the agent who moves second will be able to choose her actions contingent upon her private information as well as the observed actions of the first mover.

IV Analysis

Before we begin the analysis it is useful to derive two benchmark cases – i) the first-best socially optimal solution, and ii) the perfect information case. In the first case, a social planner chooses actions to maximize the joint surplus of the two agents:

$$W(x_i, x_j, a_i, a_j) = \sum_{k=i,j} -r(x_k - a_k)^2 - (1-r)(x_k - \bar{a})^2 \tag{2}$$

The welfare maximizing choices are $a_i^* = x_i$ and $a_j^* = x_j$. The socially optimal action for both agents is truth-telling and the joint decision shows no distortion from the preferences. Suppose now that the agents have perfect information on each other's types and move simultaneously. Denoting the agents' equilibrium actions as $\{a_i^p, a_j^p\}$, we can derive: $a_i^p = \frac{3r+1}{4r}x_i - \frac{1-r}{4r}x_j$ and $a_j^p = \frac{3r+1}{4r}x_j - \frac{1-r}{4r}x_i$. While both agents deviate from truth-telling by reporting a weighting of their own preference and the other agent's preference, the mean action, $\bar{a}^p = \frac{x_i+x_j}{2}$, perfectly reflects the mean preferences of the group. Thus, with perfect information too the group's joint decision is not distorted.

A Simultaneous Actions

Consider the game in which agents choose their actions without observing the other agent's type and actions. We proceed to derive the Bayesian Nash equilibrium of this game and focus without loss of generality on agent i. Let $\hat{a}_j$ denote the equilibrium action of j. Because j's preference (x_j) is her private information at the time of choosing the action, i's expected utility from choosing a_i as $EU(x_i, a_i) = \frac{\int_{\mathbb{R}} u(x_i, a_i, \hat{a}_j) g(x_j) \mathrm{d}x_j}{\int_{\mathbb{R}} g(x_j) \mathrm{d}x_j}$. By differentiating $EU(x_i, a_i)$

and setting it equal to zero at i's equilibrium action $a_i = \hat{a}_i$ gives us:

$$\left.\frac{\partial EU(x_i, a_i)}{\partial a_i}\right|_{a_i=\hat{a}_i} = 2r(x_i - \hat{a}_i) - \frac{(1-r)}{2}\left[-(2x_i - \hat{a}_i) + \frac{1}{2}\int_{\mathbb{R}} \hat{a}_j g(x_j)\mathrm{d}x_j\right] = 0 \quad (3)$$

In obtaining the above first order condition, we can set $\frac{\mathrm{d}\hat{a}_j}{\mathrm{d}a_i} = 0$ because in a simultaneous equilibrium, any change in the action of agent i has no impact on the equilibrium action of agent j. Simplifying equation (3) gives us $\hat{a}_i$ as:

$$\hat{a}_i = \frac{2(1+r)}{1+3r}x_i - \frac{(1-r)}{2(1+3r)}\int_{\mathbb{R}} \hat{a}_j g(x_j)\mathrm{d}x_j \quad (4)$$

Integrating i's equilibrium action $\hat{a}_i$ over the entire range of x_i gives us:

$$\int_{\mathbb{R}} \hat{a}_i g(x_i)\mathrm{d}x_i = \frac{2(1+r)}{1+3r}\int_{\mathbb{R}} x_i g(x_i)\mathrm{d}x_i - \frac{(1-r)}{2(1+3r)}\int_{\mathbb{R}} \hat{a}_j g(x_j)\mathrm{d}x_j \int_{\mathbb{R}} g(x_i)\mathrm{d}x_i \quad (5)$$

Because $\int_{\mathbb{R}} \hat{a}_i g(x_i)\mathrm{d}x_i = \int_{\mathbb{R}} \hat{a}_j g(x_j)\mathrm{d}x_j$, and because $E(x) = 0$ for a symmetric distribution, we can uniquely identify $\int_{\mathbb{R}} \hat{a}_j g(x_j)_j \mathrm{d}x_j = 0$. We thus have $\hat{a}_i = \frac{2(1+r)}{1+3r}x_i$. These results are summarized below in the proposition below: 1.

Proposition 1 *In the simultaneous actions game, there exists a unique Bayesian Nash equilibrium, where an agent i with preference x_i chooses action $\hat{a}_i = \frac{2(1+r)}{1+3r}x_i$.*

An implication of the Proposition is that agents' actions are more extreme than their true preferences (the multiplier $\mu(r) = \frac{2(1+r)}{1+3r} > 1$, for all $r < 1$). Moreover, this shift to extremity is in the direction of their original preference: *i.e.*, those with positive x_i always move right, while those with negative x_i always move towards the left. When picking the optimal action, agent i's calculation of the expected action of the other agent will be $E(\hat{a}_j) = 0$. Consider the trade-off faced by the agent if she chooses to report her true preference and choose $a_i = x_i$: Given this choice she expects the mean of the actions to be $E_i[\bar{a}] = x_i/2$ and the distance between her preference and the mean to be $E_i[x_i - \bar{a}] = x_i/2$. We know that i's utility is decreasing both in the distance between her type and the mean action and in the distance between her type and her action. By reporting $a_i = x_i$ the agent does not incur any cost from misreporting, and her expected loss is purely the cost of the joint outcome being misaligned with her preference, $EU(x_i, x_i) = -\frac{1-r}{4}x_i^2 - \frac{\mu(r)^2}{4}E(x^2)$, where $E(x^2) = \int_{\mathbb{R}} x_j^2 g(x_j)dx_j$. If instead, she exaggerates her opinion by ϵ in the direction away from zero, she successfully moves the mean closer to her own preference x_i. However in doing so, she also incurs an extra cost from lying, which is increasing with ϵ. Overall, her expected utility is $EU(x_i, x_i+\epsilon) = -r\epsilon^2 - \frac{1-r}{4}(x_i-\epsilon)^2 - \frac{\mu(r)^2}{4}E(x^2)$. For small values of ϵ, $EU(x_i, x_i+\epsilon) > EU(x_i, x_i)$ and the converse is true for ϵ large enough. Hence, in equilibrium, i picks the optimal value of a_i that minimizes the loss from the distance between the group's outcome and their own preference, but one that does not inflate the cost of exaggerating.[8]

Next, recall that group polarization is defined as the tendency of the joint outcome to move towards a more extreme point in the direction indicated by the members' original preferences. The equilibrium derived above satisfies this definition. The mean pre-deliberation preference of the group is $\bar{x} = \frac{x_i+x_j}{2}$ while the mean post-deliberation outcome is $\bar{a} = \frac{\hat{a}_i+\hat{a}_j}{2} = \frac{1+r}{1+3r}(x_i+x_j)$. If $\bar{x} > 0$, then $\bar{a} > \bar{x}$; else if $\bar{x} < 0$, then $\bar{a} < \bar{x}$. Hence, if the preferences of the two agents in the group is initially predisposed towards the right, the joint outcome is even more rightwards. Alternatively, if the group is predisposed towards the left, then its joint outcome is even more leftwards.

To investigate the comparative statics, consider the extent to which an agent i shades her opinion in equilibrium defined as $s_i = |\hat{a}_i - x_i| = \frac{1-r}{1+3r}|x_i|$. We can see that $\frac{\mathrm{d}s_i}{\mathrm{d}r} \leq 0$ and as would be expected agents shade their actions less if the cost associated with lying is higher. Second, $\frac{\mathrm{d}s_i}{\mathrm{d}|x_i|} > 0$ suggests that agents near the extremes shade more than moderates, who are closer to the center. It also implies that the overall group shift is proportional to the initial tendency of the group. The mean shift of a group is given by $\bar{s} = |\bar{a} - \bar{x}| = \frac{1-r}{1+3r}|\bar{x}|$ and so the shift exhibited by an extreme group of agents is higher than that exhibited by a relatively moderate group. While all groups tend towards the extremes in their decisions, this effect is exacerbated in extreme groups.

[8] We have not explicitly included abstention as part of the players' strategy set. Abstention can be seen as equivalent to not voicing any opinion. The analysis above would hold if we assume that not voicing any opinion implies an action that is consistent with true preferences. When agent i abstains, her actions are aligned with her true preferences and so her utility from the first term to $-r(x_i - a_i)^2 = 0$. However, by abstaining, she has no effect on the group's final outcome and $\bar{a} = a_j$. So the utility from abstaining is $-(1-r)(x_i - a_j)^2$ which is strictly lower than the utility of voicing her true preferences (and obtaining $-(1-r)\left(x_i - \frac{x_i+a_j}{2}\right)^2$.) Thus abstaining is always a dominated choice.

An interesting point is that the extent of shading, s_i, is independent of the distribution $g(x)$ as long as it is symmetric. Agents' shade the same amount irrespective of whether they are drawn from a uniform distribution or from a more polarized preference distributions where the masses are near the extrema. This suggests that polarization in this model does not stem from preference polarization. Rather it is a strategic choice made by agents in order to influence the group decisions.

Asymmetric Type Distribution: If we consider any general asymmetric distribution $g(x)$, then we can derive: $a_i = \frac{2(1+r)}{1+3r} x_i - \frac{1-r}{1+3r} E(x)$. The effect of the asymmetry of the type distribution is intuitive. Suppose $x_i > 0$ and agent i has right leaning preferences, but that the distribution of the agent types is skewed in the opposite direction, i.e., $E(x) < 0$. In this case, agent i will have the incentive to be more extreme than in the symmetric case and to shade her action even more to the right. In contrast, if the distribution is skewed in the same direction as the agent's preference, this leads to moderation.

A.1 Group Size Effects

We now consider interactions between $m > 2$ agents to investigate the role of the group size. Recall that the preferences of the agents are private information and are independently drawn and all agents simultaneously choose their public action $\{a_{\mathrm{m},i}, a_{\mathrm{m},j}, ...\} \in \mathbb{R}$, with agent i's action denoted as $a_{\mathrm{m},i}$. Agent i's utility is given by $u(x_i, a_{\mathrm{m},i}, a_{\mathrm{m},-i}) = -r(x_i - a_{\mathrm{m},i})^2 - (1-r)(x_i - \bar{a}_{\mathrm{m}})^2$, where $a_{\mathrm{m},-i}$ denotes the actions of all agents except i and $\bar{a}_{\mathrm{m}} = \frac{1}{m} \sum_{i=1}^{m} a_{\mathrm{m},i}$.

In the Bayesian Nash equilibrium of this game agent i chooses action $\hat{a}_{\mathrm{m},i} = \frac{rm^2+(1-r)m}{rm^2+(1-r)} x_i$. This shows the robustness of the equilibrium of the two-agent group. Agent i's strategy is linear in her type x_i with the multiplier $\mu_{\mathrm{m}}(r) = \frac{rm^2+(1-r)m}{rm^2+(1-r)}$. The extent of shading by agent i is $s_{\mathrm{m},i} = |\hat{a}_{\mathrm{m},i} - x_i| = \frac{(1-r)(m-1)}{rm^2+(1-r)} |x_i|$ and so as in the basic model, for all m, r, agents near the extreme shade their opinions more than those near the center. The extent of shading is also increasing in r, and independent of $g(\cdot)$.

The m-agent case provides an additional insight – it shows that the level of shading s_i is increasing in m up to $\frac{1}{\sqrt{r}} + 1$, but decreasing after that (see Appendix). In other words, after a certain point, as the number of players increases, agents tend to shade less. As $m \to \infty$, shading goes to zero, *i.e.*, players report the truth. Agents exaggerate in order to pull the mean ($\bar{a}$) closer to their preference. But when the number of agents in the group becomes large, the marginal impact of any one agent's action on the group mean outcome becomes negligible. Hence, in very large groups, the incentive to exaggerate is low. This indicates a plausible solution to the polarization problem – a social planner wishing to reduce polarization of actions may do so by picking larger deliberation groups. However, while group size can be a potential remedy, it may also have practical limits. For example, involving large groups in decision-making is likely to be costly to implement in many cases and may simply be infeasible in others (such as in juries and political committees). Therefore, later in the paper we will analyze the role of timing of the actions and ask whether sequential choices may be a potential solution to the polarization problem.

A.2 Sub-group Interactions

In many socio-political situations deliberations occur between sub-groups, where each subgroup consists of many agents having the same preferences over an issue, but different from that of the other sub-group. Political deliberations in the U.S., *e.g.*, in the Senate, occur between multi-agent Democratic and Republican sub-groups. On issues such as abortion rights, gun control, and taxation, conservatives groups have different preferences than liberals, but citizens within each sub-group have similar preferences.

Consider an extension of the basic model with a population of $m > 2$ agents that are divided into two subgroups 1 and 2 of sizes n_1 and n_2 so that $n_1 + n_2 = m$. The sub-group sizes are common knowledge. The preferences x_1 and x_2 of the sub-groups are independently drawn from $g(x)$. All agents within a sub-group know their individual preferences (and that of the others within their sub-group), but they do not observe the preferences of the other sub-group. We can write the expected utility of an individual agent i from sub-group 1 and agent j from sub-group 2 as:

$$EU_1^i = -r(x_1 - a_1^i)^2 - (1-r)\left(x_1 - \frac{1}{m}\left(a_1^i + \sum_{k_1=1(\neq i)}^{n_1} a_1^{k_1} + \sum_{k_2=1}^{n_2} \int_{\mathbb{R}} a_2^{k_2} g(x_2)dx_2\right)\right)^2 \tag{6a}$$

$$EU_2^j = -r(x_2 - a_2^j)^2 - (1-r)\left(x_2 - \frac{1}{m}\left(a_2^j + \sum_{k_2=1(\neq i)}^{n_2} a_2^{k_2} + \sum_{k_1=1}^{n_1} \int_{\mathbb{R}} a_1^{k_1} g(x_1)dx_1\right)\right)^2 \tag{6b}$$

In the Appendix we present the solution for the symmetric (for agents within a sub-group) Bayesian Nash equilibrium and show that the equilibrium actions are $\hat{a}_i(x_i, m, n_i) = x_i \frac{m(mr+(1-r))}{m^2r+(1-r)n_i}$ and $\hat{a_j}(x_j, m, n_j) = x_j \frac{m(mr+(1-r))}{m^2r+(1-r)n_j}$. The main results of the two-agent model continue to hold: the sub-group's action is linear in its preferences and sub-groups near the extremes shade more.

The main point of this analysis is to understand the role of the sub-group size on actions. Specifically, for a given population size m, how would a sub-group's size affect actions. It can be seen that for a given m, $\frac{\partial \hat{a_i}}{\partial n_i}$ and $\frac{\partial \hat{a_j}}{\partial n_j}$ are both negative. The implication is that smaller sub-groups can become even more extreme. This result can be seen as being consistent with the role of the Tea party movement in U.S. politics which was associated with pulling the Republican party more to right and with adopting increasingly conservative economic and social positions. For example, the Tea party members in the senate adopted increasingly conservative positions on environment, trade, budget, and immigration (Todd et al., 2014). This has happened even as the percentage of tea party supporters reported by polls diminished from 30% at the beginning of 2011 to 17% in October 2015 (Gallup, 2015).

B Sequential Actions

We now consider the case in which players may voice their opinions in sequence. Indeed in many legal or political institutions members typically take turns to speak. As described in the introduction in the Federal Open Market Committee (FOMC) meetings the members of the committee express their preferred policy position sequentially. The committee Chairman summarizes these positions into an overall group directive. Similarly, in juries and legislative bodies the order of speaking is often pre-determined by the institutional rules. Accordingly, we consider a two-period model in which one of the agents is randomly picked to speak in the first period and the other follows in the second period. We refer to this model as the "exogenous" sequential choice model where the order of agent actions is exogenously determined and is uncorrelated to the agents' preferences. The speaking order can be interpreted as being either determined by institutional rules or by a third-party. Then in a subsequent section we will consider the case when the agents bid to endogenously choose the speaking order.

B.1 Equilibrium in the Sequential Game

Let $a_{\mathrm{xt},i}$ denote agent i's action in period t, in this e**x**ogenous sequential actions game. Without loss of generality, suppose that agent i speaks in the first period and j in the second period. We solve for the Perfect Bayesian equilibrium (PBE) for this game, and derive the equilibrium actions of both players starting with the second player. A PBE consists of strategy profile (and associated beliefs) for the two agents that specify their optimal actions given their beliefs and the strategies of the other agent. Further, the beliefs of each agent are consistent with the strategy profile and are determined by Bayes rule where possible. In this game, the first agent i's strategy $a_{\mathrm{x1},i}(x_i, \hat{a}_{\mathrm{x2},j})$ is a function of her type x_i and her (consistent) beliefs about the optimal actions of j in period 2, whereas the second agent j's strategy $a_{\mathrm{x2},j}(x_j, a_{\mathrm{x1},i})$ is a function of her type and the action of player i that she observes.

Period 2 – The utility of the second player j when she chooses action $a_{\mathrm{x2},j}$ in response to the first player's observed action $a_{\mathrm{x1},i}$ is $u\left(x_j, a_{\mathrm{x2},j}, a_{\mathrm{x1},i}\right) = -r(x_j - a_{\mathrm{x2},j})^2 - (1-r)\left(x_j - \bar{a}_x\right)^2$ where $\bar{a}_{\mathrm{x}} = \frac{a_{\mathrm{x1},i} + a_{\mathrm{x2},j}}{2}$. The optimal choice of agent j given the first period choice of i can be derived as $\hat{a}_{\mathrm{x2},j} = \frac{2(1+r)}{1+3r} x_j - \frac{(1-r)}{1+3r} a_{\mathrm{x1},i}$.

Period 1 – We can now solve for i's first period choice. While i doesn't know the second player's type, her belief will be that j will choose an optimal action $\hat{a}_{\mathrm{x2},j}$ in response to her action. So her expected utility from choosing action $a_{\mathrm{x1},i}$ is obtained by taking the expectation of $u\left(x_i, a_{\mathrm{x}1,i}, \hat{a}_{\mathrm{x}2,j}\right)$ over the full range of x_j which gives us:

$$EU_{\mathrm{x}1}\left(x_i, a_{\mathrm{x}1,i}\right) = -r\left(x_i - a_{\mathrm{x}1,i}\right)^2 - (1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{\mathrm{x}1,i}\right)^2 + \left(\frac{1+r}{1+3r}\right)^2 \int_{\mathbb{R}} x_j^2 g(x_j)\mathrm{d}x_j\right] \tag{7}$$

Taking the F.O.C of equation (7) and following a similar analysis to that in section A gives us the equilibrium action of i as $\hat{a}_{\mathrm{x}1,i} = \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i$. Therefore, the first player i has a unique optimal response that is both linear in her type x_i and is symmetric around zero.

B.2 Characterizing the Group Outcome

Having derived the individual equilibrium actions, we now proceed to characterize the mean equilibrium outcome. For a given x_i and x_j, the mean equilibrium outcome is $\bar{a}_{\mathrm{x}} = \frac{\hat{a}_{\mathrm{x}1,i}+\hat{a}_{\mathrm{x}2,j}}{2} = \frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i + \frac{1+r}{1+3r}x_j$. Before stating the results, we define the following relationships:

- *Polarization*: $|\bar{a}_{\mathrm{x}}| > |\bar{x}|$ and $\bar{a}_{\mathrm{x}}\bar{x} > 0$. Polarization is said to have occurred if the mean of equilibrium actions ($\bar{a}_{\mathrm{x}}$) is more extreme (farther away from zero) than the mean of preferences $\bar{x}$, and this shift is in the direction of the group's initial tendency $\bar{x}$.

- *Reverse Polarization*: $|\bar{a}_{\mathrm{x}}| > |\bar{x}|$ and $\bar{a}_{\mathrm{x}}\bar{x} \leq 0$. Reverse Polarization is the case where the mean equilibrium outcome ($\bar{a}_{\mathrm{x}}$) is more extreme than the mean of preferences $\bar{x}$, and the shift is in the direction *opposite* to the group's initial tendency $\bar{x}$.

- *Moderation*: $|\bar{a}_{\mathrm{x}}| \leq |\bar{x}|$. Moderation refers to the case where the mean of the equilibrium actions ($\bar{a}_{\mathrm{x}}$) lies closer to zero than the mean of preferences ($\bar{x}$).

The following proposition summarizes the equilibrium extent of shading as measured by the relationship between the mean actions and preferences:

Proposition 2 *Let* $k_1(r) = \frac{(1+3r)(1-r)}{(1+3r)^2+4r(1-r)}$, $k_2(r) = \frac{(1+3r)\left[(1+3r)^2+16r\right]}{[(1+3r)^2+4r(1-r)][2(1+r)+(1+3r)]}$, *and* $k_3(r) = \frac{(1+3r)^3+8r(1-r)(1+r)}{2(1+r)[(1+3r)^2+4r(1-r)]}$, *and without loss of generality, let* $x_i \geq 0$. *Comparison of the mean equilibrium outcome* ($\bar{a}_{\mathrm{x}}$) *with the mean of preferences* ($\bar{x}$):

- *Polarization occurs if* $x_j > k_1(r)x_i$ *or* $x_j < -x_i$.

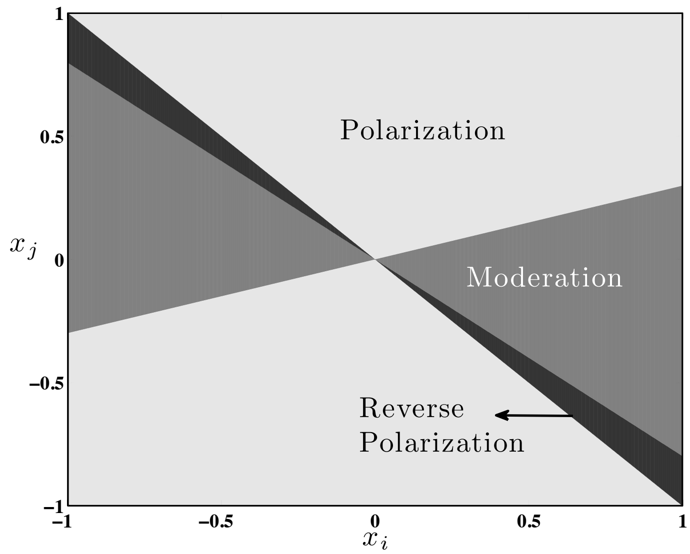

Figure 1: Regions of polarization, reverse polarization, and moderation in an exogenous sequential choice game; shown for $\{x_i, x_j\}$ drawn from $U[1, 1]$.

- *Reverse Polarization occurs if* $-x_i \le x_j < -k_2(r)x_i$.
- *Moderation occurs if* $-k_2(r)x_i \le x_j \le -k_1(r)x_i$

Proof: See Appendix □

The effect of sequential actions on group polarization is summarized in Figure 1. Polarization occurs whenever the second player's preference is relatively extreme or comparable to that of the first player *i.e.*, $x_j > k_1(r)x_i$ or $x_j < -k_2(r)x_i$. In a sequential game, the second player can condition her action on that of the first player and is therefore always able to pull the mean outcome $\bar{a}_{\mathrm{x}}$ close to her own preference. When j is extreme, she pulls the mean outcome to the extreme too, thereby leading to polarization. Note that within this region when $-x_i \le x_j \le -k_2(r)x_i$, the polarization is reverse in the sense that it is in the direction opposite to that implied by $\bar{x}$. This happens when x_i and x_j lie on opposite sides of zero, and $|x_i|$ is slightly greater than $|x_j|$. In other words, while the agents have preferences that are on opposite side of the issue, the first mover's preference is only slightly more intense. This implies that the mean group preference $\bar{x}$ lies on the same side of zero as the first mover i. However, in the second period, agent j's optimal action is able to ensure that the mean action $\bar{a}_{\mathrm{x}}$ is closer to her than to i, *i.e.*, lies on the same side of zero as her own preference x_j – opposite to that of x_i and $\bar{x}$. Therefore, in this region, the group's mean action or outcome can be seen as being polarized but in the reverse direction.

In contrast, when the second mover j's preference is closer to zero compared to i, then she can choose her second period action so as to bring the group's outcome closer to her preference. This provides a moderating influence, and the overall outcome is closer to zero than the mean preferences. Thus both polarization and moderation are possible in this exogenous sequential game with the actual outcome depending upon on the relative preferences of both players and leans in the direction of the second player. So if the second player is relatively extreme, the outcome is also extreme; however if she is moderate, the outcome is moderate too.

B.3 Comparing Simultaneous and Sequential Games

We compare the simultaneous and sequential action games to understand how the extent of group polarization is affected by the timing of actions.

Proposition 3 *Let $\{\hat{a}_i, \hat{a}_j\}$ and $\{\hat{a}_{\mathrm{x}1,i}, \hat{a}_{\mathrm{x}2,j}\}$ denote the equilibrium actions of i and j in the simultaneous choice and exogenous sequential choice games, respectively. Without loss of generality, let $x_j > 0$. Then:*

a) $\hat{a}_{\mathrm{x}2,j} \geq \hat{a}_j$ if $x_i \leq 0$ and $\hat{a}_{\mathrm{x}2,j} < \hat{a}_j$ if $x_i > 0$.

b) $|\hat{a}_{\mathrm{x}1,i}| \geq |\hat{a}_i|$ and $\frac{\mathrm{d}|\hat{a}_{\mathrm{x}1,i}|}{\mathrm{d}r} < 0$

Proof: See Appendix □

Consider first the action of the second player j in the sequential game $\hat{a}_{\mathrm{x}2,j} = \frac{2(1+r)}{1+3r} x_j - \frac{(1-r)}{1+3r} \hat{a}_{\mathrm{x}1,i}$. Part ($a$) of the Proposition shows that if the players lie on opposite sides of zero, then j becomes more extreme in the sequential actions game as compared to the simultaneous game. In contrast, when the two players preferences on the same side of zero, then in the second period j is less extreme, in response to i's action. That is, when the first player i chooses an action close to j's own preference, then she is more moderate in comparison to the simultaneous case. This is because exaggeration in the simultaneous case is driven by the anticipation of the other players' opinion. But in the sequential case player j already observes an action that shows that player i is not from the opposite camp and so the incentive to exaggerate decreases.

In contrast, the first player i's action in the sequential game is always more extreme than that in the simultaneous case (*i.e.*, $|\hat{a}_{\mathrm{x}1,i}| \geq |\hat{a}_i|$) because i knows that the second player j can compensate for her action in either direction. This is not an issue when j's preferences are similar to her own. But if the preferences happen to be very different, then by virtue of speaking second, j can nullify the effect of i's actions. Because this effect does not exist in the simultaneous game, i's action in the sequential game is more extreme.

Next, we compare the equilibrium outcomes in the two game formats.

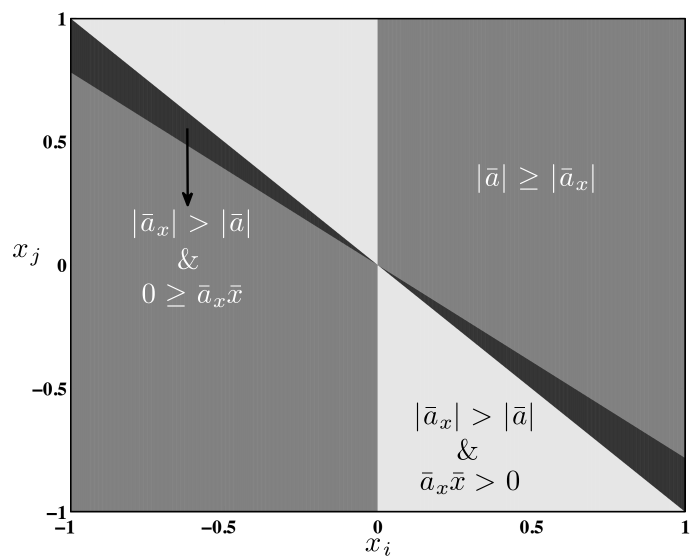

Figure 2: Comparison of mean outcome in the exogenous sequential game with that in the simultaneous game; shown for $\{x_i, x_j\}$ drawn from $U[1,1]$.

Proposition 4 *Comparison of the mean equilibrium outcome in the exogenous sequential game* $\bar{a}_{\mathrm{x}}$ *with that from the simultaneous game* ($\bar{a}$)*:*

- $|\bar{a}_{\mathrm{x}}| > |\bar{a}|$ *and* $\bar{a}_{\mathrm{x}}\bar{x} > 0$ *if* $x_j < -x_i$*, i.e., the mean outcome in the exogenous sequential game is more extreme than that in the simultaneous game, in the direction of the initial tendency* $\bar{x}$*.*

- $|\bar{a}_{\mathrm{x}}| > |\bar{a}|$ *and* $\bar{a}_{\mathrm{x}}\bar{x} \leq 0$ *if* $-x_i \leq x_j < -k_3(r)x_i$*, i.e., the mean outcome in the exogenous sequential game is more extreme than that in the simultaneous game, but in the "opposite" direction of the initial tendency* $\bar{x}$*.*
- $|\bar{a}_{\mathrm{x}}| \leq |\bar{a}|$ *if* $x_j \geq -k_3(r)x_i$*, i.e., the mean outcome in the exogenous sequential game is moderate compared to the mean outcome in the simultaneous game,*

Proof: see Appendix □

If i and j are relatively similar, then the mean outcome is more moderate in the sequential game (*i.e.*, $|\bar{a}_{\mathrm{x}}| < |\bar{a}|$). On the other hand, if i and j lie on opposite sides of zero (are relatively different) and j is relatively extreme, then the mean outcome in the sequential game is more extreme (see Figure 2). Overall, the propositions above reveal the insight that sequential actions may make agents more polarized than in the simultaneous case if they have divergent preferences. But if the individuals have relatively similar preferences, sequential actions has the potential to lead to less polarization of actions. These results highlight how the timing of the game may be exploited to combat polarization. For example, a social planner with an objective to reduce polarization can do so by assigning speaking orders if she expects players to be similar. However, if she expects them to be dissimilar, she may instead opt for a simultaneous choice format.

B.4 Value of the Speaking Order

We now analyze the relative value of the speaking order for the players by comparing the ex-ante expected utilities from speaking in the first and second periods. Agent i's a priori expected utility from speaking first and choosing action $a_{\mathrm{x},i}$ is given by equation (7). In equilibrium, i optimally chooses action $\hat{a}_{\mathrm{x}1,i}$. Substituting this into equation (7) gives us:

$$EU_{\mathrm{x}1}\left(x_i, \hat{a}_{\mathrm{x}1,i}\right) = -\frac{(1-r)(1+r)^2}{(1+3r)^2+4r(1-r)}x_i^2 - \frac{(1-r)(1+r)^2}{(1+3r)^2}\int_{\mathbb{R}} x^2 g(x)\mathrm{d}x \tag{8}$$

Similarly, we can calculate i's a priori expected utility from speaking second as follows:

$$\begin{aligned} EU_{\mathrm{x}2}\left(x_i, \hat{a}_{\mathrm{x}2,i}\right) &= \frac{\int_{\mathbb{R}} u\left(x_i, \hat{a}_{\mathrm{x}2,i}, \hat{a}_{\mathrm{x}1,j}\right) g(x_j)\mathrm{d}x_j}{\int_{\mathbb{R}} g(x_j)\mathrm{d}x_j} \\ &= -\frac{r(1-r)}{1+3r}x_i^2 - \frac{r(1-r)(1+3r)(3+r)^2}{\left[(1+3r)^2+4r(1-r)\right]^2}\int_{\mathbb{R}} x^2 g(x)\mathrm{d}x \end{aligned} \tag{9}$$

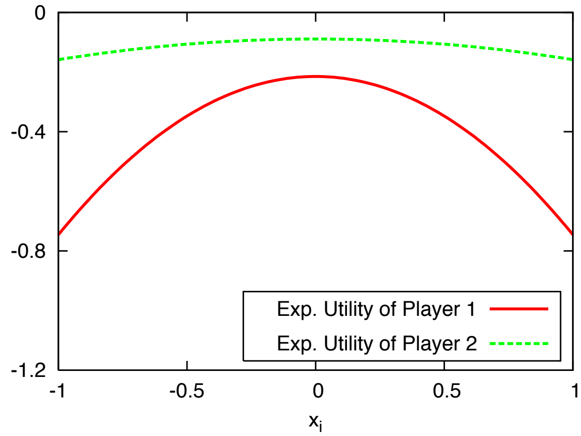

Figure 3: Players' expected utilities in an exogenous sequential choice game. Shown for $\{x_i, x_j\}$ drawn from $U[1, 1]$.

The following proposition compares the equilibrium expected utilities of speaking in the first and second periods, for a player i of type x_i.

Proposition 5 *Let $D_{\mathrm{x}}(x_i) = EU_{\mathrm{x}1}\left(x_i, \hat{a}_{\mathrm{x}1,i}\right) - EU_{\mathrm{x}2}\left(x_i, \hat{a}_{\mathrm{x}2,i}\right)$ denote the difference between the equilibrium expected utilities of agent i from speaking in the first and second periods.*
a) $D_{\mathrm{x}}(x_i) \leq 0 \Rightarrow$ for any agent i, the expected utility from speaking in the second period is greater than or equal to that from speaking in the first period.
b) $\frac{\mathrm{d}D_{\mathrm{x}}(x_i)}{\mathrm{d}|x_i|} \leq 0 \Rightarrow$ the difference in expected utilities of speaking in the first and second periods is increasing in $|x_i|$.
Proof: See Appendix $\square$

The proposition highlights an important trade-off in incentives: Moving first allows an agent to "set the agenda" by committing to an observable action, whereas moving second allows the agent the flexibility to optimally adjust to the first period actions. The analysis indicates a rationale for why agents would wait to react to the actions of other agents, rather than to act first and set the group's agenda. The general point is that, irrespective of the type of the agent, the social influence motive makes the value of flexibility that comes from speaking second to be higher than the commitment value of speaking first and setting the agenda. A player who speaks second observes the first player's action and has the opportunity, if needed, to compensate for it. This works to the second player's advantage irrespective of whether or not the first player's action was close to or not from her preference. If the first player chose an action very different from the second player's preference, then she can compensate by picking a more extreme action in the opposite direction. But if the first player were to choose an action which is already close to her own preference, then the second player can also choose an action close to her preference and thereby not incur the cost of exaggerating.

Because the second player can can adjust her action based on the observed actions of the first player, the first player in anticipation of this behavior has the incentive to be more extreme, which in turn reduces her utility even more. Thus when agents care about influencing the overall group outcome towards their true preference, they prefer if given the choice to wait and delay their actions. The benefit of speaking second is higher for players who are more extreme – moderates have less to lose from speaking first. In general, moderates suffer less from decisions which are away from the middle. However, a player whose preference is more extreme on the right suffers a lot if the final outcome is more extreme to the left (and vice-versa). In sum, the analysis suggests that there are inherent advantages to waiting, especially for agents with more extreme preferences.

C Endogenous Sequential Actions: Bidding to Speak

As described above the trade-off faced between truth-telling and group influence leads to a preference among agents to wait and speak in the second period and further such a strategy is more beneficial for players with extreme preferences. The natural question is what would happen in a group where the speaking order is endogenous. Given that speaking second is the dominant choice, there would exist a market for the order of speaking which may be characterized by allowing agents to endogenously bid for the right to determine the speaking order. In reality such an endogenous choice game implies the idea that group members may be willing to take costly actions to determine whether they are able to speak in the most favorable position.

Consider then an extension to the game where in a prior period 1 both agents participate in a first-price sealed bid auction for the opportunity to decide the speaking order.

A neutral organizer/auctioneer receives agents' bids and announces the winner: if $b_i > b_j$, then i is the winner and in the event of a tie the winner is randomly chosen. The organizer announces the winner (but not the bid amounts) and so each player's beliefs will be based on inferences about the other's type depending upon who won the auction. In period 2, the winner chooses the preferred speaking order. In period 3 the players act based on the speaking-order determined by the winner. Players' have the same utility as in equation (1), except now the winner of the auction (say i) also pays her bid b_i to the organizer in period 1 for the right to choose the speaking order.

We derive the symmetric equilibrium bidding strategies of this game where the equilibrium bidding functions $\beta(x)$ are symmetric around zero. Unlike in the standard auction models where the bidder valuations are exogenously specified, the challenge in deriving the equilibrium strategies in this model stems from the fact that a bidder's valuation for the speaking order is endogenous to the outcome of the auction itself.

Proposition 6 *In the game where the agents participate in a first-price sealed-bid auction to decide on the right to determine the speaking order, there exists a unique symmetric PBE in which, agent i*

- *has a bidding strategy* $\beta(x_i) = f(r)\frac{\int_0^{x_i} x^2 g(x)\mathrm{d}x}{\int_0^{x_i} g(x)\mathrm{d}x}$, *where* $f(r) > 0 \ \forall \ r > 0$ *and chooses to speak second if she wins the auction.*
- *If agent i speaks first, then she chooses* $\hat{a}_{\mathrm{n}1,i} = \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i$, *whereas if she speaks second, she chooses* $\hat{a}_{\mathrm{n}2,i} = \frac{2(1+r)}{1+3r}x_i - \frac{(1-r)}{1+3r}a_{\mathrm{n}1,j}$.

Proof: See Appendix □

Note that the equilibrium actions of the agents in this endogenous game end up being the same as that in the exogenous sequential game. Clearly, the agent who moves second faces the same game as the agent in the exogenous sequential game because she always chooses her response $a_{\mathrm{n}2,i}$ in response to the first player's action $a_{\mathrm{n}1,j}$. The incentives facing the first player is more subtle. If she is speaking first; this could either be because she won the auction and chose to go first, or because she lost and was asked to go first by the other player (the former case turns out to be off the equilibrium path). Regardless, i does not know j's type because j has not yet spoken. Therefore, i's actions will depend on her beliefs about j's type which will be based on the observed outcome of the auction and j's choice of speaking order (if j had the opportunity to decide it).

In a symmetric PBE, the region, say, W to which i can expect j to belong to is symmetric around zero, irrespective of the exact scenario under which i is speaking first. Hence, i's expected utility from speaking first is obtained by taking the expectation of $u(x_i, a_{\mathrm{n1},i}, \hat{a}_{\mathrm{n2},j})$ over $x_j \in W$, *i.e.*, $EU_{\mathrm{n1}}(x_i, a_{\mathrm{n1},i}) = \frac{\int_W u(x_i, a_{\mathrm{n1},i}, \hat{a}_{\mathrm{n2},j}) g(x_j) \mathrm{d}x_j}{\int_W g(x_j)\mathrm{d}x_j}$. This can be simplified by substituting for $\hat{a}_{\mathrm{n2},i}$:

$$\begin{aligned} EU_{\mathrm{n1}}(x_i, a_{\mathrm{n1},i}) = & - \quad r(x_i - a_{\mathrm{n1},i})^2 - (1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{\mathrm{n1},i}\right)^2 + \left(\frac{1+r}{1+3r}\right)^2 \frac{\int_W x_j^2 g(x_j)\mathrm{d}x_j}{\int_W g(x_j)\mathrm{d}x_j}\right] \\ & - \quad 2\frac{1+r}{1+3r}\left(x_i - \frac{2r}{1+3r}a_{\mathrm{n1},i}\right)\frac{\int_W x_j g(x_j)\mathrm{d}x_j}{\int_W g(x_j)\mathrm{d}x_j} \end{aligned} \tag{10}$$

The last term vanishes because W is symmetric around zero. By setting $\left.\frac{\mathrm{d}EU_{\mathrm{n1}}(x_i, a_{\mathrm{n1},i})}{\mathrm{d}a_{\mathrm{n1},i}}\right|_{a_{\mathrm{n1},i}=\hat{a}_{\mathrm{n1},i}} = 0$, we can solve for $\hat{a}_{\mathrm{n1},i} = \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i$ which turns out to be the same as in the exogenous sequential case.

Players' equilibrium beliefs are that upon losing, they will forfeit the right to decide the speaking order, and the right to move second. Given this, the equilibrium bidding strategy can be specified. In deriving the equilibrium bidding strategy, note that a players' value from winning the auction is endogenous, unlike a traditional first-price sealed-bid auction, where players' valuations are exogenously given. The approach to deriving the equilibrium is to show that equilibrium bidding strategies are increasing strictly monotonically in $|x_i|$. The equilibrium bidding strategy is derived in the Appendix to be $\beta(x_i) = f(r)\frac{\int_0^{x_i} x^2 g(x)\mathrm{d}x}{\int_0^{x_i} g(x)\mathrm{d}x}$ and it is monotonically increasing in $|x_i|$. Thus agents with more extreme preferences have higher value for choosing the speaking order and will accordingly bid higher. The multiplier, $f(r)$, of the equilibrium bidding function is monotonically decreasing in r, i.e., as players' need to pull the final outcome ($\bar{a}_{\mathrm{n}}$) closer to own preference increases (as r decreases), their bid increases. At $r = 0$, the bidding strategy simplifies to $\beta(x_i)|_{r=0} = 2\frac{\int_0^{x_i} x^2 g(x)\mathrm{d}x}{\int_0^{x_i} g(x)\mathrm{d}x}$, which is the highest, whereas at $r = 1$, the bidding strategy devolves to $\beta(x_i)|_{r=1} = 0$.

Consider now the mean equilibrium outcome of the endogenous sequential actions game. For a given x_i and x_j suppose $|x_i| < |x_j|$, without loss of generality. Then j

wins the auction and chooses to speak second, and the mean equilibrium outcome will be $\bar{a}_{\mathrm{n}} = \frac{\hat{a}_{\mathrm{n}1,i}+\hat{a}_{\mathrm{n}2,j}}{2} = \frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i + \frac{1+r}{1+3r}x_j \ \ \forall \ |x_i| < |x_j|$. Proposition 7 compares the mean outcome, $\bar{a}_{\mathrm{n}}$, with the mean of the preferences $\bar{x}$ and the mean outcome in the simultaneous choice game $\bar{a}$.

Proposition 7 *In the equilibrium of the endogenous sequential actions game polarization always occurs ($|\bar{a}_{\mathrm{n}}| > |\bar{x}|$ and $\bar{a}_{\mathrm{n}}\bar{x} > 0$). Comparison of the extent of polarization across the different games yields:*

- *If $x_i \cdot x_j < 0$, then $|\bar{a}_{\mathrm{n}}| > |\bar{a}|$.*
- *If $x_i \cdot x_j > 0$, then $|\bar{a}_{\mathrm{n}}| \leq |\bar{a}|$*

Allowing the agents to compete for the right to speak always leads to the polarization of actions. When the speaking order is endogenous, the agent who wins the right to speak always prefers to wait. Further, it is the agents with more extreme preferences who have the incentive to bid more for the right to determine the speaking order. This leads to the important point that if agents were to bid for the right to speak then the actions of the agents and the group outcome is always polarized. *Unlike in the exogenous sequential game where moderation is a possibility, allowing agents to choose the speaking order always leads to polarization.*

It is also useful to compare the outcome of the endogenous sequential actions game with that of the simultaneous game. If the two players lie on opposite sides of zero, then endogenous sequential actions produces more polarization than the simultaneous actions game. On the other hand, if both players lie on the same side of zero, then the outcome is less polarized than that in the simultaneous game. The basic mechanism at play is that the second player in the sequential actions game can condition her action to that of the first player and accordingly pull the group outcome closer to her own preference. Recall that the first player's action in the sequential case is always more extreme than in the simultaneous case because of a compensation effect: i.e., she knows that the second player can observe and compensate for her action. In the endogenous sequential actions game, it is the more extreme player who ends up winning the right to be the second player. Given this, the player who loses the auction and speaks first can infer that the other player has more extreme preferences. This inference induces her to be even more extreme. When the two players' preferences are on opposite sides of zero then not only does the second player have the incentive to be more extreme (after observing the first player's actions) in order to pull the joint outcome towards her preference, but the inference effect also induces the first player to be more extreme. Consequently, the group becomes more polarized than in the simultaneous actions game. In contrast, when the players' preferences are on the same side of zero, the second player's knowledge of first's actions implies that she does not need to shade and take too extreme actions. The implication is that when the players are similarly inclined, endogenizing the speaking order can help to reduce polarization.

D Welfare Comparisons

We start with the social planner's problem to understand how a benevolent principal would design the group interaction. The welfare in the two player system for any x_i and x_j is given by equation (2). Note that any pecuniary transfers (such as bids) are canceled out since they remain within the system and hence have no impact on the total welfare. In discussion forums, the speaking formats (or the timing game-forms) are design decisions and are often chosen before the agents' preferences are drawn. Therefore we can consider the expected welfare for a given game-form across the distribution of player types as a relevant measure for making welfare comparisons, i.e., $EW = \frac{\int_{x_i}\int_{x_j} W(x_i,x_j)g(x_i)g(x_j)\mathrm{d}x_i\mathrm{d}x_j}{\int_{x_i}\int_{x_j} g(x_i)g(x_j)\mathrm{d}x_i\mathrm{d}x_j}$. Denote the expected welfare for the first-best case to be EW_{FB}, the simultaneous case by EW_s, the exogenous sequential by EW_x and the endogenous sequential by EW_n.

Figure 4 shows the relationship between expected welfare functions as a function of r for the case of $U[-1,1]$. It can be seen that $EW_n > EW_x > EW_s$. Within the sequential action formats allowing agents to endogenously bid for the speaking order increases expected welfare as compared to the exogenous assignment of speaking order across the agents. The market clearing for the speaking order through the first-price auction mechanism improves efficiency. Further, it can be seen that the expected welfare under the sequential game (irrespective of whether it is exogenous or endogenous) is higher than that for the simultaneous game. It is interesting that even though the endogenous sequential

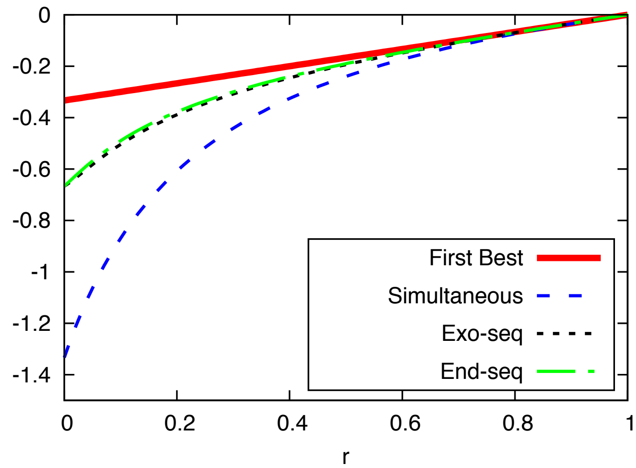

Figure 4: Expected social welfare for the four game forms. Shown for an uniform distribution from -1 to 1.

choice game produces higher polarization, it also increases the welfare by allowing those with more extreme preferences to obtain the outcomes they desire.

V Conclusion

There are many formal and informal forums in society that facilitate group interactions between citizens and shape their views on important and often contentious issues like gun control, the size of government, climate change, and abortion rights. One might expect such deliberations to help in the exchange of information and better coordination. However, one only need to look the current socio-political landscape in the U.S. to observe that deliberations on issues such as gun control, if anything, lead to more polarization. Understanding the mechanisms that make opinions more polarized is important, because they can create conflict and impede effective policy making.

We develop a theory that links the polarization of actions to a basic trade-off faced by agents between influencing others in a deliberation and representing their true preferences. When agents take actions with the objective to have group influence, deliberations can lead to polarization. Further, it is the more extreme agents who end up becoming more polarized in their actions. We analyze the role of simultaneous versus sequential timing of the actions of agents. With sequential actions polarization occurs whenever the agent who moves later is more extreme compared to the first agent. With sequential actions we also highlight the tradeoff between the commitment value of moving first versus the value of flexibility to optimally adjust to the first period actions that comes from moving later. The group influence incentive makes flexibility valuable and induces agents to wait.

The comparison of the different timings shows that if the preferences of the agents are dissimilar, then the sequential actions game produces less polarization as compared to simultaneous actions, whereas if the agents have similar preferences it is the simultaneous actions game that leads to less polarization. When the timing of sequential actions is endogenous and agents bid for the right to choose the order of actions, agents with more extreme preferences bid more, and the winning agents regardless of their preferences prefer to wait. We also examine the effect of the interaction group size and show that larger groups will show less polarization. Further, when the interactions are between sub-groups each with agents who have similar preferences, smaller sub-groups have the incentive to become even more extreme.

In many contexts there might be principals who are be interested in influencing the extent of polarization to satisfy their private objectives. Such a principal could be a firm interested in managing consumer opinions on social media or convenors of government/public institutions (such as juries, regulatory and legislative bodies) who are interested in the functioning of these institutions. Our paper considered one general mechanism which affects the extent of polarization, namely the timing of actions of the agents. But the role of incentives designed by interested principals to manage the degree of group polarization seems to be a rich area for analysis.

One might also consider connecting our setup to research that has examined why senders exaggerate messages compared to their private information in cheap talk games.[9]

[9] For example, Kartik et al. (2007) show that senders might use inflated communication in the presence of credulous receivers. Chakraborty and Harbaugh (2007) show the possibility for exaggeration of messages when a sender has multidimensional comparative information. There are also models with multiple senders; e.g., Krishna and Morgan (2001), Chen (2011), and Rausser et al. (2015) in which senders exaggerate in order to counter the actions of other senders.

It might be interesting to understand whether there is a role for costless pre-play communication to affect the polarization of actions within the model presented in this paper.

References

D. Acemoglu, V. Chernozhukov, M. Yildiz, et al. Fragility of Asymptotic Agreement Under Bayesian Learning. Unpublished paper, Department of Economics, Massachusetts Institute of Technology, 2009.

J. Aguilar. At Shooting Range, Abbott Signs "Open Carry" Bill. The Texas Tribune, 2016. Online, posted 13-June-2015, retrieved 27-March-2016.

A. Alesina. Credibility and Policy Convergence in a Two-party System with Rational Voters. *The American Economic Review*, pages 796–805, 1988.

J. Andreoni and T. Mylovanov. Diverging Opinions. *American Economic Journal: Microeconomics*, 4(1):209–232, 2012.

Armed Campuses, 2016. URL `http://www.armedcampuses.org/`. Online, acessed 27-March-2016.

S. Baliga, E. Hanany, and P. Klibanoff. Polarization and Ambiguity. *The American Economic Review*, 103(7):3071–3083, 2013.

J.-P. Benoit and J. Dubra. A Theory of Rational Attitude Polarization. Working Paper, 2016.

Campus Carry Policy. Campus Carry Policy Working Group – Final Report. Final Report, University of Texas at Austin, 2015. Online, retrieved 27-March-2016.

A. Chakraborty and R. Harbaugh. Comparative Cheap Talk. *Journal of Economic Theory*, 132(1): 70–94, 2007.

Y. Chen. Perturbed Communication Games with Honest Senders and Naive Receivers. *Journal of Economic Theory*, 146(2):401–424, 2011.

N. Cohn. Polarization Is Dividing American Society, Not Just Politics. New York Times, 2014. Online, posted 12-June-2014, retrieved 27-March-2016.

J. M. Darley and P. H. Gross. A Hypothesis-Confirming Bias in Labeling Effects. *Journal of Personality and Social Psychology*, 44(1):20, 1983.

A. K. Dixit and J. W. Weibull. Political Polarization. *Proceedings of the National Academy of Sciences*, 104(18):7351–7356, 2007.

J.-Y. Duclos, J. Esteban, and D. Ray. Polarization: Concepts, Measurement, Estimation. *Econometrica*, 72(6):1737–1772, 2004.

J.-M. Esteban and D. Ray. On the Measurement of Polarization. *Econometrica*, pages 819–851, 1994.

J. D. Fearon and D. D. Laitin. Ethnicity, Insurgency, and Civil War. *American Political Science Review*, 97(01):75–90, 2003.

D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, Cambridge, MA, 1991.

Gallup. Tea Party Movement, 2015. URL `http://www.gallup.com/poll/147635/tea-party-movement.aspx`. Online, retrieved 27-March-2016.

M. Gentzkow and J. M. Shapiro. What Drives Media Slant? Evidence from US Daily Newspapers. *Econometrica*, 78(1):35–71, 2010.

E. L. Glaeser, G. A. Ponzetto, and J. M. Shapiro. Strategic Extremism: Why Republicans and Democrats Divide on Religious Values. *The Quarterly journal of economics*, 120(4):1283–1330, 2005.

S. Huitlin. Guns on Campus: An Overview, 2015. URL `http://www.ncsl.org/research/education/guns-on-campus-overview.aspx`. Online, posted 5-October-2015, retrieved 27-March-2016.

D. J. Isenberg. Group Polarization: A Critical Review and Meta-Analysis. *Journal of Personality and Social Psychology*, 50(6):1141, 1986.

Y. Kamada and F. Kojima. Voter Preferences, Polarization, and Electoral Policies. *American Economic Journal: Microeconomics*, 6(4):203–236, 2014.

Kansas State Education Board. Kansas Science Education Standards. Web Archive, 2006. Online, archived on 6-February-2007, retrieved on 27-March-2016.

N. Kartik, M. Ottaviani, and F. Squintani. Credulity, Lies, and Costly Talk. *Journal of Economic theory*, 134(1):93–116, 2007.

M. Kaufman. A Deeper Divide: The Gun Control Debate After Newtown. Hartford Courant, 2013. Online, posted 12-September-2016, retrieved 8-Jan-2017.

D. R. Kinder and W. Mebane. The Political Process and Economic Change, 1983.

P. Kondor. The More We Know About The Fundamental, The Less We Agree On The Price. *The Review of Economic Studies*, 79(3):1175–1207, 2012.

V. Krishna and J. Morgan. A Model of Expertise. *The Quarterly Journal of Economics*, 116(2): 747–775, 2001.

G. Lopez-Moctezuma. Sequential Deliberation in Collective Decision-Making: The Case of the FOMC. Working Paper, 2016.

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

J. G. Montalvo and M. Reynal-Querol. Ethnic Polarization, Potential Conflict, and Civil Wars.

American Economic Review, (95 (3)), 2005.

S. Mullainathan and A. Shleifer. The Market for News. *The American Economic Review*, 95(4): 1031–1053, 2005.

S. Plous. Biases in the Assimilation of Technological Breakdowns: Do Accidents Make Us Safer? *Journal of Applied Social Psychology*, 21(13):1058–1082, 1991.

M. Rabin and J. L. Schrag. First Impressions Matter: A Model of Confirmatory Bias. *Quarterly journal of Economics*, pages 37–82, 1999.

G. C. Rausser, L. K. Simon, and J. Zhao. Rational Exaggeration and Counter-Exaggeration in Information Aggregation Games. *Economic Theory*, 59(1):109–146, 2015.

J. Rawls. A Theory of Justice. *Bamforth N, Malik M and OCinneide C*, 1971.

J. E. Roemer. The Strategic Role of Party Ideology When Voters Are Uncertain about How the Economy Works. *American Political Science Review*, 88(02):327–335, 1994.

P. Slevin. Kansas Education Board First to Back "Intelligent Design". The Washington Post, 2005. Online, posted 9-November-2005, retrieved 27-March-2016.

C. R. Sunstein. The Law of Group Polarization. *Journal of Political Philosophy*, 10(2):175–195, 2002.

C. Todd, M. Murray, and C. Dann. Theres The Tea Party – And Then Theres Everyone Else. NBCNews, 2014. URL `http://www.nbcnews.com/politics/first-read/tea-party-stands-alone-key-issues-n135426`. Online, published = 19-June-2014, retrived = 27-March-2016.

VI APPENDIX

Equilibrium of the m-Player Game

To derive the Bayesian Nash equilibrium for a m-player simultaneous game, we can write the utility of a player i as:

$$u(x_i, a_{\mathrm{m},i}, a_{\mathrm{m},-i}) = -r(x_i - a_{\mathrm{m},i})^2 - (1-r)\left(x_i - \frac{a_{\mathrm{m},i}}{m} - \frac{\sum_{j=1}^{m-1} a_{\mathrm{m},j}}{m}\right)^2 \tag{A.1}$$

To obtain i's expected utility for any $a_{\mathrm{m},i}$ in equilibrium, we can take the expectation of equation (A.1) over all the other $m-1$ agents' equilibrium actions:

$$EU_{\mathrm{m}}\left(x_i, a_{\mathrm{m},i}\right) = \frac{\int_{\mathbb{R}} \dots \int_{\mathbb{R}} \int_{\mathbb{R}} \left(u(x_i, a_{\mathrm{m},i}, \hat{a}_{\mathrm{m},-i})\right) g(x_1)g(x_2)\dots g(x_{m-1})\mathrm{d}x_1\mathrm{d}x_2\dots\mathrm{d}x_{m-1}}{\int_{\mathbb{R}} \dots \int_{\mathbb{R}} \int_{\mathbb{R}} g(x_1)g(x_2)\dots g(x_{m-1})\mathrm{d}x_1\mathrm{d}x_2\dots\mathrm{d}x_{m-1}} \tag{A.2}$$

Substituting for $u(x_i, a_{\mathrm{m},i}, a_{\mathrm{m},-i})$ from equation (A.1) and simplifying, we have:

$$\begin{aligned} EU_{\mathrm{m}}\left(x_i, a_{\mathrm{m},i}\right) &= -r(x_i - a_{\mathrm{m},i})^2 \\ + \quad & \frac{2(1-r)}{m}\left(x_i - \frac{a_{\mathrm{m},i}}{m}\right) \frac{\int_{\mathbb{R}} \dots \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\sum_{j=1}^{m-1} a_{\mathrm{m},j}\right) g(x_1)g(x_2)\dots g(x_{m-1})\mathrm{d}x_1\mathrm{d}x_2\dots\mathrm{d}x_{m-1}}{\int_{\mathbb{R}} \dots \int_{\mathbb{R}} \int_{\mathbb{R}} g(x_1)g(x_2)\dots g(x_{m-1})\mathrm{d}x_1\mathrm{d}x_2\dots\mathrm{d}x_{m-1}} \\ - \quad & (1-r)\left[\left(x_i - \frac{a_{\mathrm{m},i}}{m}\right)^2 + \frac{\int_{\mathbb{R}} \dots \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\sum_{j=1}^{m-1} a_{\mathrm{m},j}\right)^2 g(x_1)g(x_2)\dots g(x_{m-1})\mathrm{d}x_1\mathrm{d}x_2\dots\mathrm{d}x_{m-1}}{m^2 \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \int_{\mathbb{R}} g(x_1)g(x_2)\dots g(x_{m-1})\mathrm{d}x_1\mathrm{d}x_2\dots\mathrm{d}x_{m-1}}\right] \end{aligned} \tag{A.3}$$

We know that $\int_{\mathbb{R}} g(x)dx = 1$. Also, from i's perspective, all the other agents' types are i.i.d from the distribution $g(x)$. So we can simplify equation (A.3) to:

$$\begin{aligned} EU_{\mathrm{m}}\left(x_i, a_{\mathrm{m},i}\right) = \quad - \quad & r(x_i - a_{\mathrm{m},i})^2 + 2(1-r)\frac{m-1}{m}\left(x_i - \frac{a_{\mathrm{m},i}}{m}\right)\int_{\mathbb{R}} a_{\mathrm{m},j} g(x_j)\mathrm{d}x_j \\ - \quad & (1-r)\left[\left(x_i - \frac{a_{\mathrm{m},i}}{m}\right)^2 + \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\sum_{j=1}^{m-1} a_{\mathrm{m},j}\right)^2 g(x_1)g(x_2)\dots g(x_{m-1})\mathrm{d}x_1\mathrm{d}x_2\dots\mathrm{d}x_{m-1}\right] \end{aligned} \tag{A.4}$$

Taking the F.O.C of Equation (A.4) w.r.t $a_{\mathrm{m},i}$ at $\hat{a}_{\mathrm{m},i}$ gives us:

$$\left.\frac{\mathrm{d}EU_{\mathrm{m}}\left(x_i, a_{\mathrm{m},i}\right)}{\mathrm{d}a_{\mathrm{m},i}}\right|_{a_{\mathrm{m},i}=\hat{a}_{\mathrm{m},i}} = 2r(x_i - \hat{a}_{\mathrm{m},i}) - \frac{(1-r)(m-1)}{m^2}\int_{\mathbb{R}} \hat{a}_{\mathrm{m},j} g(x_j)\mathrm{d}x_j + \frac{1-r}{m}\left(x_i - \frac{\hat{a}_{\mathrm{m},i}}{m}\right) = 0 \tag{A.5}$$

This gives us $\hat{a}_{\mathrm{m},i}$ as:

$$\hat{a}_{\mathrm{m},i} = \frac{\left[rm^2 + (1-r)m\right]x_i - (1-r)(m-1)\int_{\mathbb{R}} \hat{a}_{\mathrm{m},j} g(x_j)\mathrm{d}x_j}{rm^2 + (1-r)} \tag{A.6}$$

Integrating $\hat{a}_{\mathrm{m},i}$ w.r.t x_i, we have:

$$\int_{\mathbb{R}} \hat{a}_{\mathrm{m},i} g(x_i)\mathrm{d}x_i = \frac{\left[rm^2 + (1-r)m\right]\int_{\mathbb{R}} x_i g(x_i)\mathrm{d}x_i - (1-r)(m-1)\int_{\mathbb{R}} \hat{a}_{\mathrm{m},j} g(x_j)\mathrm{d}x_j \int_{\mathbb{R}} g(x_i)\mathrm{d}x_i}{rm^2 + (1-r)} \tag{A.7}$$

Since $\int_{\mathbb{R}} g(x_i)\mathrm{d}x_i = 1$ and $\int_{\mathbb{R}} \hat{a}_{\mathrm{m},i} g(x_i)\mathrm{d}x_i = \int_{\mathbb{R}} \hat{a}_{\mathrm{m},j} g(x_j)\mathrm{d}x_j = E(a_m)$ this simplifies to $E(a_m) = E(x)$. Since $E(x) = 0$ in a symmetric distribution, we have:

$$\hat{a}_{\mathrm{m},i} = \frac{rm^2 + (1-r)m}{rm^2 + (1-r)} x_i \tag{A.8}$$

We now derive the comparative statics regarding the extent of shading $s_{\mathrm{m},i}$.
i) $\frac{\mathrm{d}s_{\mathrm{m},i}}{\mathrm{d}r} = (m-1)\frac{-(1-r)(m^2-1)-[rm^2+(1-r)]}{[rm^2+(1-r)]^2}|x_i|$. This simplifies to $\frac{\mathrm{d}s_{\mathrm{m},i}}{\mathrm{d}r} = -\frac{(m-1)m^2}{[rm^2+(1-r)]^2}|x_i|$. We know that $m, m-1 > 0$, because by definition $m \geq 2$. Also, it is clear that $[rm^2 + (1-r)]^2 > 0$ and that $|x_i| \geq 0$. It therefore follows that $\frac{\mathrm{d}s_{\mathrm{m},i}}{\mathrm{d}r} \geq 0$.
ii) $\frac{\mathrm{d}s_{\mathrm{m},i}}{\mathrm{d}|x_i|} = \frac{(1-r)(m-1)}{rm^2+(1-r)}$. Since $0 < r < 1$ and $m \geq 2$, this value is always positive. So $\frac{\mathrm{d}s_{\mathrm{m},i}}{\mathrm{d}|x_i|} > 0$.
iii) $\frac{\mathrm{d}s_{\mathrm{m},i}}{\mathrm{d}m} = (1-r)\frac{1-r(m-1)^2}{[rm^2+(1-r)]^2}$. If $m < \frac{1}{\sqrt{r}} + 1$, then $1 - r(m-1)^2 > 0$; so in this range $\frac{\mathrm{d}s_{\mathrm{m},i}}{\mathrm{d}m} > 0$. Else if $m > \frac{1}{\sqrt{r}} + 1$, then $1 - r(m-1)^2 < 0$; so in this range $\frac{\mathrm{d}s_{\mathrm{m},i}}{\mathrm{d}m} < 0$.

Equilibrium for Sub-group Interactions

The expected utility of agent i from sub-group 1 and agent j in sub-group 2 are given in Equations (6). The first-order condition for agent i is $\frac{\mathrm{d}EU_1^i}{\mathrm{d}a_1^i} = 0$ which can be calculated to be:

$$r(x_1 - a_1^i) + \frac{1-r}{m}\left(x_1 - \frac{a_1^i + \sum_{k_1=1(\neq i)}^{n_1} a_1^{k_1} + \sum_{k_2=1}^{n_2}\int_{\mathbb{R}} a_2^{k_2} dx_2}{m}\right) = 0 \tag{A.9}$$

As we are looking for a symmetric in actions with sub-group equilibrium, we can set $a_1^i = a_1^k = \hat{a}_1$, and $a_2^k = \hat{a}_2$, and simplify (A.9) to:

$$x_1(r + \frac{1-r}{m}) - \hat{a}_1(r + \frac{n_1(1-r)}{m^2}) - \frac{(1-r)n_2}{m^2}\int_{\mathbb{R}} \hat{a}_2 g(x_2) dx_2 = 0 \tag{A.10}$$

Thus the equilibrium action $\hat{a}_1$ for agents from sub-group 1 is:

$$\hat{a}_1 = x_1\left(\frac{m(mr + (1-r))}{m^2 r + n_1(1-r)}\right) - \frac{(1-r)n_2}{m^2 r + n_1(1-r)}\int_{\mathbb{R}} \hat{a}_2 g(x_2) dx_2 \tag{A.11}$$

Integrating the equilibrium action over the entire distribution of x_1 we can get:

$$\int_{\mathbb{R}} \hat{a}_1 g(x_1) dx_1 = E(x_1)\left(\frac{m(mr + (1-r))}{m^2 r + n_1(1-r)}\right) - \frac{(1-r)n_2}{m^2 r + n_1(1-r)}\int_{\mathbb{R}} \hat{a}_2 g(x_2) dx_2 \tag{A.12}$$

By noting that $E(x_1) = 0$ for a symmetric distribution, we have that $\int_{\mathbb{R}} \hat{a}_1 g(x_1) dx_1 = \int_{\mathbb{R}} \hat{a}_2 g(x_2) dx_2$. Therefore the equilibrium $\hat{a}_1$ is $\hat{a}_1 = x_1\left(\frac{m(mr+(1-r))}{m^2 r + n_1(1-r)}\right)$. Using similar analysis for sub-group 2 we can also derive $\hat{a}_2 = x_2\left(\frac{m(mr+(1-r))}{m^2 r + n_2(1-r)}\right)$.

Equilibrium of the Exogenous Sequential Game

Let i and j be the first and second players, respectively. The second player j's equilibrium action has already been derived in the main text (see section B.1), and is equal to:

$$\hat{a}_{\mathrm{x2},j} = \frac{2(1+r)}{1+3r} x_j - \frac{(1-r)}{1+3r} a_{\mathrm{x1},i} \tag{A.13}$$

So we now derive the first player's optimal action $\hat{a}_{\mathrm{x}1,i}$. In equilibrium, i's utility from choosing action $a_{\mathrm{x}1,i}$ is given by:

$$u\left(x_i, a_{\mathrm{x}1,i}, \hat{a}_{\mathrm{x}2,j}\right) = -r(x_i - a_{\mathrm{x}1,i})^2 - (1-r)\left(x_i - \bar{a}_{\mathrm{x}}\right)^2 \tag{A.14}$$

where $\bar{a}_{\mathrm{x}} = \frac{a_{\mathrm{x}1,i}+\hat{a}_{\mathrm{x}2,j}}{2}$. We know that when i chooses $a_{\mathrm{x}1,i}$, in response, j chooses $\hat{a}_{\mathrm{x}2,j} = \frac{2(1+r)}{1+3r}x_j - \frac{(1-r)}{1+3r}a_{\mathrm{x}1,i}$. This in turn gives us $\bar{a}_{\mathrm{x}}$ as:

$$\bar{a}_{\mathrm{x}} = \frac{(1+r)x_j + 2ra_{\mathrm{x}1,i}}{1+3r} \tag{A.15}$$

Substituting this value of $\bar{a}_{\mathrm{x}}$ into Equation (A.14), we have:

$$\begin{aligned} u\left(x_i, a_{\mathrm{x}1,i}, \hat{a}_{\mathrm{x}2,j}\right) = -r(x_i - a_{\mathrm{x}1,i})^2 \quad &- \quad (1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{\mathrm{x}1,i}\right)^2 + \left(\frac{1+r}{1+3r}x_j\right)^2\right] \\ &+ \quad 2(1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{\mathrm{x}1,i}\right)\left(\frac{1+r}{1+3r}x_j\right)\right] \end{aligned} \tag{A.16}$$

Hence, the expected utility of agent i in from choosing action $a_{\mathrm{x}1,i}$ in period 1 is:

$$EU_{\mathrm{x}1}\left(x_i, a_{\mathrm{x}1,i}\right) = \frac{\int_{\mathbb{R}} u\left(x_i, a_{\mathrm{x}1,i}, \hat{a}_{\mathrm{x}2,j}\right) g(x_j)\mathrm{d}x_j}{\int_{\mathbb{R}} g(x_j)\mathrm{d}x_j} \tag{A.17}$$

We can simplify the above as follows:

$$\begin{aligned} EU_{\mathrm{x}1}\left(x_i, a_{\mathrm{x}1,i}\right) = \quad &- \quad r(x_i - a_{\mathrm{x}1,i})^2 + 2(1-r)\left(x_i - \frac{2r}{1+3r}a_{\mathrm{x}1,i}\right)\left(\frac{1+r}{1+3r}\right)\int_{\mathbb{R}} x_j g(x_j)\mathrm{d}x_j \\ &- \quad (1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{\mathrm{x}1,i}\right)^2 + \left(\frac{1+r}{1+3r}\right)^2 \int_{\mathbb{R}} x_j^2 g(x_j)\mathrm{d}x_j\right] \end{aligned} \tag{A.18}$$

Computing the integrals in Equation (A.18), we have:

$$EU_{\mathrm{x}1}\left(x_i, a_{\mathrm{x}1,i}\right) = -r\left(x_i - a_{\mathrm{x}1,i}\right)^2 - (1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{\mathrm{x}1,i}\right)^2 + \left(\frac{1+r}{1+3r}\right)^2 \int_{\mathbb{R}} x_j^2 g(x_j)\mathrm{d}x_j\right] \tag{A.19}$$

Solving the first-order condition for agent i's action gives us the equilibrium choice $\hat{a}_{\mathrm{x}1,i}$:

$$\hat{a}_{\mathrm{x}1,i} = \frac{(1+3r)(3+r)}{(1+3r)^2 + 4r(1-r)}x_i \tag{A.20}$$

Thus, given x_i and x_j, there exists a unique exogenous sequential choice equilibrium, where $\hat{a}_{\mathrm{x}1,i} = \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i$ and $\hat{a}_{\mathrm{x}2,j} = \frac{2(1+r)x_j-(1-r)a_{\mathrm{x}1,i}}{1+3r}$.

Proof of Proposition 2

Recall that the mean of the equilibrium outcome for the exogenous sequential game can be expressed as:

$$\bar{a}_{\mathrm{x}} = \frac{\hat{a}_{\mathrm{x}1,i} + \hat{a}_{\mathrm{x}2,j}}{2} = \frac{2ra_{\mathrm{x}1,i} + (1+r)x_j}{1+3r} = \frac{2r(3+r)}{(1+3r)^2 + 4r(1-r)}x_i + \frac{1+r}{1+3r}x_j \tag{A.21}$$

Next we show that $0 < k_1(r), k_2(r), k_3(r) < 1$.

- $k_1(r) < 1$ if $\frac{(1+3r)(1-r)}{(1+3r)^2+4r(1-r)} < 1 \Rightarrow (1+3r)(1-r) < (1+3r)^2+4r(1-r) \Rightarrow -4r(1+3r) < 4r(1-r)$, which is always true if $0 < r < 1$. Hence $k_1(r) < 1$.
- $k_2(r) < 1$ if $\frac{(1+3r)\left[(1+3r)^2+16r\right]}{[(1+3r)^2+4r(1-r)][2(1+r)+(1+3r)]} < 1, \Rightarrow 4r(1+3r)(3+r) < 2(1+3r)^2(1+r)+4r(1-r)(1+r) \Rightarrow -(1+3r)(1-r)^2 < 4r(1-r)(1+r)$, which is always true since $0 < r < 1$.
- $k_3(r) < 1$ if $\frac{(1+3r)^3+8r(1-r)(1+r)}{2(1+r)[(1+3r)^2+4r(1-r)]} < 1 \Rightarrow (1+3r)^3 < 2(1+r)(1+3r)^2, \Rightarrow r < 1$, which we know is always true.

Further, $k_1(r), k_2(r), k_3(r) > 0$ since all the terms in their numerators and denominators are positive. Thus $0 < k_1(r), k_2(r), k_3(r) < 1$. Next, we present the proofs of Proposition 3a and 3b.

a) First, we compare $\bar{a}_{\mathrm{x}}$ with $\bar{x}$. Without loss of generality, let $x_i \geq 0$.

- Polarization – $|\bar{a}_{\mathrm{x}}| > |\bar{x}|$ and $\bar{a}_{\mathrm{x}}\bar{x} > 0$.
  - First, consider the case where $x_j > k_1(r)x_i$. In this case, $\bar{a}_{\mathrm{x}}\bar{x} > 0$ because both $\bar{a}_{\mathrm{x}} > 0$ and $\bar{x} > 0$. The condition $|\bar{a}_{\mathrm{x}}| > |\bar{x}|$ therefore simplifies to $\bar{a}_{\mathrm{x}} > \bar{x}$. This can be expressed as:

$$\frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i + \frac{1+r}{1+3r}x_j > \frac{1+r}{1+3r}x_i + \frac{1+r}{1+3r}x_j$$

$$\Rightarrow \quad x_j > \frac{(1+3r)(1-r)}{(1+3r)^2+4r(1-r)}x_i \tag{A.22}$$

  where the multiplier of x_i is $k_1(r) = \frac{(1+3r)(1-r)}{(1+3r)^2+4r(1-r)}$. By definition,(A.22) is satisfied.
  - Second, consider the case where $-x_i < x_j \leq k_1(r)x_i$. Here, $\bar{x} > 0$. So the condition $|\bar{a}_{\mathrm{x}}| > |\bar{x}|$ simplifies to $\bar{a}_{\mathrm{x}} < \bar{x}$. From the previous case, we know that $\bar{a}_{\mathrm{x}} > \bar{x}$ if and only if $x_j > k_1(r)x_i$. This is not possible here since by definition $x_j \leq k_1(r)x_i$.
  - Third, consider the case where $x_j = -x_i$. Here $\bar{x} = 0$. So, we cannot have $\bar{a}_{\mathrm{x}}\bar{x} > 0$.
  - Fourth, consider the case where $x_j < -x_i$. Here, $\bar{x} < 0$. Since we require $\bar{a}_{\mathrm{x}}\bar{x}$ to be greater than zero, it follows that $\bar{a}_{\mathrm{x}} < 0$. So the condition $|\bar{a}_{\mathrm{x}}| > |\bar{x}|$ simplifies to $\bar{a}_{\mathrm{x}} < \bar{x} \Rightarrow x_j < k_1(r)x_i$, which is always true since by definition $x_j < -x_i$.

  Hence, polarization occurs in the first and third cases, *i.e.*, when $x_j > k_1(r)x_i$ or when $x_j < -x_i$.
- Reverse Polarization – $|\bar{a}_{\mathrm{x}}| > |\bar{x}|$ and $\bar{a}_{\mathrm{x}}\bar{x} \leq 0$.
  - First, consider the case where $x_j > -k_2(r)x_i$. Here both $\bar{a}_{\mathrm{x}} > 0$ and $\bar{x} > 0$ since $k_2(r) < 1$. So it cannot be that $\bar{a}_{\mathrm{x}}\bar{x} \leq 0$. Therefore, this case is ruled out.
  - Second, consider the case where $-x_i \leq x_j < -k_2(r)x_i$. Here $\bar{x} \geq 0$. So the condition $|\bar{a}_{\mathrm{x}}| \geq |\bar{x}|$ simplifies to $\bar{a}_{\mathrm{x}} \leq -\bar{x}$. This can be expressed as:

$$\frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i + \frac{1+r}{1+3r}x_j < -\frac{1+r}{1+3r}x_i - \frac{1+r}{1+3r}x_j$$

$$\Rightarrow \quad x_j < \frac{(1+3r)\left[(1+3r)^2+16r\right]}{[(1+3r)^2+4r(1-r)]\left[2(1+r)+(1+3r)\right]}x_i \tag{A.23}$$

  where the multiplier of x_i is labeled $k_2(r) = \frac{(1+3r)\left[(1+3r)^2+16r\right]}{[(1+3r)^2+4r(1-r)][2(1+r)+(1+3r)]}$. Since by definition $x_j < -k_2(r)x_i$, condition (A.23) is always satisfied.
  - Third, consider the case where $x_j < -x_i$. Here $\bar{x} < 0$. So the condition $|\bar{a}_{\mathrm{x}}| \geq |\bar{x}|$ simplifies to $\bar{a}_{\mathrm{x}} \geq \bar{x}$, which we know is the same as $x_j \geq k_1(r)x_i > 0$. However, this is not possible since by definition $x_j < -x_i$.

  Hence, Reverse Polarization only occurs when $-x_i \leq x_j < -k_2(r)x_i$.
- Moderation – $|\bar{a}_{\mathrm{x}}| \leq |\bar{x}|$.
  - First, consider the case where $x_j < -x_i$. Here, $\bar{x} \leq 0$. So the condition $|\bar{a}_{\mathrm{x}}| \leq |\bar{x}|$ simplifies to $-\bar{x} \leq \bar{a}_{\mathrm{x}} \leq \bar{x}$. If $-\bar{x} \leq \bar{a}_{\mathrm{x}}$, it then follows that $x_j \geq k_2(r)x_i \geq 0$, which is impossible since by definition $x_j < -x_i$.

- Second, consider the case where $-x_i \leq x_j < -k_2(r)x_i$. Here, $\bar{x} \geq 0$. So the condition $|\bar{a}_{\mathrm{x}}| \leq |\bar{x}|$ simplifies to $-\bar{x} \leq \bar{a}_{\mathrm{x}} \leq \bar{x}$. We know that this condition can be expressed as $-k_2(r)x_i \leq x_j \leq k_1(r)x_i$. This is not possible, since by definition $x_j < -k_2(r)x_i$.
- Third, consider the case where $-k_2(r)x_i \leq x_j \leq k_1(r)x_i$. Here, $\bar{x} > 0$. So the condition $|\bar{a}_{\mathrm{x}}| \leq |\bar{x}|$ simplifies to $\bar{x} \leq \bar{a}_{\mathrm{x}} \leq -\bar{x}$. This in turn can be expressed as $-k_2(r)x_i \leq x_j \leq k_1(r)x_i$, which we know is true by definition.
- Fourth, consider the case where $x_j > k_1(r)x_i$. Here, $\bar{x} > 0$. So the condition $|\bar{a}_{\mathrm{x}}| \leq |\bar{x}|$ simplifies to $\bar{x} \leq \bar{a}_{\mathrm{x}} \leq -\bar{x}$. This in turn can be expressed as $-k_2(r)x_i \leq x_j \leq k_1(r)x_i$, which cannot be true, since by definition $x_j > k_1(r)x_i$.

Hence, Moderation occurs only when $-k_2(r)x_i \leq x_j \leq k_1(r)x_i$.

Proof of Proposition 3

Required to show that, for $x_j > 0$, $\hat{a}_{\mathrm{x}2,j} \geq \hat{a}_j$ if $x_i \leq 0$, and $\hat{a}_{\mathrm{x}2,j} < \hat{a}_j$ if $x_i > 0$.
i) $x_i \leq 0$
$\hat{a}_{\mathrm{x}2,j} - \hat{a}_j$ can be simplified to $\hat{a}_{\mathrm{x}2,j} - \hat{a}_j = -\frac{(1-r)}{1+3r}\hat{a}_{\mathrm{x}1,i}$. We know that $\hat{a}_{\mathrm{x}1,i} = \mu_x(r)x_i \leq 0$ because $\mu_x(r) > 0$, $x_i \leq 0$. It therefore follows that $-\frac{(1-r)}{1+3r}\hat{a}_{\mathrm{x}1,i} \geq 0 \Rightarrow \hat{a}_{\mathrm{x}2,j} \geq \hat{a}_j$.
ii) $x_i > 0$
As before $\hat{a}_{\mathrm{x}2,j} - \hat{a}_j = -\frac{(1-r)}{1+3r}\hat{a}_{\mathrm{x}1,i}$. However, here $\hat{a}_{\mathrm{x}1,i} = \mu_x(r)x_i > 0$ because $\mu_x(r), x_i > 0$. Hence it follows that $\hat{a}_{\mathrm{x}2,j} - \hat{a}_j < 0 \Rightarrow \hat{a}_{\mathrm{x}2,j} < \hat{a}_j$.
b) The derivative of $\hat{a}_{\mathrm{x}1,i}$ w.r.t r can be calculated and simplified to $\frac{\mathrm{d}\hat{a}_{\mathrm{x}1,i}}{\mathrm{d}r} = -\frac{4(5r^2+6r+5)}{[(1+3r)^2+4r(1-r)]^2}x_i$.

Since $r > 0$, it follows that $\frac{\mathrm{d}\hat{a}_{\mathrm{x}1,i}}{\mathrm{d}r} < 0$

c) To show that $|\hat{a}_{\mathrm{x}1,i}| \geq |\hat{a}_i|$, we need to show that $\mu_{\mathrm{x}}(r) \geq \mu(r)$.

$$\begin{aligned}
\mu_{\mathrm{x}}(r) - \mu(r) &= \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)} - \frac{2(1+r)}{1+3r} \\
&= \frac{(1-r)^3}{(1+3r)[(1+3r)^2+4r(1-r)]} > 0 \text{ if } r > 0 \qquad \text{(A.24)}
\end{aligned}$$

Therefore, $\mu_{\mathrm{x}}(r) \geq \mu(r) \Rightarrow |\hat{a}_{\mathrm{x}1,i}| \geq |\hat{a}_i|$. □

Proof of Proposition 4

We now compare $\bar{a}_{\mathrm{x}}$ with $\bar{a}$. Without loss of generality, let $x_i \geq 0$.

- $|\bar{a}_{\mathrm{x}}| > |\bar{a}|$ and $\bar{a}_{\mathrm{x}}\bar{x} > 0$.
  - First, consider the case where $x_j \geq -x_i$. Here, $\bar{x} \geq 0$. So the condition $|\bar{a}_{\mathrm{x}}| > |\bar{a}|$ simplifies to $\bar{a}_{\mathrm{x}} > \bar{a} \Rightarrow \frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i > \frac{1+r}{1+3r}x_i$, which is impossible since $0 < r < 1$ and $x_i \geq 0$.
  - Second, consider the case where $-x_j < x_i$. Here $\bar{x} < 0$. So the condition $|\bar{a}_{\mathrm{x}}| > |\bar{a}|$ simplifies to $\bar{a}_{\mathrm{x}} < \bar{a} \Rightarrow$, $\frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i < \frac{1+r}{1+3r}x_i$, which is always true for $0 < r < 1$, $x_i \geq 0$.

  Hence, for $-x_j < x_i$, the mean outcome in the exogenous sequential game is more polarized than that in the simultaneous game, and this polarization is in the same direction as $\bar{x}$.
- $|\bar{a}_{\mathrm{x}}| > |\bar{a}|$ and $\bar{a}_{\mathrm{x}}\bar{x} \leq 0$.
  - First, consider the case where $x_j < -x_i$. Then $\bar{a}, \bar{x} < 0$. So for $\bar{a}_{\mathrm{x}}\bar{x} \leq 0$ to be true, we require $\bar{a}_{\mathrm{x}} \leq 0$, which is not possible since $\bar{a}_{\mathrm{x}} < \bar{x} < 0$.
  - Second, consider the case where $-x_i \leq x_j < -k_3(r)x_i$. Here, $\bar{x} \geq 0$ and the condition

$|\bar{a}_\mathrm{x}| > |\bar{a}|$ simplifies to $\bar{a}_\mathrm{x} > -\bar{a}$. This can be expressed as:

$$\begin{aligned} \frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i &< -\frac{1+r}{1+3r}x_i \\ \Rightarrow \quad x_j &< \frac{(1+3r)^3+8r(1-r)(1+r)}{2(1+r)\left[(1+3r)^2+4r(1-r)\right]}x_i \end{aligned} \tag{A.25}$$

where the multiplier of x_i is labeled $k_3(r) = \frac{(1+3r)^3+8r(1-r)(1+r)}{2(1+r)[(1+3r)^2+4r(1-r)]}x_i$. Since by definition, $x_j < k_3(r)x_i$, condition (A.25) is always satisfied.

- Third, consider the case where $x_j \geq k_3(r)x_i$. Then, $\bar{x} \geq 0$ and the $|\bar{a}_\mathrm{x}| > |\bar{a}|$ simplifies to $\bar{a}_\mathrm{x} < -\bar{a}$. However, from the second case, we know that this condition can only be satisfied when $x_j < k_3(r)x_i$, which cannot hold here, since by definition $x_j \geq k_3(r)x_i$.

Hence, for $-x_i \leq x_j < -k_3(r)x_i$, the mean outcome in the exogenous sequential game is more extreme than that in the simultaneous game, but in the direction opposite to that indicated by the mean preference $\bar{x}$.

- $|\bar{a}_\mathrm{x}| \leq |\bar{a}|$.
  - First, consider the case where $x_j \geq -k_3(r)x_i$. Here, $\bar{a}_\mathrm{x}, \bar{x}, \bar{a} \geq 0$. So the condition $|\bar{a}_\mathrm{x}| \leq |\bar{a}|$ simplifies to $\bar{a}_\mathrm{x} \leq \bar{a} \Rightarrow$, $\frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i \leq \frac{1+r}{1+3r}x_i$, which is always true for $0 < r < 1$, $x_i \geq 0$.
  - Second, consider the case where $-x_i \leq x_j < k_3(r)x_i$. Here also $\bar{x}, \bar{a} \geq 0$. So the condition $|\bar{a}_\mathrm{x}| \leq |\bar{a}|$ simplifies to $-\bar{a} \leq \bar{a}_\mathrm{x} \leq \bar{a}$. From the case before, we know that $\bar{a}_\mathrm{x} \leq \bar{a}$ for $x_i \geq 0$. However, to ensure that $-\bar{a} \leq \bar{a}_\mathrm{x}$, we need $x_j \geq k_3(r)x_i$, which cannot be true since by definition $-x_i \leq x_j < k_3(r)x_i$.
  - Third, consider the case where $x_j < -x_i$. Here, $\bar{x}, \bar{a} < 0$. So the condition $|\bar{a}_\mathrm{x}| \leq |\bar{a}|$ simplifies to $\bar{a} \leq \bar{a}_\mathrm{x} \leq -\bar{a}$. The condition $\bar{a}_\mathrm{x} \geq \bar{a}$ reduces to $x_j \geq -k_3(r)x_i$, which we know is not possible since $x_j < -x_i$ and $0 < k_3(r) < 1$.

Hence, for $x_j \geq -k_3(r)x_i$, the mean outcome in the exogenous sequential game is less extreme (moderate) compared to that in the simultaneous game.

Proof of Proposition 5

The expected utility of the first player i in equilibrium is given by (7). Substituting for $\hat{a}_{\mathrm{x}1,i}$ gives us:

$$\begin{aligned} EU_{\mathrm{x}1}\left(x_i, \hat{a}_{\mathrm{x}1,i}\right) = \quad & - \quad r\left(x_i - \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i\right)^2 \\ & - \quad (1-r)\left[\left(x_i - \frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i\right)^2 + \frac{(1+r)^2}{3(1+3r)^2}\int_{\mathbb{R}} x_j^2 g(x_j)\mathrm{d}x_j\right] \end{aligned} \tag{A.26}$$

This in turn simplifies to:

$$EU_{\mathrm{x}1}\left(x_i, \hat{a}_{\mathrm{x}1,i}\right) = -\frac{(1-r)(1+r)^2}{(1+3r)^2+4r(1-r)}x_i^2 - \frac{(1-r)(1+r)^2}{(1+3r)^2}\int_{\mathbb{R}} x_j^2 g(x_j)\mathrm{d}x_j \tag{A.27}$$

Next, consider the a priori expected utility of player j, in equilibrium. It is obtained by integrating the utility of the second player over the range of x_i. That is, $EU_{\mathrm{x}2}\left(x_j, \hat{a}_{\mathrm{x}2,j}\right) =$

$\int_{\mathbb{R}} u\left(x_j, \hat{a}_{\mathrm{x}2,j}, \hat{a}_{\mathrm{x}1,i}\right) g(x_i)\mathrm{d}x_i$. Substituting for $\hat{a}_{\mathrm{x}2,j}$ as $\frac{2(1+r)}{1+3r}x_j - \frac{(1-r)}{1+3r}a_{\mathrm{x}1,i}$, we have:

$$EU_{\mathrm{x}2}\left(x_j, \hat{a}_{\mathrm{x}2,j}\right) = - r\int_{\mathbb{R}}\left(x_j - \frac{2(1+r)}{1+3r}x_j - \frac{1-r}{1+3r}\hat{a}_{\mathrm{x}1,i}\right)^2 g(x_i)\mathrm{d}x_i - (1-r)\int_{\mathbb{R}}\left(x_j - \frac{1+r}{1+3r}x_j - \frac{2r}{1+3r}\hat{a}_{\mathrm{x}1,i}\right)^2 g(x_i)\mathrm{d}x_i \tag{A.28}$$

Substituting for $\hat{a}_{\mathrm{x}1,i}$ and integrating, we have:

$$EU_{\mathrm{x}2}\left(x_j, \hat{a}_{\mathrm{x}2,j}\right) = -\frac{r(1-r)}{1+3r}x_i^2 - \frac{r(1-r)(1+3r)(3+r)^2}{\left[(1+3r)^2+4r(1-r)\right]^2}\int_{\mathbb{R}} x^2 g(x)\mathrm{d}x \tag{A.29}$$

Next, we compare the difference in the expected utilities for a specific player i, where $D_{\mathrm{x}}(x_i) = EU_{\mathrm{x}1}\left(x_i, \hat{a}_{\mathrm{x}1,i}\right) - EU_{\mathrm{x}2}\left(x_i, \hat{a}_{\mathrm{x}2,i}\right)$. We can show that $D_{\mathrm{x}}(x_i) \leq 0$ for all i if the following two conditions are satisfied:

$$\frac{r(1-r)(1+3r)(3+r)^2}{\left[(1+3r)^2+4r(1-r)\right]^2} > -\frac{(1-r)(1+r)^2}{(1+3r)^2} \tag{A.30}$$

and

$$-\frac{r(1-r)}{1+3r} \geq -\frac{(1-r)(1+r)^2}{(1+3r)^2+4r(1-r)} \tag{A.31}$$

First, consider the inequality (A.30), which can be simplified to:

$$(1+r)^2\left[(1+3r)^2+4r(1-r)\right]^2 > r(3+r)^2(1+3r)^3 \tag{A.32}$$

This in turn simplifies to:

$$\begin{aligned} & (1+3r)^3\left[(1+r)^2(1+3r) - r(3+r)^2\right] + (1+r)^2\left[16r^2(1-r)^2 + 8r(1-r)(1+3r)^2\right] > 0 \\ \Rightarrow\ & 16r^2(1+r)^2(1-r)^2 + (1+3r)^2(1-r)\left[2r(1+r)^2 + (1+3r)(1-r)\right] > 0 \end{aligned} \tag{A.33}$$

Since both the terms in the R.H.S of inequality (A.33) are non-negative, the inequality is always true. Therefore, (A.30) is always true. Next, consider the inequality (A.31), which can be expressed as:

$$\frac{(1-r)(1+r)^2}{(1+3r)^2+4r(1-r)}x_i^2 \geq \frac{r(1-r)}{1+3r}x_i^2 \tag{A.34}$$

This is true if:

$$\begin{aligned} & (1-r)\left[(1+r)^2(1+3r) - r[(1+3r)^2+4r(1-r)]\right]x_i^2 \geq 0 \\ \Rightarrow\ & (1-r)^2\left[2r^2+5r+1\right]x_i^2 \geq 0 \end{aligned} \tag{A.35}$$

We know that $x_i^2 \geq 0$ and that both $(1-r)^2$ and $2r^2+5r+1$ are positive for $0 < r < 1$. Hence this inequality is always true too. Further, since both (A.30) and (A.31) are always true, it follows that $EU_{\mathrm{x}2}\left(x_i, \hat{a}_{\mathrm{x}2,i}\right) > EU_{\mathrm{x}1}\left(x_i, \hat{a}_{\mathrm{x}1,i}\right)$.
b) Now we prove the second part of the Proposition. Let $x_i > 0$, then:

$$\frac{\mathrm{d}D_{\mathrm{x}}(x_i)}{\mathrm{d}x_i} = 2x_i(1-r)\left[\frac{r}{1+3r} - \frac{(1+r)^2}{(1+3r)^2+4r(1-r)}\right] \tag{A.36}$$

Since we have already shown that (A.31) is true, we know that $(1-r)\left[\frac{r}{1+3r} - \frac{(1+r)^2}{(1+3r)^2+4r(1-r)}\right] > 0$. It therefore follows that $2x_i(1-r)\left[\frac{r}{1+3r} - \frac{(1+r)^2}{(1+3r)^2+4r(1-r)}\right] \leq 0$. Hence, $\frac{\mathrm{d}D_{\mathrm{x}}(x_i)}{\mathrm{d}x_i} \leq 0$. Similar proof applies for $x_i < 0$. □

Proof of Proposition 6

The solutions for the optimal actions for periods 3 and 4 are analogous to that in the exogenous sequential choice game and are outlined in the main text. Below, we derive the players' optimal action for the first two periods.

Period 2 – A player who has lost the auction makes no decisions in period 2. So we only consider the actions of a player who won the auction in period 1. Suppose player j bids according to the symmetric bidding function $\beta(\cdot)$, then player i belief upon winning is that player j must belong to a some symmetric region W for her to have won the auction. In that case, i's expected utility from speaking first is:

$$\begin{aligned} EU_{\mathrm{n1}}\left(x_i, a_{\mathrm{n1},i}\right) = & - r(x_i - a_{\mathrm{n1},i})^2 - (1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{\mathrm{n1},i}\right)^2 + \left(\frac{1+r}{1+3r}\right)^2 \frac{\int_W x_j^2 g(x_j)\mathrm{d}x_j}{\int_W g(x_j)\mathrm{d}x_j}\right] \\ & - 2\frac{1+r}{1+3r}\left(x_i - \frac{2r}{1+3r}a_{\mathrm{n1},i}\right)\frac{\int_W x_j g(x_j)\mathrm{d}x_j}{\int_W g(x_j)\mathrm{d}x_j} \end{aligned} \tag{A.37}$$

The last term vanishes because W is symmetric around zero. So:

$$\begin{aligned} EU_{\mathrm{n1}}\left(x_i, a_{\mathrm{n1},i}\right) &= -r(x_i - a_{\mathrm{n1},i})^2 - (1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{\mathrm{n1},i}\right)^2 + \left(\frac{1+r}{1+3r}\right)^2 \frac{\int_W x_j^2 g(x_j)\mathrm{d}x_j}{\int_W g(x_j)\mathrm{d}x_j}\right] \\ &= -\frac{(1-r)(1+r)^2}{(1+3r)^2+4r(1-r)}x_i^2 - \frac{(1-r)(1+r)^2}{(1+3r)^2}\frac{\int_W x_j^2 g(x_j)\mathrm{d}x_j}{\int_W g(x_j)\mathrm{d}x_j} \end{aligned} \tag{A.38}$$

Similarly, simplifying i's expected utility from speaking second gives us:

$$EU_{\mathrm{n2}}\left(x_i, a_{\mathrm{n2},i}\right) = -\frac{r(1-r)}{(1+3r)}\left[x_i^2 + \left(\frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}\right)^2 \int_W g(x_j)x_j^2 \mathrm{d}x_j\right] \tag{A.39}$$

i prefers to speak second if:

$$EU_{\mathrm{n2}}\left(x_i, a_{\mathrm{n2},i}\right) > EU_{\mathrm{n1}}\left(x_i, a_{\mathrm{n1},i}\right) \tag{A.40}$$

Let $A(r) = \frac{(1-r)(1+r)^2}{(1+3r)^2+4r(1-r)}$, $B(r) = \frac{(1-r)(1+r)^2}{(1+3r)^2}$, $C(r) = \frac{r(1-r)}{(1+3r)}$, and $D(r) = \frac{r(1-r)}{(1+3r)}\left(\frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}\right)^2$. It is trivial to show that the multiplier of x_i^2 in $EU_{\mathrm{n2}}\left(x_i, a_{\mathrm{n2},i}\right)$ is always greater than that in $EU_{\mathrm{n1}}\left(x_i, a_{\mathrm{n1},i}\right) \Rightarrow A(r) > C(r)$. Similarly, we can show that the multiplier of $\int_W x_j^2 g(x_j)\mathrm{d}x_j$ in $EU_{\mathrm{n2}}\left(x_i, a_{\mathrm{n2},i}\right)$ is also greater than that in $EU_{\mathrm{n1}}\left(x_i, a_{\mathrm{n1},i}\right) \Rightarrow B(r) > D(r)$. Therefore, in a symmetric bidding equilibrium, for all x_is, the expected value from speaking second is higher than that from speaking first. Therefore, upon winning the auction, all types will choose to speak second.

Period 1 At the beginning of period 1, before placing her bid, player i knows that if she wins, she will choose to go second and if player j wins, she will go first. Hence, her expected utility from choosing a bid b_i and then choosing the optimal action in the subsequent periods is

$$EU(x_i, b_i) = \frac{\int_{L_i} u(x_i, \hat{a}_{\mathrm{n1},i}, \hat{a}_{\mathrm{n2},j})g(x_j)\mathrm{d}x_j + \int_{W_i}\left[u(x_i, \hat{a}_{\mathrm{n1},j}, \hat{a}_{\mathrm{n2},i}) - b_i\right]g(x_j)\mathrm{d}x_j}{\int_{\mathbb{R}} g(x_j)\mathrm{d}x_j} \tag{A.41}$$

where $j \in W_i$ for i to win the auction and $j \in L_i$ for her to lose the auction, if she chooses a bid b_i.

This simplifies to:

$$\begin{aligned}2EU(x_i,b_i) = & - \quad A(r)x_i^2\int_{L_i} g(x_j)\mathrm{d}x_j - B(r)\int_{L_i} x_j^2 g(x_j)\mathrm{d}x_j \\ & - \quad C(r)x_i^2\int_{W_i} g(x_j)\mathrm{d}x_j - D(r)\int_{W_i} x_j^2 g(x_j)\mathrm{d}x_j - \int_{W_i} b_i g(x_j)\mathrm{d}x_j \\ = & - \quad 2x_i^2 + [A(r)-C(r)]x_i^2\int_{W_i} g(x_j)\mathrm{d}x_j + [B(r)-D(r)]\int_{W_i} x_j^2 g(x_j)\mathrm{d}x_j - \int_{W_i} b_i g(x_j)\mathrm{d}x_j \end{aligned} \tag{A.42}$$

Now consider any two types x' and x'' and a bidding function $\beta(\cdot)$. In equilibrium, x' can do no better by playing x'''s strategy $\beta(x'')$ over her own strategy $\beta(x')$ and vice-versa. That is:

$$EU(x',\beta(x')) \geq EU(x',\beta(x'')) \tag{A.43}$$
$$EU(x'',\beta(x'')) \geq EU(x',\beta(x'')) \tag{A.44}$$

Substituting the simplified expressions for the expected utilities into the above inequalities and adding them up gives us:

$$[A(r)-C(r)]\,(x'^2 - x''^2)\left(\int_{W_1} g(x_j)\mathrm{d}x_j - \int_{W_2} g(x_j)\mathrm{d}x_j\right) \geq 0 \tag{A.45}$$

We know that $A(r) - C(r) > 0$. So if $x'^2 > x''^2$, then for the above inequality to hold, we require that $\int_{W_1} g(x_j)\mathrm{d}x_j - \int_{W_2} g(x_j)\mathrm{d}x_j \geq 0 \Rightarrow$ the region over which a player wins upon bidding $\beta(x')$ is greater than that over which she wins when she bids $\beta(x'')$. In other words, the equilibrium bidding strategies are monotonically increasing in $|x|$. Further, following the technique as that outlined in Fudenberg and Tirole (1991) p. 217, we can show strict monotonicity, *i.e.*, if $|x'| > |x''|$, then $\beta(x') > \beta(x'')$.

Now that we have shown that the bidding strategies are monotonically increasing in $|x|$, for the specific bid b_i by player i (when the other player uses the bidding function $\beta(\cdot)$), we re-write Equation (A.42) as:

$$EU(x_i,b_i) = -2x_i^2 + 2[A(r)-C(r)]x_i^2\int_0^{\beta^{-1}(b_i)} g(x)\mathrm{d}x + 2[B(r)-D(r)]\int_0^{\beta^{-1}(b_i)} x^2 g(x)\mathrm{d}x - 2\int_0^{\beta^{-1}(b_i)} b_i g(x)\mathrm{d}x \tag{A.46}$$

Further, we specify the following expression for the derivatives:

$$\frac{\mathrm{d}\left[\int_0^{\beta^{-1}(b_i)} F(x)\right]}{\mathrm{d}b_i}\mathrm{d}x = F(V)\frac{\mathrm{d}V}{\mathrm{d}b_i} \tag{A.47}$$

where $V = \beta^{-1}(b_i)$. To obtain the equilibrium bidding function, we can calculate the F.O.C of Equation (A.46) as $\left.\frac{\mathrm{d}EU(x_i,b_i)}{\mathrm{d}b_i}\right|_{b_i=\hat{b}_i} = 0$. This simplifies to:

$$[A(r)-C(r)]\,x_i^2\frac{g(\beta^{-1}(\hat{b}_i))}{\beta'(\beta^{-1}(\hat{b}_i))} + [B(r)-D(r)]\frac{g(\beta^{-1}(\hat{b}_i))\left[\beta^{-1}(\hat{b}_i)\right]^2}{\beta'(\beta^{-1}(\hat{b}_i))} - \left[\hat{b}_i\frac{g(\beta^{-1}(\hat{b}_i))}{\beta'(\beta^{-1}(\hat{b}_i))} + \int_0^{\beta^{-1}(\hat{b}_i)} g(x)\mathrm{d}x\right] = 0$$

In equilibrium $\hat{b}_i = \beta(x_i)$ and so $\beta^{-1}(\hat{b}_i) = x_i$. So the above equation simplifies to:

$$[A(r)-C(r)+B(r)-D(r)]\,x_i^2 g(\beta^{-1}(\hat{b}_i)) = \frac{\mathrm{d}\left[\beta(x_i)\int_0^{x_i} g(x)\mathrm{d}x\right]}{\mathrm{d}x_i} \tag{A.48}$$

Integrating this from 0 to x_i, we have:

$$\hat{\beta} = \beta(x_i) = [A(r) - C(r) + B(r) - D(r)] \left[\frac{\int_0^{x_i} x^2 g(x)\mathrm{d}x}{\int_0^{x_i} g(x)\mathrm{d}x} \right] \tag{A.49}$$

Since $A(r) - C(r) > 0$ and $B(r) - D(r) > 0$, the multiplier of $\left[\frac{\int_0^{x_i} x^2 g(x)\mathrm{d}x}{\int_0^{x_i} g(x)\mathrm{d}x} \right]$ is positive which recovers the assumption that the bidding function is symmetric around zero. $\square$

For Online Publication: Welfare Comparisons

Welfare Under First-Best Planner's Choice

The social planner's choice involve agents choosing their true preferences as their actions, *i.e.*, $a_i = x_i$ and $a_j = x_j$, and $\bar{a} = \frac{x_i + x_j}{2}$. Then:

$$\begin{aligned} W_{FB}(x_i, x_j, a_i, a_j) &= W_{FB}(x_i, x_j) \\ &= -r(x_i - a_i)^2 - (1-r)(x_i - \bar{a})^2 - r(x_j - a_j)^2 - (1-r)(x_j - \bar{a})^2 \\ &= -(1-r)\frac{(x_i - x_j)^2}{2} \end{aligned} \tag{A.50}$$

Then the expected welfare is given by:

$$\begin{aligned} EW_P &= \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} W_1(x_i, x_j) g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j}{\int_{\mathbb{R}} \int_{\mathbb{R}} g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j} \\ &= -\frac{(1-r)}{2 \int_{\mathbb{R}} \int_{\mathbb{R}} g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j} \left[\int_{\mathbb{R}} \int_{\mathbb{R}} (x_i^2 + x_j^2 - 2x_i x_j) g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j \right] \end{aligned}$$

We know that $\int_{\mathbb{R}} \int_{\mathbb{R}} g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j = 1$ and $\int_{\mathbb{R}} \int_{\mathbb{R}} x_i x_j g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j = 0$ because $g(\cdot)$ is a symmetric distribution. Therefore, EW_1 simplifes to:

$$\begin{aligned} EW_{FB} &= \frac{-(1-r)}{2} \int_{\mathbb{R}} \int_{\mathbb{R}} \left[x_i^2 + x_j^2 \right] g(x_i) g(x_j) \mathrm{d}x_j \mathrm{d}x_j \\ &= -(1-r) \int_{\mathbb{R}} x^2 g(x) \mathrm{d}x \end{aligned} \tag{A.51}$$

Welfare in the Simultaneous Game

In the simultaneous game, the actions are $a_i = \frac{2(1+r)}{1+3r} x_i$, $a_j = \frac{2(1+r)}{1+3r} x_j$, and the mean action is $\bar{a} = \frac{(1+r)}{1+3r}(x_i + x_j)$. Substituting this in the welfare function we can obtain:

$$\begin{aligned} W_s(x_i, x_j, a_i, a_j) &= W_s(x_i, x_j) \\ &= -\frac{r(1-r)^2}{(1+3r)^2}(x_i^2 + x_j^2) - \frac{1-r}{(1+3r)^2} \left[2r x_i - (1+r) x_j \right]^2 \\ &\quad - \frac{1-r}{(1+3r)^2} \left[2r x_j - (1+r) x_i \right]^2 \\ &= -\frac{r(1-r)^2}{(1+3r)^2}(x_i^2 + x_j^2) \\ &\quad - \frac{1-r}{(1+3r)^2} \left[4r^2 (x_i^2 + x_j^2) + (1+r)^2 (x_i^2 + x_j^2) - 8r(1+r) x_i x_j \right] \end{aligned} \tag{A.52}$$

The expected welfare is then given by:

$$EW_s = \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} W_s(x_i, x_j) g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j}{\int_{\mathbb{R}} \int_{\mathbb{R}} g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j}$$

As before, $\int_{\mathbb{R}} \int_{\mathbb{R}} g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j = 1$ and $\int_{\mathbb{R}} \int_{\mathbb{R}} x_i x_j g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j = 0$. So:

$$EW_s = -\frac{2(1-r)}{(1+3r)^2} [r(1-r) + 4r^2 + (1+r)^2] \int_{\mathbb{R}} x^2 g(x) \mathrm{d}x \tag{A.53}$$

Welfare in the Exogenous Sequential Choice Game

Without loss of generality, assume that i speaks first and j speaks second. Then, we have the actions of the two agents as $a_i = \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)} x_i$, $a_j = \frac{2(1+r)}{1+3r} x_j - \frac{1-r}{1+3r} a_i$, and the mean action as $\bar{a} = \frac{2r(3+r)}{(1+3r)^2+4r(1-r)} x_i + \frac{1+r}{1+3r} x_j$. Using these, we can further derive the following expressions:

$$\begin{aligned}
x_i - a_i &= -\frac{2(1-r)(1+r)}{(1+3r)^2+4r(1-r)} x_i \\
x_j - a_j &= -\frac{1-r}{1+3r}\left[x_j - \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)} x_i\right] \\
x_i - \bar{a} &= \frac{(1+3r)(1+r)}{(1+3r)^2+4r(1-r)} x_i - \frac{1+r}{1+3r} x_j \\
x_j - \bar{a} &= \frac{2r}{1+3r} x_j - \frac{2r(3+r)}{(1+3r)^2+4r(1-r)} x_i
\end{aligned}$$

Substituting the above terms in the welfare equation, we have:

$$\begin{aligned}
W_x(x_i, x_j, a_i, a_j) &= W_x(x_i, x_j) \\
&= -\frac{4r(1-r)^2(1+r)^2}{((1+3r)^2+4r(1-r))^2} x_i^2 - (1-r)\left[\frac{(1+3r)(1+r)}{(1+3r)^2+4r(1-r)} x_i - \frac{1+r}{1+3r} x_j\right]^2 \\
&\quad -\frac{r(1-r)^2}{(1+3r)^2}\left[x_j - \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)} x_i\right]^2 \\
&\quad -(1-r)\left[\frac{2r}{1+3r} x_j - \frac{2r(3+r)}{(1+3r)^2+4r(1-r)} x_i\right] \qquad \text{(A.54)}
\end{aligned}$$

As before, the expected welfare is given by:

$$EW_x = \frac{\int_{\mathbb{R}}\int_{\mathbb{R}} W_3(x_i, x_j) g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j}{\int_{\mathbb{R}}\int_{\mathbb{R}} g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j}$$

As before, this implies that the integrals of the $x_i x_j$ terms are canceled out. Further, and $\int_{\mathbb{R}}\int_{\mathbb{R}} g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j = 1$.

$$\begin{aligned}
EW_x = &- \frac{4r(1-r)^2(1+r)^2}{\left[(1+3r)^2+4r(1-r)\right]^2}\int_{\mathbb{R}}\int_{\mathbb{R}} x_i^2 g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j \\
&- \frac{r(1-r)^2}{(1+3r)^2}\int_{\mathbb{R}}\int_{\mathbb{R}}\left[x_j^2 + \frac{(1+3r)^2(3+r)^2}{\left[(1+3r)^2+4r(1-r)\right]^2} x_i^2\right] g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j \\
&- (1-r)\int_{\mathbb{R}}\int_{\mathbb{R}}\left[\frac{(1+3r)^2(1+r)^2+4r^2(3+r)^2}{\left[(1+3r)^2+4r(1-r)\right]^2} x_i^2 + \frac{(1+r)^2+4r^2}{(1+3r)^2} x_j^2\right] g(x_i) g(x_j) \mathrm{d}x_i \mathrm{d}x_j
\end{aligned}$$

This simplifies to:

$$\begin{aligned}
EW_x = &-\frac{(1-r)\left[(1+r)^2+4r^2+r(1-r)\right]}{3(1+3r)^2}\int_{\mathbb{R}} x^2 g(x) \mathrm{d}x \\
&- \left[\frac{(1-r)\left[4r(1-r)(1+r)^2+4r^2(3+r)^2+(1+3r)^2(1+r)^2+r(1-r)(3+r)^2\right]}{3\left[(1+3r)^2+4r(1-r)\right]^2}\right]\int_{\mathbb{R}} x^2 g(x) \mathrm{d}x \qquad \text{(A.55)}
\end{aligned}$$

Welfare in the Endogenous Sequential Choice Game

As before, assume that i speaks first and j speaks second. Recall that the players' actions here are the same as that in the exogenous sequential choice game. However, we know that $|x_i| < |x_j|$. So while the welfare equation remains the same, the integrations regions are different. Thus, we have:

$$\begin{aligned}
W_n(x_i, x_j, a_i, a_j) &= W_n(x_i, x_j) \\
&= -\frac{4r(1-r)^2(1+r)^2}{((1+3r)^2+4r(1-r))^2}x_i^2 - (1-r)\left[\frac{(1+3r)(1+r)}{(1+3r)^2+4r(1-r)}x_i - \frac{1+r}{1+3r}x_j\right]^2 \\
&\quad -\frac{r(1-r)^2}{(1+3r)^2}\left[x_j - \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i\right]^2 \\
&\quad -(1-r)\left[\frac{2r}{1+3r}x_j - \frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i\right]
\end{aligned} \tag{A.56}$$

and the expected welfare is:

$$EW_n = \frac{\iint\limits_{R1\cup R2} W_4(x_i, x_j)g(x_i)g(x_j)\mathrm{d}x_i\mathrm{d}x_j}{\iint\limits_{R1\cup R2} g(x_i)g(x_j)\mathrm{d}x_i\mathrm{d}x_j} \tag{A.57}$$

where the two regions R1 and R2 are defined as follows:

$$\begin{aligned}
R_1 &\equiv x_i \in [0,\infty], \quad x_j \in [x_i,\infty] \cup [-\infty,-x_i] \\
R_2 &\equiv x_i \in [-\infty,0], \quad x_j \in [-x_i,\infty] \cup [-\infty,x_i]
\end{aligned}$$

As before, we can show that the integral of the $x_i x_j$ terms over $R_1 \cup R_2$ is zero. So we now consider the integrals of the x_j^2 and x_i^2 terms. Since inference on j's type is conditional on i, we first integrate over j's type. Because of the symmetry of the distrbution, it is easy to show that:

$$\iint\limits_{R1\cup R2} x_i^2 g(x_i)g(x_j)\mathrm{d}x_i\mathrm{d}x_j = 4\int_0^\infty x^2 g(x)(1-G(x))\mathrm{d}x \tag{A.58}$$

$$\iint\limits_{R1\cup R2} g(x_i)g(x_j)\mathrm{d}x_i\mathrm{d}x_j = 4\int_0^\infty g(x)(1-G(x))\mathrm{d}x \tag{A.59}$$

Substituting these expressions back in the expected welfare function, we have:

$$\begin{aligned}
EW_n = &\left(-\frac{(1-r)\left[(1+r)^2+4r^2+r(1-r)\right]}{(1+3r)^2}\right. \\
&-\left.\frac{(1-r)\left[4r(1-r)(1+r)^2+4r^2(3+r)^2+(1+3r)^2(1+r)^2+r(1-r)(3+r)^2\right]}{3\left[(1+3r)^2+4r(1-r)\right]^2}\right) \\
&\cdot\left(\frac{\int_0^\infty x^2 g(x)(1-G(x))\mathrm{d}x}{\int_0^\infty x^2 g(x)(1-G(x))\mathrm{d}x}\right)
\end{aligned} \tag{A.60}$$