AMERICAN MARKETING ASSOCIATION

Article

Journal of Marketing Research
2021, Vol. 58(4) 782-800

DOI: 10.1177/00222437211016389
journals.sagepub.com/home/mrj
SAGE

Strategic Polarization in Group Interactions

Ganesh Iyer and Hema Yoganarasimhan

Abstract

The authors study the phenomenon of strategic group polarization, in which members take more extreme actions than their preferences. The analysis is relevant for a broad range of formal and informal group settings, including social media, online platforms, sales teams, corporate and academic committees, and political action committees. In the model, agents with private preferences choose a public action (voice opinions), and the mean of their actions represents the group's realized outcome. The agents face a trade-off between influencing the group decision and truth-telling. In a simultaneous-move game, agents strategically shade their actions toward the extreme. The strategic group influence motive can create substantial polarization in actions and group decisions even when the preferences are relatively moderate. Compared with a simultaneous game, a randomized-sequential-actions game lowers polarization when agents' preferences are relatively similar. Sequential actions can even lead to moderation if the later agents have moderate preferences. Endogenizing the order of moves (through a first-price sealed-bid auction) always increases polarization, but it is also welfare enhancing. These findings can help group leaders, firms, and platforms design mechanisms that moderate polarization, such as the choice of speaking order, the group size, and the knowledge members have of others' preferences and actions.

Keywords

polarization, collective decision making, social choice, lobbying, group interactions

Online supplement: https://doi.org/10.1177/00222437211016389

Many business and organizational settings involve group interactions and decisions. We observe formal and informal group discussions in contexts ranging from social media groups to sales teams, corporate personnel committees, academic committees, community groups, and political action committees. The overarching premise and goal of group interactions is to allow agents to exchange opinions, increase the alignment of their views, and reduce conflict (Rawls 1971). However, we often see the opposite: individuals engaged in group interactions often become more divergent in their views and behaviors. Indeed, a large body of experimental work shows that group deliberations often lead to more polarized behavior instead of enabling greater alignment (Isenberg 1986). This kind of polarized behavior is especially common when group members' actions impact their joint decision or group verdict.

We study the phenomenon of polarization of agents' observed actions in group settings; that is, where members of a group take actions (or voice opinions) that are more extreme than their true preferences (Sunstein 2002). Consider the following examples that illustrate relevant aspects of the phenomenon studied in this article:

- **Brand Perceptions:** Advertising campaigns of brands that take positions on sociopolitical issues (e.g., Nike's "Believe in Something" campaign) are often actively discussed on social media forums such as Twitter and Facebook. In these online discussions, consumers with liberal and conservative political preferences compete to make their views about the brands heard (Taylor 2017). How do these interactions shape overall brand perceptions?
- **Corporate Lobbying and Climate Change:** During the heated climate change debates of 2008 in Congress, the most intensive lobbying efforts in the electric utilities industry were by two companies at the opposite and extreme ends of the environmental performance spectrum. Southern Company, one of the highest polluting utilities in the United States, spent $14 million in 2008 on climate change lobbying. At the same time, PG&E, one of the greenest utilities in terms of carbon emissions, had an even more intense campaign and spent an estimated $27 million (Hulac 2016). Existing research (Delmas, Lim, and Nairn-Birch 2016) provides empirical evidence for a systematic U-shaped relationship

Ganesh Iyer is Professor, Haas School of Business, University of California, Berkeley, USA (email: giyer@haas.berkeley.edu). Hema Yoganarasimhan is Professor, Foster School of Business, University of Washington, USA (email: hemay@uw.edu).

between greenhouse emissions and lobbying intensity: the companies with the least and most carbon emissions are the most aggressive and extreme voices in the climate debate. What might account for this pattern of behavior?

- **Diversity and Inclusion:** Human resources groups within firms have struggled with the debate on balancing merit versus diversity in recruiting decisions. Often, the final decisions look more extreme than the individual preferences of group members. For example, a former technical recruiter alleges that Google's senior executives implemented policies stating that certain hires from the third quarter of 2017 must be "all diverse," meaning that meant all hires had to be Black, Hispanic, or female (none could be a White or Asian man). This decision was considered more extreme than expected even by some within the company (Eastland 2018).
- **Faculty Hiring:** Faculty hiring decisions in marketing departments often bring into play motivations to influence the group. Consider the potential debate about which subarea to hire in (e.g., behavioral vs. quantitative, theory vs. empirical) or which emerging area to focus on (e.g., machine learning, field experiments). Group members may have different private preferences on how to allocate scarce faculty slots across different research streams. What mechanisms can a department chair use to reduce polarization and division in the hiring deliberations?
- **Gun Control:** In 2015, Texas passed the "campus carry" law, formally known as Senate Bill 11 (Aguilar 2015). Following this, the University of Texas assembled a 19-member working group to discuss how to implement this law into practice. The working group's members had disparate views and 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 extreme outcomes, such as the conclusion that the committee would not recommend a ban on guns in the classroom (Campus Carry Policy 2015).[1]
- **Online Reviews:** Research shows that reviewers on platforms such as Amazon and Yelp place significant weight on the reviews of others (Godes and Silva 2012). Furthermore, some reviewers on Yelp seem to have a greater motivation to deviate from the ratings posted by others and overweight their own existing beliefs (Dai et al. 2018). Recent work also points out that more extreme beliefs may be overrepresented in online reviews (Klein et al. 2018). How does the temporal sequencing of reviews affect overall product ratings?

[1] For comprehensive discussions of campus carry laws across different states and the recent developments in this area, see Huitlin (2015) and Armed Campuses (http://www.armedcampuses.org/).

Common Themes

Some common themes emerge from these examples, and these are the focus of our analysis. First, they represent contexts in which agents have strongly held preferences. For instance, in hiring decisions, agents often have strong preferences on the role of diversity or the importance of research areas. Similarly, consumers may have strong political preferences that shape their reactions to brands' sociopolitical stances in their advertising campaigns. Issues that dominate our public policy discussions also fall under this category (e.g., abortion, immigration, the size of government). Across the spectrum of these cases, when participating in a group discussion, an agent's motivation is to achieve a group outcome that closely aligns with her preferences. This is distinct from settings where agents care about some true underlying state of the world and have uncertainty about this state. In those settings, agents typically try to aggregate their information from the group to uncover the uncertainty and condition their decision on it. In contrast, our research captures settings where the actions/voiced opinions represent agents' stated preferences rather than beliefs/information about an unknown true state of the world.

Second, unlike standard voting models, individuals' actions or voiced opinions are not necessarily binary/discrete. Rather, they are typically a choice of the extent or the magnitude of an action (i.e., continuous choice).[2] In the faculty hiring example, it could be the strength and the number of arguments presented by group members to make a behavioral versus quantitative faculty hire. In the campus carry example, the choice was not a simple "Should guns be allowed on campus or not?" Rather, it was a nuanced decision on issues such as where guns would be allowed (classrooms), where they should not be allowed (child-care units), where they should be allowed with discretion (single-user offices), where they can be stored (in-person or locked vehicle), and who would be allowed to bring guns (those over 21 years old, with license).

Third, in many of these cases, agents voice their opinions sequentially and can be influenced by prior opinions in the system. For example, in online reviews on digital platforms, earlier reviews can influence (and can be influenced by[3]) future reviews. Similarly, in social media discussions, users often respond to prior comments, and this sequential voicing of opinions can shape the overall tendency of the group (De-Wit, Van der Linden, and Brick 2019).

Further, note that the outcomes in these examples share two key characteristics. First, both individual opinions and the group's joint decision/outcome end up being more extreme than one would expect a priori based on the initial distribution of the group members' preferences. This pattern

[2] There is a stream of theoretical literature that establishes criteria to compare different voting methods. For a discussion of this topic, see Pacuit (2011); for an early simulation study of different voting schemes, see Bordley (1983).

[3] Earlier reviewers may change the content of their reviews in anticipation of the content in future reviews.

Table 1. Summary of the Key Features and Outcomes in the Examples.

	Features			Outcomes	
Example	Strongly Held Preferences	Nonbinary/ Continuous Actions	Sequential Actions	Extreme Group Outcomes	Extreme Users More Polarized
Brand perceptions	✓	✓	✓	✓	✓
Online reviews	✓	✓	✓	✓	
Corporate lobbying	✓	✓			✓
Diversity and inclusion	✓	✓		✓	
Faculty hiring	✓		✓		✓
Gun control	✓	✓		✓	

is not unique to the examples discussed previously. Indeed, it is prevalent in many sociopolitical discussions. Even a cursory reading of current news suggests that discussions of social and political issues show evidence of polarization (Cohn 2014). Second, the more extreme individuals often exhibit more polarized actions. For example, in the case of the corporate lobbying example, we saw that the firms with more extreme positions invested the most in lobbying. Table 1 presents an overview of the common features of setup and outcomes in the aforementioned examples and summarizes the main takeaways.

Research Agenda and Approach

Previous research in economics mainly attributes polarization to imperfect information aggregation and polarization of agents' beliefs in group settings (for details, see the "Related Literature" section). A parallel stream of literature in psychology attributes polarization of group decisions to behavioral biases stemming from social comparison and persuasive argumentation (Baron 2005; Zuber, Crott, and Werner 1992). Our idea is distinct from these prior theories based on either imperfect information aggregation or behavioral biases. We ask, "Can polarization of group decisions stem from strategic interactions between agents, even if the agents are rational and there is no imperfect information on some true state of the world?" Our analysis then links agents' preferences to the resulting polarization in their observed actions in group settings.

We propose a theory of group polarization with two related objectives. First, our theory connects the emergence of polarized group outcomes to individuals' strategic motives for group influence. We develop a model of group decision making, where agents have heterogeneous preferences over an issue. The basic analysis starts with a group of two agents. Each agent's utility function has two components: First, an agent incurs disutility if she chooses an action (or 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 her true preference. This can be interpreted as a reputational (or even a psychological) cost of misreporting her true preference. Second, an agent cares about the distance of the group's decision/outcome from her true preference. This represents the group influence motive: individuals would like to move the group's eventual outcome toward 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 decision or outcome, and all the agents' actions influence this outcome. Our model thus connects the trade-off between the desire for group influence and truth-telling to the polarization of actions at the individual and 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).

Next, we examine some key variables that can influence the existence and extent of polarization: group size, subgroup interactions, partial knowledge of the other agents' types, and the game structure. With respect to the last variable, we investigate the timing of actions or the order in which agents voice opinions. In a simultaneous game, each agent chooses an action without observing the actions of other group members. Alternatively, agents can express their opinions sequentially, in which case those who speak/act later can observe the actions of those who spoke before. Thus, the main difference between these two timings is the "observability" of others' actions. We examine whether group decisions are more polarized when agents speak simultaneously or when they speak sequentially. If individuals were to speak sequentially, who has a greater incentive to speak first—those with more extreme preferences or the moderates?

Questions pertaining to the timing and observability of actions are important because we see examples of both models in practice. Standard secret ballot models, where each agent submits their opinion without observing others' actions, can be interpreted as a simultaneous-actions model. For example, a department chair can survey everyone's opinion simultaneously (e.g., through online survey tools) and aggregate their opinions to make a decision. However, on a social media site, the opinions posted by previous members are visible to everyone and can affect the actions of later members. A recent study on Twitter finds that when individuals express their opinions in sequence, exposure to opposing views leads users to become

more entrenched (extreme) in their views when compared with their original positions (Bail et al. 2018).[4] Indeed, with the advent of online ballots and opinion-sharing forums, both timing formats are equally easy to design and implement. However, we do not have good answers to which of these models lead to more polarized decisions.

Further, we ask whether endogenizing the timing of actions by allowing agents to influence the speaking order affects polarization, and if so, how? We consider a game in which agents can participate in a first-stage auction to bid on the right to decide when they speak. The first-stage auction can be interpreted as agents lobbying with a principal (e.g., a department chair, a policy maker) to influence the speaking order. There is a long history in economics of modeling lobbying activities as auctions (Che and Gale 1998). Our endogenous choice model follows this tradition, and we capture agents' costs to influence the game rules through lobbying (Harstad and Svensson 2006; Potters and Van Winden 1992). Within this context, we examine which agents will have higher incentives to bid for the right to mandate the speaking order and when they will prefer to speak. Finally, we aim to compare social welfare under different game forms and derive the relationship between the extent of polarization in the system and overall welfare.

Results and Contribution

First, we show that in a simultaneous game, agents engage in strategic shading toward the extremes (i.e., take actions [voice opinions] that are more extreme than their true preferences). Moreover, agents with extreme preferences shade more than moderates because they expect the equilibrium outcome to be further away from their preference (as highlighted in the example on the electrical utility industry). Notably, an agent's incentive to shade and the extent of polarization in the group outcome are independent of the preference distribution. In other words, we show that polarized actions or behavior in group interactions do not necessarily stem from polarized preferences. Instead, they result from the strategic motivations of the agents that come into play in group settings. We also see that shading at the individual level causes the joint group decision to be more extreme (compared with the average preference of group members).

Extending the analysis to many players shows that the extent of shading goes down with group size, which suggests that group size can be a mechanism to control polarization. We also examine the interaction between subgroups in which agents within a subgroup have homogeneous preferences while there is heterogeneity across subgroups to show that smaller subgroups have the incentive to become even more extreme.

[4] On Twitter and other popular online opinion-sharing platforms, individuals with more extreme views often contribute more than those with more moderate views. In such cases, the observed distribution of opinions can be heavily skewed, with many extreme opinions and only a few moderate ones.

The analysis and comparison of simultaneous- and sequential-choice games establishes some of our main results. In an exogenous sequential speaking setting, polarization occurs if the agent who moves later is relatively more extreme than the first agent. In contrast, 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 first agent's observed action and pull the group decision closer to her preference. Indeed, this pattern is often visible in online forums, where agents who come later tend to express progressively more extreme opinions (Bail et al. 2018). Thus, given the group influence motive, the informational benefit of waiting and responding to others' actions is more attractive than moving first and setting the agenda. We then show that if agents' preferences are relatively similar, then the group decision is more moderate in the sequential actions game. In contrast, the simultaneous actions game leads to more moderate group outcomes when agents' preferences are dissimilar.

Next, we discuss the findings from the endogenous choice game, where agents participate in a first-price sealed-bid auction for the right to determine the speaking order. Here, we find that agents with more extreme preferences bid more for the right to decide the speaking order, and upon winning, all agents, regardless of their preferences, prefer to wait. More importantly, because the more extreme agents bid more and speak second, the group outcome is always polarized in this game format. However, the extent of polarization can be greater or lower compared with the simultaneous game. We find that when players' preferences are relatively similar, endogenizing the speaking order leads to less polarization compared with the simultaneous game. This implies that in settings where players are similarly inclined, endogenizing the speaking order can mitigate group polarization. Interestingly, we also find that the endogenous sequential game has the highest total welfare even though it can lead to more polarized decisions because it allocates the right to decide the speaking order more efficiently.

To summarize, our article makes three key contributions to the literature on group polarization. First, we propose a theory of strategic polarization based on users' incentive to influence the group decision. Our theory provides a rational explanation for the polarization of observed actions and is distinct from prior research that has focused on the polarization of beliefs arising either from imperfect updating or from behavioral biases. We show that the group influence motive can polarize users' actions even when their preferences are moderate. Second, we quantify the role of speaking order and observability of actions on polarization. We show that the observability of prior agents' actions mitigates polarization when agents' preferences are similar, but allowing agents to influence the speaking order always exacerbates polarization. Third, we identify levers that a group coordinator can employ to moderate polarization: group size, the speaking order, and the amount of knowledge that agents have about others. Finally, we show that game

formats that lead to more polarized outcomes do not necessarily lead to lower welfare.

Related Research

Research in psychology starting with Stoner (1961) shows evidence that group deliberation can make both individuals and the overall group decision extreme in the direction of their original proclivities. Polarization has been demonstrated in a variety of contexts including jury decisions (Main and Walker 1973), faculty evaluations and pay, attitudes toward women (Myers 1975), and judgments of attractiveness (Myers 1982), to name a few. This stream of work has described polarization using psychological explanations. In contrast, we identify a strategic rationale for group polarization that can accommodate and explain the different studies in this literature.

A more recent stream of literature focuses on providing different economic explanations for polarization of "beliefs" in group interactions. Dixit and Weibull (2007) analyze a model of Bayesian updating by agents with heterogeneous normally distributed priors about a true (policy) state and a common noise. In this setup, while the mean belief of the group may diverge under Bayesian updating after observing the common signal, individual-level polarization does not occur. Nevertheless, Baliga, Hanany, and Klibanoff (2013) show that polarization of individual beliefs can occur if individuals who observe a common signal are ambiguity averse. Similarly, Zimper and Ludwig (2009) consider a model of Bayesian learning with psychological bias in a setting where agents have ambiguous beliefs and show that this can lead to diverging posterior beliefs even if agents receive identical information. Acemoglu, Chernozhukov, and Yildiz (2009) consider a Bayesian learning problem for agents with different priors about the distribution of signals and show that even a tiny amount of signal uncertainty leads to significant disagreement in asymptotic beliefs.[5] More recently, Nielsen and Stewart (2020) show that polarization can occur in a Bayesian setting where two rational agents learn a finite amount of shared evidence. In contrast to this literature, our analysis is about the polarization of observed actions resulting from the strategic incentives of agents rather than opinions or beliefs. This allows us to establish a rationale for the polarization of group actions even in circumstances where preferences and beliefs of the agents are relatively moderate.

A parallel stream of research examines the role of behavioral biases or non-Bayesian updating on polarization. An early article by Rabin and Schrag (1999) formalizes a model of confirmatory bias where agents ignore signals that do not confirm with their initial impression and update in the direction of their current beliefs, generating polarization.[6] Bénabou (2012) investigates the emergence of collective denial in groups as agents form overoptimistic beliefs by ignoring negative signals. Glaeser and Sunstein (2009) analyze non-Bayesian behavior in which agents fail to account for the common sources of information of others' opinions. We complement this literature by identifying the role of a general group influence motive and how it interacts with the timing and the observability of actions in determining the extent of polarization. Whether the outcome of the group is more or less polarized depends on strength of the group influence motive as well as whether agents who move later are relatively more or less extreme.

There is also a related literature on strategic communication and cheap talk in persuasion games with multiple senders (experts) who try to influence a decision maker. Early research by Gilligan and Krehbiel (1989) and Austen-Smith (1990) model debates as cheap talk messages from multiple senders with different interests to show that such debates will only affect the outcome if the agents' preferences are not too dissimilar. Within this stream, Krishna and Morgan (2001) show that consulting two perfectly informed experts rather than one is beneficial when the experts are biased in opposite directions. A group of extremists does not have informational value in this framework. However, Bhattacharya and Mukherjee (2013) show that if the experts are uncertain about their information, then a decision maker may indeed prefer to hear from more extreme experts. Our article does not deal with strategic information transmission but, rather, with the strategic effect of the group influence motive in creating an incentive for polarized actions.

Our research is also related to the literature in marketing on group decision making that focuses on linking group behavior to that of individual members. Rao and Steckel (1991) develop an empirical model where group preferences are a weighted linear model of individual preferences. Their model attempts to account for observed group polarization in the data. Eliashberg and Winkler (1981) study group decision making and examine how uncertain group payoffs should be divided among the members in a Pareto optimal manner, given their risk attitudes and preferences. More broadly, our research also adds to the literature on social effects in marketing. A stream of empirical research documents the existence of social effects (for an overview, see Hartmann et al. [2008]; for a recent documentation of social effects using data from field experiments, see Sun, Zhang, and Zhu [2019]). A related stream considers the impact of these social effects on firms' strategies: For example, Amaldoss and Jain (2005) analyze competitive pricing strategies of conspicuous goods when consumers have preferences for uniqueness and conformity, and Yoganarasimhan (2012) analyzes a monopolist's advertising decisions in a market where consumers engage in social signaling. Similarly, Iyer and Soberman (2016) analyze the role of social comparison preferences in the context of socially responsible innovations.

[5] Other articles 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).

[6] This is related to research in social psychology, starting from Lord, Ross, and Lepper (1979), that provides experimental evidence that groups of individuals who hold differing opinions about sociopolitical issues use information in a biased manner by incorporating confirming evidence more readily than disconfirming evidence.

Model

We first present the basic model of group interactions, where the mean of actions of the agents is considered 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. In the context of our examples, this could be an agent's preferences for a brand or her true stance on the gun control. Both agents simultaneously choose a publicly observable action, $\{a_i, a_j\} \in \mathbb{R}$. In a group interaction, an agent's action can be interpreted as her voiced opinion and the action is continuous (e.g., writing a tweet about a brand's political position, voicing her opinion in a meeting on how the campus carry law should be implemented). After both agents have spoken, assume that a neutral third party or principal implements the mean of their voiced preferences or actions as the group outcome or decision.

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

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

where $\bar{a} = (a_i + a_j)/2$ and $r \in (0, 1)$ represents the relative weights that the agents places on the different components of their utility. In the case of the University of Texas working group on the campus carry issue, $\bar{a}$ would denote the recommendation that the committee chair makes to the university administrators based on the summary of the average group opinions. In the case of online reviews, $\bar{a}$ denotes the average rating of the product/restaurant.

Agents obtain disutility 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 (cost of lying) or a reduced form representation of credibility of actions arising from potential reputational concerns.[8] Second, their utility is decreasing in the distance between the group outcome ($\bar{a}$) and their true preference. For example, in the Google recruiting case, some stakeholders likely felt unhappy about the decision to make all recruiting within certain categories diversity driven.

Our main results are not dependent on this assumption of the mean as the group outcome. They hold qualitatively for any decision rule that uses a linear combination of agents' preferences and puts nonzero weights on the actions of all players (for details, see Web Appendix A.1).[9] What is necessary for our results to qualitatively hold is that the outcome measure is a function of the actions of all the agents in the group. In other words, agents have a taste for influencing the group's outcome. A greater value of r represents issues for which agents have stronger relative preference for voicing opinions that are consistent with their true preferences. Overall, the agent's utility function displays the single-peakedness property, where each agent has an ideal point (in this case, x_i). Actions and outcomes away from this point are less than ideal and are strictly monotonically decreasing in both directions.

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 on her private information as well as the observed actions of the first mover.

Benchmark Cases

Before we analyze the private information game, it is useful to derive two benchmark cases: (1) the first-best socially optimal solution and (2) the perfect information case. In the first case, a social planner chooses actions to maximize the joint surplus of the two agents:

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

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. Next, suppose 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, the group's joint decision is not distorted.

Simultaneous Actions

Equilibrium

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

[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 $\{a_i, a_j\} \in \mathbb{R}$).

[8] One could consider a model of reputation in which any misreporting today has consequences for the future. The first term in the utility function can be seen as the reduced-form equivalent of this model. This reduced-form representation captures the credibility of actions; thus, when r approaches 1, the players report their true preferences.

[9] The idea of using a linear combination of individual preferences to represent the group preferences/outcomes has a long history in consumer research and marketing and has been shown to have significant empirical validity (Corfman and Lehmann 1987; Elisahberg et al. 1986; Rao and Steckel 1991).

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 is $EU(x_i, a_i) = \frac{\int_{\mathbb{R}} u(x_i, a_i, \hat{a}_j) g(x_j) dx_j}{\int_{\mathbb{R}} g(x_j) dx_j}$.

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

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

In obtaining the first-order condition, we can set $d\hat{a}_j/da_i = 0$ 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

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

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

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

Because $\int_{\mathbb{R}} \hat{a}_i g(x_i) dx_i = \int_{\mathbb{R}} \hat{a}_j g(x_j) dx_j$, and because $E(x) = 0$ for a symmetric distribution, we can uniquely identify $\int_{\mathbb{R}} \hat{a}_j g(x_j)_j dx_j = 0$. We thus have $\hat{a}_i = \frac{2(1+r)}{1+3r} x_i$. These results are summarized in Proposition 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 Proposition 1 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, whereas those with negative x_i always move 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 $EU(x_i, x_i + \varepsilon) = -r\varepsilon^2 - \frac{1-r}{4}(x_i - \varepsilon)^2 - \frac{\mu(r)^2}{4} E(x^2)$. For small values of ε, $EU(x_i, x_i + \varepsilon) > EU(x_i, x_i)$ and the converse is true for ε large enough. Thus, 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.[10]

Next, recall that group polarization is defined as the tendency of the joint outcome to move toward a more extreme point in the direction indicated by the members' original preferences. The equilibrium derived previously satisfies this definition. The mean predeliberation preference of the group is $\bar{x} = \frac{x_i + x_j}{2}$ while the mean postdeliberation 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 $\bar{a} < \bar{x}$. Thus, if the preferences of the two agents in the group is initially predisposed toward the right, the group decision is even more rightward. Alternatively, if the group is predisposed toward the left, then its joint decision is even more leftward.

Comparative Statics

To investigate the comparative statics, we denote the extent to which an agent i shades her opinion in equilibrium as $s_i = |\hat{a}_i - x_i| = \frac{1-r}{1+3r}|x_i|$. Figure 1 depicts an example of the equilibrium actions of agents whose preferences are drawn from $U[-1, 1]$ and their shading as a function of their preference x_i, for $r = .1$. We can see that $ds_i/dr \leq 0$, and as would be expected, agents shade their actions less if the cost associated with lying is higher. Second, $ds_i/(d|x_i|) > 0$ suggests that agents near the extremes shade more than moderates, who are closer to the center. For example, as discussed previously, consumers who have strong liberal or conservative political preferences are the ones who are more vocal in their Twitter activity about the advertising campaigns of brands. Similarly, this also

[10] We have not explicitly included abstention as part of the players' strategy set. Abstention can be seen as equivalent to not voicing any opinion. This analysis 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 $-(1-r)\left[x_i - (x_i + a_j)/2\right]^2$.) Thus, abstaining is always a dominated choice.

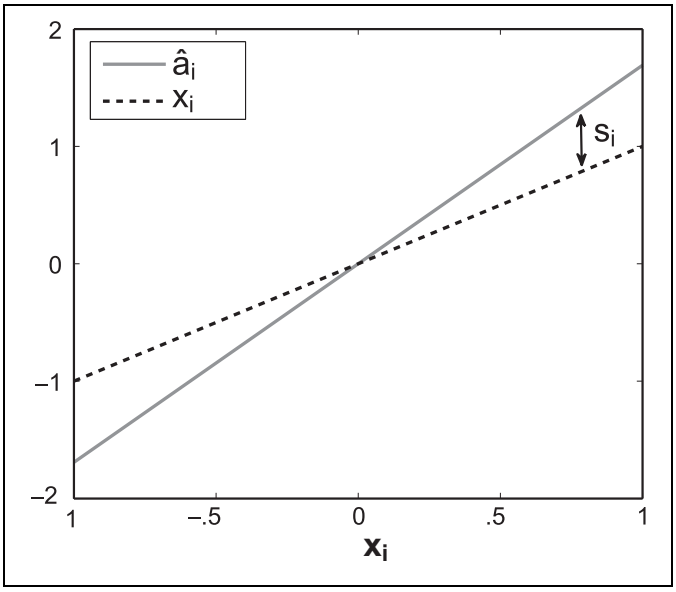

Figure 1. Equilibrium actions and shading in a two-player simultaneous game with r = .1.

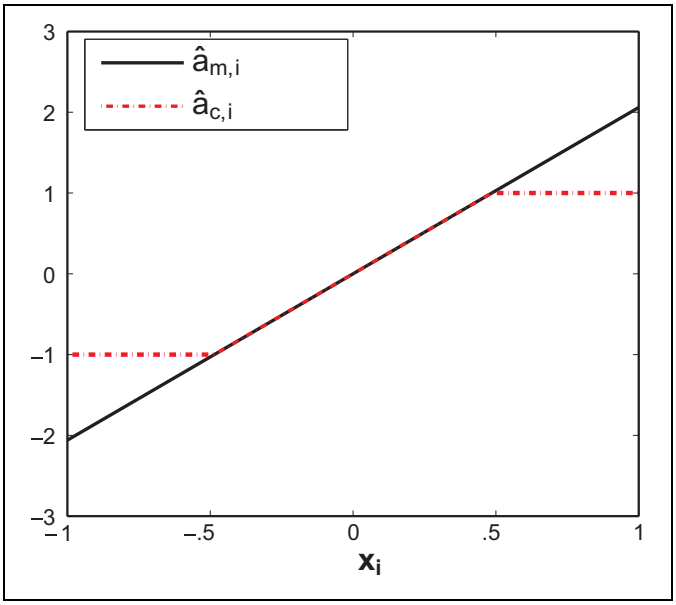

Figure 2. Equilibrium actions in unconstrained and constrained m-player simultaneous-choice games with r = .1 and m = 5.

provides a rationale for the U-shaped relationship described in Delmas, Lim, and Nairn-Birch (2016) between greenhouse emissions and lobbying intensity: the more extreme companies with the least and most carbon emissions are also the more extreme voices in the climate debate.

The result 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 toward the extremes in their decisions, this effect is exacerbated in extreme groups.

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 distribution 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 to influence group decisions.

Extensions and Robustness

Next, we consider several extensions of the main model to illustrate the robustness of the results and to develop a comprehensive theoretical account by examining how polarization is influenced by the game structure, group size and characteristics, preference characteristics, and the informational endowment of the agents.

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

In the Bayesian Nash equilibrium of this game, agent i chooses action $\hat{a}_{m,i} = \frac{rm^2+(1-r)m}{rm^2+(1-r)}x_i$. 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 $\mu_m(r) = \frac{rm^2+(1-r)m}{rm^2+(1-r)}$. The extent of shading by agent i is $s_{m,i} = |\hat{a}_{m,i} - x_i| = \frac{(1-r)(m-1)}{rm^2+(1-r)}|x_i|$, 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 $1/\sqrt{r} + 1$ but decreasing after that (see Web Appendix A.2.2). 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 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. Thus, 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 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 some cases and may simply be infeasible in others. For example, faculty groups in many schools are small. Therefore, we subsequently analyze

the role of timing of the actions and ask whether sequential choices may be a potential solution to the polarization problem.

Subgroup interactions. In many situations, interactions occur between subgroups, where each subgroup consists of many agents having the same preferences over an issue, but different from that of the other subgroup. For example, academic marketing departments consist of quantitative and behavioral subgroups, political deliberations in the United States (e.g., in the Senate) occur between multi-agent Democratic and Republican subgroups. On issues such as abortion rights, gun control, and taxation, conservatives groups have different preferences than liberals, but citizens within each subgroup tend to 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 subgroup sizes are common knowledge. The preferences x_1 and x_2 of the subgroups are independently drawn from $g(x)$. All agents within a subgroup know their individual preferences (and that of the others within their subgroup), but they do not observe the preferences of the other subgroup. We can write the expected utility of an individual agent i from subgroup 1 and agent j from subgroup 2 as

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

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

In Web Appendix A.3, we present the solution for the symmetric (for agents within a subgroup) 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 subgroup's action is linear in its preferences, and subgroups near the extremes shade more.

The main point of this analysis is to understand the role of the subgroup size on actions. Specifically, for a given population size m, how would a subgroup's size affect actions? It can be seen that for a given m, $\partial\hat{a}_i/\partial n_i$ and $\partial\hat{a}_j/\partial n_j$ are both negative. The implication is that smaller subgroups can become even more extreme. For example, in the academic hiring scenario, the department chair may expect to see more extreme opinions and actions if one of the subgroups is smaller than the other. 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, Murray, and Dann 2014). This 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).

Constrained-choice model. So far, we have allowed agents to choose actions beyond the range of types. However, this may not always be feasible. For example, if the public action is supposed to be a revelation of private type, then it is impossible for agents to proffer a type that does not exist. Thus, we now consider a model where agents' actions are bounded within a credible range. This model can be interpreted as a setting where agents' actions are truncated or discounted if they are too extreme. In particular, voicing opinions outside the range of agent types can be inferred as lacking in credibility (or lying about one's type) and therefore discounted.

Specifically, consider a scenario in which agents' preferences are drawn from a uniform distribution bounded at -1 and 1 (i.e., $U[-1, 1]$), and they are required to choose an action that lies between $[-1, 1]$.[11] As in the previous models, agents simultaneously choose actions, and the mean of their actions is implemented as the group's decision. That said, this analysis is a reduced-form way to consider credibility, and alternatively one could consider a dynamic model of reputational concerns with polarization, which would affect the credibility of actions.

In Web Appendix A.4, we show that in an m-player game constrained-choice model, the optimal response function continues to be $\frac{rm^2+(1-r)m}{rm^2+(1-r)} x_i$ until it hits the bounds (i.e., 1 or -1) and then is constrained to be at the bounds of the distribution. For a pictorial depiction of the equilibrium actions in this bounded setting (denoted by $\hat{a}_{c,i}$), see Figure 1.

In summary, expanding the number of players, constraining the choice set, or considering subgroups does not affect the key results. Thus, moving forward, we retain the two-player, unconstrained-choice model and focus our attention on other interesting modifications such as information revelation and sequential choices.

Asymmetric type distribution. Next, we consider a situation in which the distribution of types, $g(x)$, is asymmetric (i.e., $E(x) \neq 0$). In Web Appendix A.5, we show that for any general asymmetric distribution $g(x)$, 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 that $x_i>0$ 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

[11] The results translate directly to any symmetric bounded preference distribution. We use $U[-1, 1]$ mainly to illustrate the actions pictorially in Figure 2.

direction as the agent's preference, this works against the tendency to be extreme and may even lead to moderation.

Partial knowledge. In many instances, players may have some knowledge about the preferences of their rivals. This may especially be the case in smaller groups such as faculty groups or corporate teams, in which members have a history of prior interactions. For example, suppose that each agent i knows whether the other agent j is in $\mathbb{R}_+$ or $\mathbb{R}_-$ but actual locations of (x_i, x_j) are still private information for the respective players. Thus, each agent knows which side the other "leans" toward, but not their exact preference.

Given the setup, there are two possible cases of partial knowledge: First, the case in which each agent knows that the other player's preference is on the opposite side (i.e., opposite leaning). The alternative case is one in which each agent knows that the other is on the same side (i.e., similar leaning). The equilibrium analysis is in Web Appendix A.6.

Opposite-leaning agents: Suppose, without loss of generality, that $x_i \in \mathbb{R}_+$ and $x_j \in \mathbb{R}_-$. In other words, agent i is known to be left leaning and agent j to be right leaning. The equilibrium actions in this case turn out to be

$$\hat{a}_i = \frac{2(1+r)}{1+3r}x_i - \frac{(1-r^2)}{2r(1+3r)}\int_{\mathbb{R}_-} x_j g(x_j)dx_j, \quad (7)$$

$$\hat{a}_j = \frac{2(1+r)}{1+3r}x_j - \frac{(1-r^2)}{2r(1+3r)}\int_{\mathbb{R}_+} x_i g(x_i)dx_i. \quad (8)$$

Recall from Proposition 1 that in the full private information case, each agent's actions are a function of only their own preferences. With partial knowledge, the equilibrium actions of agent i are a response not only to her own preference x_i but also to $E(x_j) = \int_{\mathbb{R}_-} x_j g(x_j)dx_j$ conditional on the knowledge that the other agent is opposite leaning. Thus, each agent's actions are now also a function of the knowledge they possess about their rival.

Note that $\frac{2(1+r)}{1+3r} > 1$, which means that with opposite-leaning agents, the actions are still more extreme in response to the own preference. Further, the knowledge that x_j is opposite leaning also adds to the polarization of a_i. Note that $\frac{1-r^2}{2r(1+3r)} < \frac{2(1+r)}{1+3r}$, as long as $r > 1/5$. Thus, an agent responds more to her own private type than to partial information about the rival as long as the truth-telling motive is strong enough. Finally, we have $\bar{a} = \frac{1+r}{1+3r}(x_i + x_j)$; so, similar to Proposition 1, $\bar{x} > 0$, $\bar{a} > \bar{x}$ and $\bar{x} < 0$, $\bar{a} < \bar{x}$. Thus, the extent of polarization in the group outcome remains the same as in the main model.

Similar-leaning agents: Next, consider the alternative case in which both agents know that they are on the same side of zero (i.e., they are similar leaning). Without loss of generality, let $(x_i, x_j) \in R_+$. As derived in the Web Appendix, the equilibrium actions are

$$\hat{a}_i = \frac{2(1+r)}{1+3r}x_i - \frac{(1-r)}{(1+3r)}E(x_j), \quad (9)$$

$$\hat{a}_j = \frac{2(1+r)}{1+3r}x_j - \frac{(1-r)}{(1+3r)}E(x_i), \quad (10)$$

where $E(x) = \int_{\mathbb{R}_+} x g(x)dx$. When the group members are similar leaning and on the same side of zero, the presence of partial knowledge can have a moderating influence, and the equilibrium actions will be less extreme. Indeed, when a player's preference is sufficiently small when compared with the expected value of the preference distribution (e.g., x_i is small enough compared with $E(x)$), it is even possible that the player will choose a moderating action away from her direction of preference and toward zero. We can also note that the partial knowledge of the rival's preference (which is expressed through the effect of the expected value of the preference $E(x)$) has a greater effect on actions when agents have opposite (vs. similar) preference leanings. Finally, the mean of the group actions is $\bar{a} = \frac{2(1+r)}{1+3r}(\bar{x}) - \frac{(1-r)}{1+3r}E(x)$. Thus, the group outcome may be more extreme than the mean realized preferences, but it is also moderated by partial knowledge to the extent of the expected preference.

Thus, in general, partial knowledge induces agents to consider the available information about the other player, and this can be a force for moderation of group actions when agents end up being similar leaning. More generally, suppose that each agent has a noisy (but better than the prior) information signal about the private information of the other agent. Then, as the precision of the signal improves, the other player's preferences will have a greater effect in moderating actions. At the extreme, with perfect information, we will get the aforementioned benchmark case in which agents have full information. The availability of information about group members' private preferences can thus be used as a strategic instrument by the principal to moderate group behavior.

Incentive to disclose preferences. If agents had the opportunity to verifiably communicate their preference, would they have the incentive to do so? Communication that makes preferences common knowledge has the potential to reduce polarization in actions. Accordingly, consider an ex ante disclosure game in which each agent i simultaneously chooses whether to reveal the private information about x_i prior to the agents choosing their public actions a_i. In Web Appendix A.7, we solve for the equilibrium of the disclosure game and show that the scenario where both agents choose not to disclose their type is an equilibrium. Thus, our basic model with private information emerges as an equilibrium even when players can communicate their preferences.

Alternative preferences. To have a better perspective of the role of the group influence motive in our model, we compare it with some alternative preference formulations to understand what types of preferences may counter the polarization of actions.

Consider first an alternative formulation in which each agent's social preference is to minimize the distance between their actions and the true mean preference of the group. This

can be seen as a taste for conformity with group preferences. Might this help reduce the extent of polarization in actions? Specifically, suppose that agent i minimizes the distance between a_i and the average of the group's true preference $\bar{x}$. Thus, the alternative utility function for, say, i would be $U(x_i, a_i, a_j) = -r(x_i - a_i)^2 - (1-r)(a_i - \bar{x})^2$ and equivalently for j. The equilibrium actions, as derived in Web Appendix A.8, are $\hat{a}_i = \frac{(1+r)}{2}x_i$ and $\hat{a}_j = \frac{(1+r)}{2}x_j$. These equilibrium actions show moderation, in that they are closer to the center than the agent's true preferences. This is natural because i has an incentive to move her action closer to $\frac{x_i + E(x_j)}{2}$. Because $E(x) = 0$, this naturally leads to the agent moving closer to zero (i.e., moderating her action). Further, the equilibrium outcome is $\bar{a} = \frac{(1+r)}{4}(x_i + x_j)$ and $|\bar{a}| < |\bar{x}|$. Thus, the group outcome is more moderate than the average preferences.

As another alternative formulation, consider the case when agents care about minimizing the distance between their public actions and the group's outcome. In other words, the agent cares about how closely their actions or voiced opinions conform to the mean outcome of the group. Specifically, suppose that the agent i's utility function was $U(x_i, a_i, a_j) = -r(x_i - a_i)^2 - (1-r)(a_i - \bar{a})^2$, and equivalently for j. We derive the equilibrium in the Web Appendix and can show that for a symmetric distribution of agent types, the equilibrium actions are $\hat{a}_i = \frac{4r}{1+3r}x_i$. This implies that $|\hat{a}_i| < |x_i|$; consequently, the agents moderate their actions to be closer to center, and the group outcome is also moderate compared with the mean of preferences (i.e., $|\bar{a}| < |\bar{x}|$). Overall, if people care about conforming with their peers' actions, we get moderation in actions and group decisions, similar to the main findings in Bernheim (1994).

Together, these two extensions with alternative preferences highlight the importance and role of the group influence motive (i.e., the agent's desire to move the group's outcome closer to her true preferences) in driving polarization. Finally, note that we have assumed that the weight on the truth-telling (r) is independent of agents' preferences. However, it is possible that weights differ between agents with more versus less extreme preferences. For example, it is possible that more extreme agents place a relatively greater weight on the group influence motive rather than on truth-telling. As long as the function $r(x_i)$ is symmetric around zero, the main insights of the analysis will continue to hold.

Sequential Actions

Next, we consider the case in which players may voice their opinions in sequence. On Yelp, consumers see past reviews while providing their ratings. Similarly, in social media groups, individuals may express opinions in sequence. In a department meeting, the chair may mandate the order in which different faculty members may speak. Indeed, in many institutional settings, members typically take turns to speak. In Federal Open Market Committee meetings, committee members express their preferred policy position sequentially in an order that varies across meetings. The committee chair summarizes these positions into an overall group directive on the federal short-term interest rates. Similarly, in juries and legislative bodies, the order of speaking is often predetermined 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 upon observing the action taken by the first. 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 determined by either institutional rules or a third party. Then, we consider the case in which the agents bid to endogenously choose the speaking order.

Equilibrium in the Sequential Game

Let $a_{xt,i}$ denote agent i's action in period t in this exogenous 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 who 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 $a_{x1,i}(x_i, \hat{a}_{x2,j})$ 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 $a_{x2,j}(x_j, a_{x1,i})$ 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 $a_{x2,j}$ in response to the first player's observed action $a_{x1,i}$ is $u(x_j, a_{x2,j}, a_{x1,i}) = -r(x_j - a_{x2,j})^2 - (1-r)(x_j - \bar{a}_x)^2$, where $\bar{a}_x = \frac{a_{x1,i} + a_{x2,j}}{2}$. The optimal choice of agent j given the first period choice of i can be derived as $\hat{a}_{x2,j} = \frac{2(1+r)}{1+3r}x_j - \frac{(1-r)}{1+3r}a_{x1,i}$.

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

$$EU_{x1}(x_i, a_{x1,i}) = -r(x_i - a_{x1,i})^2 - (1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{x1,i}\right)^2 + \left(\frac{1+r}{1+3r}\right)^2 \int_{\mathbb{R}} x_j^2 g(x_j) dx_j\right]. \quad (11)$$

Taking the first-order condition of Equation 11 and following a similar analysis to that in the simultaneous equilibrium

case gives us the equilibrium action of i as $\hat{a}_{x1,i} = \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i$. Therefore, the first player i has a unique optimal response that is both linear in her type x_i and is symmetric around zero.

Characterizing the Group Outcome

Having derived the individual equilibrium actions, we proceed to characterize the mean equilibrium outcome. For a given x_i and x_j, the mean equilibrium outcome is $\bar{a}_x = \frac{\hat{a}_{x1,i}+\hat{a}_{x2,j}}{2} = \frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i + \frac{1+r}{1+3r}x_j$. Before stating the results, we define the following relationships:

- *Polarization*: $|\bar{a}_x| > |\bar{x}|$ and $\bar{a}_x\bar{x} > 0$. Polarization is said to have occurred if the mean of equilibrium actions ($\bar{a}_x$) 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*: $|\bar{a}_x| > |\bar{x}|$ and $\bar{a}_x\bar{x} \leq 0$. Reverse polarization is the case where the mean equilibrium outcome ($\bar{a}_x$) 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*: $|\bar{a}_x| \leq |\bar{x}|$. Moderation refers to the case where the mean of the equilibrium actions ($\bar{a}_x$) 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 $k_1(r) = \frac{(1+3r)(1-r)}{(1+3r)^2+4r(1-r)}$ and $k_2(r) = \frac{(1+3r)\left[(1+3r)^2+16r\right]}{\left[(1+3r)^2+4r(1-r)\right][2(1+r)+(1+3r)]}$, and without loss of generality, let $x_i \geq 0$. Comparison of the mean equilibrium outcome ($\bar{a}_x$) with the mean of preferences ($\bar{x}$):

- Polarization occurs if $x_j > k_1(r)x_i$ or $x_j < -x_i$.
- Reverse polarization occurs if $-x_i \leq x_j < -k_2(r)x_i$.
- Moderation occurs if $-k_2(r)x_i \leq x_j \leq -k_1(r)x_i$

Proof: See the Web Appendix. □

Figure 3 summarizes the effect of sequential actions on group polarization. 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 $\bar{a}_x$ 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 $-x_i \leq x_j \leq -k_2(r)x_i$, 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 $\bar{a}_x$ 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.

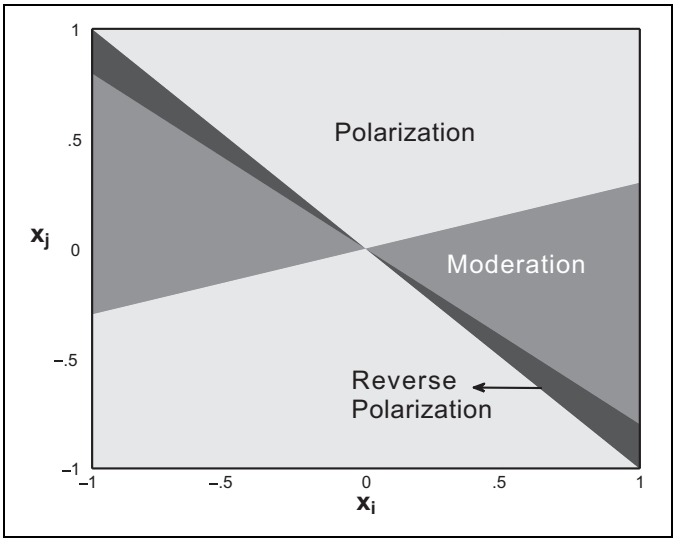

Figure 3. Regions of polarization, reverse polarization, and moderation in an exogenous-sequential-choice game; shown for $\{x_i, x_j\}$ drawn from $U[-1, 1]$.

In contrast, when the second mover j's preference is closer to zero compared with 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 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 as well.

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 $\{\hat{a}_{x1,i}, \hat{a}_{x2,j}$ 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) $\hat{a}_{x2,j} \geq \hat{a}_j$ if $x_i \leq 0$ and $\hat{a}_{x2,j} < \hat{a}_j$ if $x_i > 0$.
(b) $|\hat{a}_{x1,i}| \geq |\hat{a}_i|$ and $\frac{d|\hat{a}_{x1,i}|}{dr} < 0$.

Proof: See the Web Appendix. □

Consider first the action of the second player j in the sequential game $\hat{a}_{x2,j} = \frac{2(1+r)}{1+3r}x_j - \frac{(1-r)}{1+3r}\hat{a}_{x1,i}$. Part (a) of Proposition 3 shows that if the players lie on opposite sides of zero, then j becomes more extreme in the sequential actions game as compared with the simultaneous game. In contrast, when the two players' preferences are 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 compared with 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., $|\hat{a}_{x1,i}| \geq |\hat{a}_i|$) 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.

Proposition 4: Compare the mean equilibrium outcome in the exogenous sequential game $\bar{a}_x$ with that from the simultaneous game ($\bar{a}$). Let $k_3(r) = \frac{(1+3r)^3+8r(1-r)(1+r)}{2(1+r)\left[(1+3r)^2+4r(1-r)\right]}$. Then,

- $|\bar{a}_x| > |\bar{a}|$ and $\bar{a}_x\bar{x} > 0$ 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 $\bar{x}$).
- $|\bar{a}_x| > |\bar{a}|$ and $\bar{a}_x\bar{x} \leq 0$ 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 $\bar{x}$).
- $|\bar{a}_x| \leq |\bar{a}|$ if $x_j \geq -k_3(r)x_i$ (i.e., the mean outcome in the exogenous sequential game is moderate compared with the mean outcome in the simultaneous game).

Proof: See the Web Appendix. □

If i and j are relatively similar, then the mean outcome is more moderate in the sequential game (i.e., $|\bar{a}_x| < |\bar{a}|$). However, 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 4). Overall, the propositions 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 may lead to less polarization of actions. These results highlight how the timing of the game may be exploited to combat polarization. For example, a planner or a group coordinator 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.

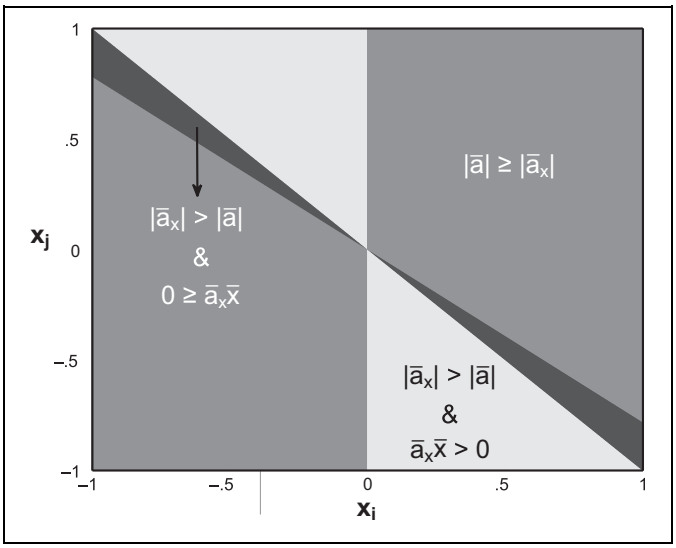

Figure 4. Comparison of mean outcome in the exogenous sequential game with that in the simultaneous game; shown for $\{x_i, x_j\}$ drawn from $U[-1, 1]$.

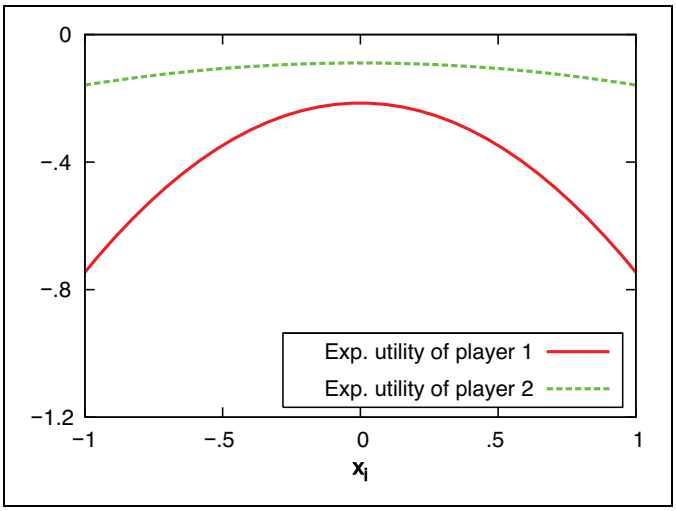

Figure 5. Players' expected utilities in the exogenous-sequential-choice game; shown for $\{x_i, x_j\}$ drawn from $U[-1, 1]$ for $r = .1$.

Value of the Speaking Order

Next, we 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 $a_{x,i}$ is given by Equation 11. In equilibrium, i optimally chooses action $\hat{a}_{x1,i}$. Substituting this into Equation 11 gives us the following:

$$EU_{x1}(x_i, \hat{a}_{x1,i}) = -\frac{(1-r)(1+r)^2}{(1+3r)^2 + 4r(1-r)} x_i^2 - \frac{(1-r)(1+r)^2}{(1+3r)^2} \int_{\mathbb{R}} x^2 g(x) dx. \quad (12)$$

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

$$EU_{x2}(x_i, \hat{a}_{x2,i}) = \frac{\int_{\mathbb{R}} u(x_i, \hat{a}_{x2,i}, \hat{a}_{x1,j}) g(x_j) dx_j}{\int_{\mathbb{R}} g(x_j) dx_j},$$

$$= -\frac{r(1-r)}{1+3r} x_i^2 - \frac{r(1-r)(1+3r)(3+r)^2}{\left[(1+3r)^2 + 4r(1-r)\right]^2} \int_{\mathbb{R}} x^2 g(x) dx. \quad (13)$$

See Figure 5 for a pictorial illustration of these expected utilities. 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 $D_x(x_i) = EU_{x1}(x_i, \hat{a}_{x1,i}) - EU_{x2}(x_i, \hat{a}_{x2,i})$ denote the difference between the equilibrium expected utilities of agent i from speaking in the first and second periods.

(a) $D_x(x_i) \leq 0 \Rightarrow$ 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) $\frac{dD_x(x_i)}{d|x_i|} \leq 0 \Rightarrow$ the difference in expected utilities of speaking in the first and second periods is increasing in $|x_i|$.

Proof: See the Web Appendix. □

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 the first player's action was close to 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 that 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 adjust her action on the basis of 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 toward their true preference, they prefer 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 that 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 summary, the analysis suggests that there are inherent advantages to waiting, especially for agents with more extreme preferences.

Endogenous Sequential Actions: Bidding to Speak

As described previously, 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 that 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 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. This first-stage auction can also be seen as agents lobbying to influence 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 on 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 $\beta(x_i) = f(r)\dfrac{\int_0^{x_i} x^2 g(x)dx}{\int_0^{x_i} g(x)dx}$ and chooses to speak second if she wins the auction. The function $f(r)$ is defined in Web Appendix F and has the property that $f(r)>0\ \forall\ r>0$.
- If agent i speaks first, then she chooses $\hat{a}_{n1,i} = \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i$, whereas if she speaks second, she chooses $\hat{a}_{n2,i} = \frac{2(1+r)}{1+3r}x_i - \frac{(1-r)}{1+3r}a_{n1,j}$.

Proof: See Web Appendix F. □

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 $a_{n2,i}$ in response to the first player's action $a_{n1,j}$. The incentives facing the first player are more subtle. If she is speaking first, this could be either 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. Thus, i's expected utility from speaking first is obtained by taking the expectation of $u(x_i, a_{n1,i}, \hat{a}_{n2,j})$ over $x_j \in W$ (i.e., $EU_{n1}(x_i, a_{n1,i}) = \dfrac{\int_W u(x_i, a_{n1,i}, \hat{a}_{n2,j})g(x_j)dx_j}{\int_W g(x_j)dx_j}$). This can be simplified by substituting for $\hat{a}_{n2,i}$:

$$EU_{n1}(x_i, a_{n1,i}) = -r(x_i - a_{n1,i})^2$$

$$-(1-r)\left[\left(x_i - \frac{2r}{1+3r}a_{n1,i}\right)^2 + \left(\frac{1+r}{1+3r}\right)^2 \frac{\int_W x_j^2 g(x_j)dx_j}{\int_W g(x_j)dx_j}\right]$$

$$-2\frac{1+r}{1+3r}\left(x_i - \frac{2r}{1+3r}a_{n1,i}\right)\frac{\int_W x_j g(x_j)dx_j}{\int_W g(x_j)dx_j}. \qquad (14)$$

The last term vanishes because W is symmetric around zero. By setting $\frac{dEU_{n1}(x_i,a_{n1,i})}{da_{n1,i}}\big|_{a_{n1,i}=\hat{a}_{n1,i}} = 0$, we can solve for $\hat{a}_{n1,i} = \frac{(1+3r)(3+r)}{(1+3r)^2+4r(1-r)}x_i$, 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 player's 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 Web Appendix to be $\beta(x_i) = f(r)\dfrac{\int_0^{x_i} x^2 g(x)dx}{\int_0^{x_i} g(x)dx}$, and it is monotonically increasing in $|x_i|$. Thus, agents with more extreme preferences place higher value on 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 $[\bar{a}_n]$ closer to their own preference increases [as r decreases], their bid increases). At $r = 0$, the bidding strategy simplifies to $\beta(x_i)|_{r=0} = 2\dfrac{\int_0^{x_i} x^2 g(x)dx}{\int_0^{x_i} g(x)dx}$, which is the highest, whereas at $r = 1$, the bidding strategy devolves to $\beta(x_i)|_{r=1} = 0$.

Consider 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 $\bar{a}_n = \frac{\hat{a}_{n1,i}+\hat{a}_{n2,j}}{2} = \frac{2r(3+r)}{(1+3r)^2+4r(1-r)}x_i + \frac{1+r}{1+3r}x_j\ \forall\ |x_i|<|x_j|$. Proposition 7 compares the mean outcome, $\bar{a}_n$, 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 ($|\bar{a}_n|>|\bar{x}|$ and $\bar{a}_n\bar{x}>0$). Comparison of the extent of polarization across the different games yields the following:

- If $x_i \times x_j<0$, then $|\bar{a}_n|>|\bar{a}|$.
- If $x_i \times x_j>0$, then $|\bar{a}_n| \le |\bar{a}|$.

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 agents' actions and the group outcome are always polarized.

Thus, 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 the endogenous sequential actions game produces more polarization than the simultaneous actions game. On the other hand, if both players lie on the same side of zero, 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: that is, 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 toward 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 the 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 reduce polarization.

Welfare Comparisons

We start with the planner's problem to understand how a principal would design the group interaction to maximize social welfare. 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 because they remain within the system and thus 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., $EW = \frac{\int_{x_i}\int_{x_j} W(x_i,x_j)g(x_i)g(x_j)dx_i dx_j}{\int_{x_i}\int_{x_j} g(x_i)g(x_j)dx_i dx_j}$). 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.

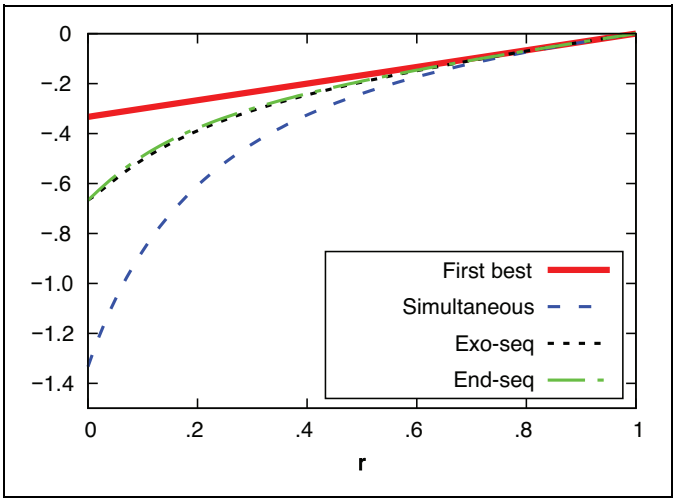

Figure 6. Expected social welfare for the four game forms; shown for $U[-1, 1]$.

Figure 6 shows the relationship between expected welfare functions as a function of r for the case of $U[-1, 1]$. It can be seen that $EW_n > EW_x > EW_s$. For a detailed derivation of this inequality, see the Web Appendix. Within the sequential action formats, allowing agents to endogenously bid for the speaking order increases expected welfare as compared with 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, the expected welfare under the sequential game (irrespective of whether it is exogenous or endogenous) is higher; even though the endogenous-sequential-choice game produces higher polarization, it also increases the welfare by allowing those with more extreme preferences to obtain the outcomes they desire.

Conclusion, Marketing Implications, and Future Research

Many formal and informal forums in business, organizational, and sociopolitical settings facilitate group interactions that shape our views and decisions on issues ranging from brand choices to faculty hiring, diversity, and gun control. We might expect such interactions to help users to exchange information and align their opinions. However, even a cursory glance at the current sociopolitical landscape in the United States suggests otherwise. Indeed, group deliberations seem to increase polarization rather than reducing it. Although we show that polarization can lead to higher social welfare, it can also create conflict and lead to other social problems (Esteban and Schneider 2008). Policy makers may therefore put positive weight on both reducing polarization and increasing welfare. Because these two objectives are not necessarily aligned, it is critical to pin down the mechanisms that make actions and opinions more polarized.

We develop a theory that links the polarization of agents' actions to a fundamental trade-off agents face between influencing others in a deliberation and expressing their true preferences.

When agents' actions affect the group's outcome, deliberations can lead to polarization. Further, we show that the more extreme agents end up becoming more polarized in their actions. Next, we analyze and compare the role of two types of speaking orders: simultaneous versus exogenous sequential. In sequential-choice settings, polarization occurs when the agent who moves later is more extreme than the first mover. In this context, we also highlight the trade-off between the commitment value of moving first versus the value of the flexibility (to respond to other user's actions) that comes from moving second. We find that the group influence incentive makes flexibility valuable and induces agents to wait. Further, we see that if agents' preferences are dissimilar, the sequential actions game produces less polarization compared with the simultaneous game (whereas the opposite is true if agents have similar preferences).

Next, we endogenize speaking order by allowing agents to bid for the right to choose the speaking order. We show that the agent with more extreme preferences bids more, and the winning agent always chooses to speak later. In addition, we examine the role of the group size and the presence of subgroups on polarization. We find that larger groups show less polarization, and smaller subgroups tend to go to extremes. Finally, we investigate alternative preference distributions and information structure to expand and clarify the role of the group influence motive in causing polarization.

Our findings have important implications for the marketing examples discussed previously. For example, Nike's "Believe in Something" ad campaign featuring Colin Kaepernick evoked highly polarized reactions, with younger (18–34 years) individuals strongly approving the ad and older individuals disapproving. Similar reaction disparities were seen across racial and political affiliations. These results are consistent with our model, where we find that users express opinions that are more extreme in the direction of their original preference in public discourses. In the branding context, this suggests that polarization can be a mechanism to increase brand differentiation, as articulated by Phil Knight: "It does not matter how many people hate your brand as long as enough people love it" (Stoll 2019). An implication for brand strategy is that firms can design advertising strategies that take a stand on important social, political, or environmental issues prevalent in society to create strong brand differentiation in competitive markets.

Our analysis is also relevant to the design of review systems in recommendation platforms. Many e-commerce platforms (e.g., Amazon, Tripadvisor, Yelp) display the mean ratings received by a product/seller on their websites. Prior research has suggested that consumers pay attention to these aggregate ratings and that this affects platform demand and revenues (Chevalier and Mayzlin 2006). However, if later reviewers react to earlier reviews and bias their ratings, then the aggregate rating measure can become biased (i.e., no longer representing the mean of consumer preferences). In turn, this can have adverse consequences for consumers' postpurchase satisfaction on a platform, thereby affecting its future reputation. Thus, an important design question for platforms is, "What are the optimal weights for early versus later consumer reviews in aggregation and recommendation algorithms?"

Finally, our article suggests several avenues for future research. First, our model captures credibility concerns in a reduced-form way. While this suffices for our purpose, future researchers might want to develop a complete model of reputation. Combining the dynamics of reputational concerns with polarization can help answer whether reputation systems can improve/exacerbate polarization. Another possible direction is to use data on group decisions to identify and isolate the different sources of polarization (e.g., strategic incentives, polarization of beliefs, behavioral biases). Field experiments or observational data with exogenous variation in these sources can improve our understanding of how these factors contribute. It would also be valuable to empirically investigate the extent and the nature of the divergence between the polarization of actions and preferences/beliefs. Finally, it may also be useful to tie polarized group decisions/outcomes to broader firm-level decisions or societal decisions.

Acknowledgments

The authors 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.

Author Contributions

Ganesh Iyer and Hema Yoganarasimhan contributed equally and are listed alphabetically.

Associate Editor

Anthony Dukes

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz (2009), "Fragility of Asymptotic Agreement Under Bayesian Learning," unpublished paper, Department of Economics, Massachusetts Institute of Technology.

Aguilar, Julian (2015), "At Shooting Range, Abbott Signs 'Open Carry' Bill," *The Texas Tribune* (June 13), https://www.texastribune.org/2015/06/13/abbott-signs-open-carry-bill/.

Amaldoss, Wilfred and Sanjay Jain (2005), "Pricing of Conspicuous Goods: A Competitive Analysis of Social Effects," *Journal of Marketing Research*, 42 (1), 30–42.

Andreoni, James and Tymofiy Mylovanov (2012), "Diverging Opinions," *American Economic Journal: Microeconomics*, 40 (1), 209–32.

Austen-Smith, David (1990), "Information Transmission in Debate," *American Journal of Political Science*, 34 (1) 124–52.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M.B. Fallin Hunzaker, et al. (2018), "Exposure to Opposing Views on Social Media Can Increase Political Polarization," *Proceedings of the National Academy of Sciences*, 1150 (37), 9216–21.

Baliga, Sandeep, Eran Hanany, and Peter Klibanoff (2013), "Polarization and Ambiguity," *American Economic Review*, 103 (7), 3071–83.

Baron, Robert S. (2005), "So Right It's Wrong: Groupthink and the Ubiquitous Nature of Polarized Group Decision Making," *Advances in Experimental Social Psychology*, 37, 219–53.

Bénabou, Roland (2012), "Groupthink: Collective Delusions in Organizations and Markets," *Review of Economic Studies*, 80 (2), 429–62.

Bernheim, B. Douglas (1994), "A Theory of Conformity," *Journal of Political Economy*, 102 (5), 841–77.

Bhattacharya, Sourav and Arijit Mukherjee (2013), "Strategic Information Revelation When Experts Compete to Influence," *RAND Journal of Economics*, 44 (3), 522–44.

Bordley, Robert F. (1983), "A Pragmatic Method for Evaluating Election Schemes Through Simulation," *American Political Science Review*, 77 (1), 123–41.

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

Che, Yeon-Koo and Ian L. Gale (1998), "Caps on Political Lobbying," *American Economic Review*, 88 (3), 643–51.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing research*, 43 (3), 345–54.

Cohn, Nate (2014), "Polarization Is Dividing American Society, Not Just Politics," *New York Times* (June 12), https://www.nytimes.com/2014/06/12/upshot/polarization-is-dividing-american-society-not-just-politics.html.

Corfman, Kim and Donald Lehmann (1987), "Models of Cooperative Group Decision-Making and Relative Influence: An Experimental Investigation of Family Purchase Decisions," *Journal of Consumer Research*, 14 (1), 1–13.

Dai, Weijia, Jungmin Lee, Ginger Jin, and Michael Luca (2018), "Aggregation of Consumer Ratings: An Application to Yelp.com," *Quantitative Marketing and Economics*, 160 (3), 289–339.

Delmas, Magali, Jinghui Lim, and Nicholas Nairn-Birch (2016), "Corporate Environmental Performance and Lobbying," *Academy of Management Discoveries*, 20 (2), 175–97.

De-Wit, Lee, Sander Van der Linden, and Cameron Brick (2019), "Are Social Media Driving Political Polarization?" *Greater Good Magazine* (January 16), https://greatergood.berkeley.edu/article/item/is_social_media_driving_political_polarization.

Dixit, Avinash K. and Jörgen W. Weibull (2007), "Political Polarization," *Proceedings of the National Academy of Sciences*, 104 (18), 7351–56.

Eastland, Terry (2018), "Google in the Dock," *Washington Examiner* (April 20), https://www.washingtonexaminer.com/weekly-standard/google-says-it-hires-on-merit-while-also-practicing-diverse-only-recruiting-a-lawsuit-is-calling-them-on-it.

Elisahberg, Jehoshua, Stephen LaTour, Arvind Rangaswamy, and Stern Louis (1986), "Assessing the Predictive Accuracy of Two Utility-Based Theories in a Marketing Channel Negotiation Context," *Journal of Marketing Research*, 23 (2), 101–10.

Eliashberg, Jehoshua and Robert Winkler (1981), "Risk Sharing and Group Decision Making," *Management Science*, 27 (11), 1221–35.

Esteban, Joan and Gerald Schneider (2008), "Polarization and Conflict: Theoretical and Empirical Issues," *Journal of Peace Research*, 45 (2), 131–41.

Gallup (2015), "Tea Party Movement," Online, retrieved March 27, 2016. Available at: http://www.gallup.com/poll/147635/tea-party-movement.aspx.

Gilligan, Thomas W. and Keith Krehbiel (1989), "Asymmetric Information and Legislative Rules with a Heterogeneous Committee," *American Journal of Political Science*, 33 (2), 459–90.

Glaeser, Edward L. and Cass R. Sunstein (2009), "Extremism and Social Learning," *Journal of Legal Analysis*, 10 (1), 263–324.

Godes, David and José C. Silva (2012), "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*, 31 (3), 448–73.

Harstad, Bård and Jakob Svensson. (2006), "Bribes, Lobbying, and Development," CEPR Discussion Papers 5759.

Hartmann, Wesley R., Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, and David Godes (2008), "Modeling Social Interactions: Identification, Empirical Methods and Policy Implications," *Marketing Letters*, 19 (3/4), 287–304.

Huitlin, Susan (2015), "Guns on Campus: An Overview," *National Conference of State Legislatures* (October), http://www.ncsl.org/research/education/guns-on-campus-overview.aspx.

Hulac, Benjamin (2016), "Clean Energy Firms Lobby Congress as Much as Dirty Firms Do," *Scientific American* (September 9), https://www.scientificamerican.com/article/clean-energy-firms-lobby-congess-as-much-as-dirty-firms-do/.

Isenberg, Daniel J. (1986), "Group Polarization: A Critical Review and Meta-Analysis," *Journal of Personality and Social Psychology*, 50 (6), 1141–51.

Iyer, Ganesh and David Soberman (2016), "Social Responsibility and Product Innovation," *Marketing Science*, 35 (5), 727–42.

Klein, Nadav, Ioana Marinescu, Andrew Chamberlain, and Morgan Smart (2018), "Online Reviews Are Biased. Here's How to Fix Them," *Harvard Business Review* (March 6), https://hbr.org/2018/03/online-reviews-are-biased-heres-how-to-fix-them.

Kondor, Péter (2012), "The More We Know About the Fundamental, the Less We Agree on the Price," *Review of Economic Studies*, 79 (3), 1175–1207.

Krishna, Vijay and John Morgan (2001), "A Model of Expertise," *Quarterly Journal of Economics*, 116 (2), 747–75.

Lord, Charles G., Lee Ross, and Mark R. Lepper (1979), "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence," *Journal of Personality and Social Psychology*, 37 (11), 2098–2109.

Main, Eleanor C. and Thomas G. Walker (1973), "Choice Shifts and Extreme Behavior: Judicial Review in the Federal Courts," *Journal of Social Psychology*, 910 (2), 215–21.

Myers, David G. (1975), "Discussion-Induced Attitude Polarization," *Human Relations*, 28 (8), 699–714.

Myers, David G. (1982), "Polarizing Effects of Social Interaction," *Group Decision Making*, 125, 137–38.

Nielsen, Michael and Rush T. Stewart (2020), "Persistent Disagreement and Polarization in a Bayesian Setting," *British Journal for the Philosophy of Science*, 72 (1), 51–78.

Pacuit, Eric (2011), "Voting Methods," *Stanford Encyclopedia of Philosophy* (August 3), https://plato.stanford.edu/entries/voting-methods/.

Potters, Jan and Frans Van Winden (1992), "Lobbying and Asymmetric Information," *Public Choice*, 74 (3), 269–92.

Rabin, Matthew and Joel L. Schrag (1999), "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, 114 (1), 37–82.

Rao, Vithala and Joel Steckel (1991), "A Polarization Model for Describing Group Preferences," *Journal of Consumer Research*, 18 (1) 108–18.

Rawls, John (1971), *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Stoll, John D. (2019), "When It Comes to Colin Kaepernick, the Flag and Nike, It's Just Business," *The Wall Street Journal* (July 3), https://www.wsj.com/articles/when-it-comes-to-colin-kaepernick-the-flag-and-nike-its-just-business-11562161561.

Stoner, James Arthur Finch (1961), "A Comparison of Individual and Group Decisions Involving Risk," doctoral thesis, Massachusetts Institute of Technology.

Sun, Monic, Xiaoquan Zhang, and Feng Zhu (2019), "U-Shaped Conformity in Online Social Networks," *Marketing Science*, 38 (3), 461–80.

Sunstein, Cass R. (2002), "The Law of Group Polarization," *Journal of Political Philosophy*, 10 (2), 175–95.

Taylor, Kate (2017), "Brands Including Papa John's and Starbucks Are Victims of a 'Consumer Awakening' as Boycotts Explode in Trump's America," *Business Insider* (November 23), https://www.businessinsider.com.au/boycott-most-polarizing-foods-in-trump-era-2017-11.

Todd, Chuck, Mark Murray, and Carrie Dann (2014), "There's the Tea Party—And Then There's Everyone Else," *NBC News* (June 19), http://www.nbcnews.com/politics/first-read/tea-party-stands-alone-key-issues-n135426.

Yoganarasimhan, Hema (2012), "Cloak or Flaunt? The Fashion Dilemma," *Marketing Science*, 31 (1), 74–95.

Zimper, Alexander and Alexander Ludwig (2009), "On Attitude Polarization Under Bayesian Learning with Non-Additive Beliefs," *Journal of Risk and Uncertainty*, 39 (2), 181–212.

Zuber, Johannes A., Helmut W. Crott, and Joachim Werner (1992), "Choice Shift and Group Polarization: An Analysis of the Status of Arguments and Social Decision Schemes," *Journal of Personality and Social Psychology*, 62 (1), 50–61.