In defense of exclusionary deliberation: communication and voting with private beliefs and values

Adam Meirowitz\(^1\)

Dept. of Politics
Princeton University
ameirowi@princeton.edu
September 7, 2004

Fundamentally, deliberative democracy is an institution in which participants communicate and then vote. We analyze strategic behavior in this type of institution when agents do not necessarily have common beliefs and values. The potential for some pairs of participants to have diametrically opposed preferences makes it difficult to support equilibria in which participants truthfully reveal their private information. Nonetheless, truthful equilibria are shown to exist for some (but not all) parameterizations in which \textit{non-common values} are likely. Truthful equilibria exist if and only if participants of all possible preference types believe that it is more likely that a majority of the group share their preference type than a majority of the group have opposed preferences. Even when truthful equilibria fail to exist, the probability that the collective choice corresponds to that which a majority would choose, with full-information, approaches one as population size tends to infinity. Despite this limiting efficiency, larger groups need not outperform smaller groups as truthful equilibria are easier to support with small deliberative bodies. The design of deliberative institutions to aggregate information involves a trade-off between the statistical benefit of more participants and the strategic problems associated with information transmission in larger settings without common values. For many reasonable parameterizations, the latter effect is dominant and excluding randomly chosen participants is desirable.

\(^1\)I appreciate comments from Chris Achen, John Duggan, Mark Fey, William Keech, John Patty, Tom Romer and Roberto Weber. I thank Natasha Zharinova for her assistance.
1 Introduction

There are again two methods of removing the causes of fraction: the one, by destroying the liberty..., the other, by giving every citizen the same opinions, the same passions, and the same interests. .... The second expedient is as impractical as the first would be unwise. As long as the reason of man continues fallible, and he is at liberty to exercise it, different opinions will be formed. As long as the connection subsists between his reason and his self-love, his opinions and his passions will have a reciprocal influence on each other; and the former will be objects to which the latter will attach themselves. The diversity in the faculties of men, from which the rights of property originate, is not less an insuperable obstacle to a uniformity of interests.... The latent causes of faction are thus sown in the nature of man; and we see them everywhere brought into different degrees of activity, according to the different circumstances of civil society. A zeal for different opinions concerning religion, concerning government, and many other points, as well as speculation as of practice; an attachment to different leaders ambitiously contending for pre-eminence and power; or to persons of other descriptions whose fortunes have been interesting to the human passions, have, in turn, divided mankind into parties, inflamed them with mutual animosity, and rendered them much more disposed to vex and oppress each other than to cooperate for their common good. (James Madison 1788 , Federalist 10, p. 131-32.)

Equally important, the republican belief in the subordination of the private interests to the public good carries a risk of tyranny and even mysticism. The belief is also threatening to those who reject the existence of a unitary public good, and who emphasize that conceptions of the good are plural, and dependent on perspective and power. (Cass Sunstein 1988 p. 1540)

In many collective choice settings, like legislatures, town-hall meetings, corporate board meetings, advisory committee meetings, and of course faculty meetings, individuals can freely communicate prior to formal voting procedures. The possibility of argument, debate and even reasoned discourse offered by these deliberative settings has not escaped the attention of prominent scholars.2 In recent scholarship, Sunstein (1988), for instance, argues that deliberation can lead to "uniquely correct outcomes", and Guttmann and Thompson (1996) "believe that a deliberative perspective can help resolve some moral disagreements in democratic politics,..." and, "... help citizens treat one another with mutual respect as they deal with the disagreements that invariably remain"(p. 9). Fishkin (1991) defends several reforms that introduce deliberation by randomly selected masses to various stages of the republican system. Despite its broad appeal to normative scholars, deliberation has been challenged by several authors3 and empirical evidence on the effectiveness of

---

2See Mendelberg (2002) for a recent review of the normative and empirical literature on deliberative democracy.
3For example Sander (1997) notes that inequities in persuasiveness may translate into inequitable conclusions.
deliberation is mixed. Mendelberg (2002) summarizes the evidence, "Group discussion sometimes meets the expectations of deliberative theorists, other times falls short. Participants can, as theorists wish, conduct themselves with empathy for others, equality and open-mindedness. But attempts to deliberate can also backfire." (p.151). Despite the presence of an extensive literature, few scholars of deliberative democracy have explicated a clear positive logic of discourse in the presence of Madison’s non-uniformity of interests and Sunstein’s private interests. Page and Shapiro (p. 111) conclude "We consider it to be very much an open question just how well deliberation works, by what mechanisms, under what circumstances."

The starting point for this project and a few other recent game-theoretic papers (Coughlan 2001; Austen-Smith and Feddersen 2003a,b, Gerardi and Yariv 2003, Hafer and Landa 2003) is that a positive theory explaining how individuals will behave in particular types of deliberative settings is a useful (if not necessary) precursor to the development of compelling normative arguments about the effectiveness of deliberation as a means for aggregating private information and making collective choices. In this study we address a basic question: In a deliberative setting in which (i) agents may have extremely heterogenous preferences (or conceptions of the good) which preclude universally-acceptable consensus, and (ii) these preferences are private information, to what extend do the incentives to deceive each other limit the potential for information transmission and aggregation?

The tension between a notion of the public good and private values has been noted by normative scholars. Guttman and Thompson write, "The aim of moral reasoning that our deliberative democracy prescribes falls between impartiality, which requires something like altruism, and prudence, which demands no more than enlightened self-interest. Its first principal is reciprocity...” They go on to define reciprocity, "[reciprocity] can be seen in the difference between acting in one’s self interest (say, taking advantage of a legal loophole or a lucky break) and acting fairly (following rules in the spirit than one expects others to adopt)” (p.2). Sunstein explains, "The republican belief in deliberation counsels political actors to achieve a measure of critical distance from prevailing desires and practices, subjecting these desires and practices to scrutiny and review,” but he concedes that “this is not to suggest that deliberation calls for some standard entirely external to private beliefs and values (as if such a thing could be imagined).” This paper is fixed in the rational choice perspective, focusing on the effectiveness of deliberation in aggregating private beliefs and values when participants are prudent (or self interested). The findings are mixed in their support for deliberative democracy. First, in some cases in which agents might have diametrically opposed preferences equilibria in which agents are truthful exist. Despite this, the presence of such truthful equilibria hinges on individual perceptions about the likelihood that most participants have similar values. As such, efficient information transmission requires that everyone believe they are likely to have the values of the majority. Second, even if truthful equilibria fail to exist, very large deliberative bodies will select decisions that are optimal for a

---

4See also Mendelberg’s review of related theoretic and experimental work in psychology on small group dynamics that also seek to understand the process of deliberation.
majority of the participants. These findings suggest that despite the existence of private beliefs and values deliberation may be effective.

The analysis demonstrates a tension between broad participation (the size of a group) and the effectiveness of its deliberation. Democratic scholars have long noted that democracies cannot function if they are too large. Madison, for instance argued that the United States was too geographically vast for direct democracy (Federalist 14). Hamilton concluded that even many of the states were too large for effective direct democracy (Federalist 9). In contrast, Condorcet’s (1786) argument concludes that the larger the population the higher the probability that a democracy will make the ”right” decision.\footnote{Social psychologist (e.g., Stasser and Titus 1985; Stasser 1992; Gigone and Hastie 1993, 1997) point out that group dynamics may result in group discussions that ignore (or at least underemphasize) information held only by a few members and instead focus on the information common to many members of the group.} When agents are uncertain of each other’s values we find that democratic performance may suffer from increased population size despite the statistical Condorcetian effect. For small groups excluding agents from the deliberative body may improve group performance as truthful equilibria become easier to support when group size decreases. In larger settings heterogeneity is likely and the fear of dishonesty reduces the value of discourse. While sufficiently large groups will tend to make desirable decisions even without effective communication, group size needs to be quite large for the statistical effect to swamp the inefficiency from poor information transmission. For group sizes that can feasibly fit under one roof and communicate, group size reductions may be desirable. Moreover, for tough problems in which individual private information is of low quality this finding is most pronounced. Interestingly, the 1970’s saw debate about reforming jury sizes in U.S. jury trials. Proponents of the reform were worried that with 12 jurors it was too likely that a peculiar juror would derail deliberation.

While the desirability of participants is ambiguous, preference homogeneity is unequivocally desirable, as increased confidence that participants are similar makes truthful behavior easier to support. Overall then, whereas some normative scholars view deliberation as a means by which heterogenous interests can come together and reach consensus, with larger groups being better, the analysis demonstrates that from the perspective of aggregating private values and beliefs to make policy decisions, small homogenous groups are likely to outperform large heterogenous groups.

1.1 Related Literature

Several strands of research have implications for the study of deliberation. Models of debate (e.g., Lipman and Seppi 1995; Glazer and Rubinstein 2001) involve communication in non-cheap talk settings. Hafer and Landa (2003) focus on deliberation when agents are endowed with a set of sentences that they can accept as valid. Also related is the literature on strategic voting in jury games (Austen-Smith and Banks 1996; Feddersen and Pesendorfer 1996, 1997, 1998; Myerson 1994; Witt 1998; McClellan 1998; Duggan and
Martinelli 2001; Meirowitz 2002). Too extensions of the jury models separately address the key issues of this paper. Kim (2004) extends the standard jury model to include two groups of voters with opposed state contingent preferences. In this setting he finds that as group size tends to infinity majority rule leads to the full-information majority preferred outcome with probability approaching one. Coughlan (2001) considers a model of jury deliberations involving one round of cheap-talk communication and then a vote, demonstrating that if all agents have sufficiently similar preferences, but different information about some unknown state, they will share their private information about the unknown state. With nearly identical preferences, after hearing the public speeches agents will agree on which course of action is most desirable. A key feature of this model is the assumption that all agents have very similar preferences (they want to convict guilty defendants, and acquit innocent ones) and that all agents know that they all have similar preferences. In the language of game theory, the problem involves common knowledge of common values.

While this first finding is promising for deliberative democrats, it should not be over-weighted. A defense for deliberative democracy that rests on the assumption that there are no pluralistic concerns is underwhelming. The positive case for deliberative democracy seems to require treatment of situations where agents might possess heterogeneous preferences. To this end, Austen-Smith and Feddersen (2003a), move beyond the basic common values problem considering agents that have private biases and care about a common good. The analysis is focused on the case of three decision makers and the comparison of equilibrium behavior under unanimity rule and majority rule. Austen-Smith and Feddersen conclude that truthful deliberation need not occur in equilibria, but that majority rule is generally a better rule for eliciting information aggregation than unanimity rule. In a subsequent piece (Austen-Smith and Feddersen 2003b) the result is generalized as the authors demonstrate that under a minimal set of assumptions the equilibrium set will be better under majority rule then under unanimity rule. This finding suggests that the appeal of unanimity rule may have been overstated by scholars of deliberative democracy.

In an alternative mechanism design approach Gerardi and Yariv (2003) establish an equivalence between equilibria for all voting rules (except unanimity rule). Their paper allows for a large class of preferences profiles by agents – including deliberation problems in which agents may have diametrically opposed preferences so that their need not be commonality about how preferences over policies vary with private information. Gerardi and Yariv tell us that aside from unanimity rule the rule choice does not matter very much, but they are silent on the potential for information transmission and aggregation in these models with possibly opposed preferences as the equilibrium set is not characterized. It is here that the current paper makes a contribution.

We consider deliberative settings where (1) agents are uncertain both about some payoff relevant state, (for example a factual question like whether a policy decision will save tax-payer money, or a corporate decision will increase short-term profits), and (2) agents possess private information about their state-policy preferences need not be identical as agents may weight type I and II errors differently (or equivalently possess distinct prior probability assessment of the defendant’s guilt).
contingent preferences. So in these examples it is not known a priori that everyone in the room wishes to reduce taxes or maximize short-term profits, some pairs of individuals may have diametrically opposed preferences (but not know it). In other words an agent may know that she prefers to choose a policy that increases corporate profits, but she is not certain that everyone else in the deliberative body shares this preference. We investigate how this fact strains reliance on information that others provide. One interpretation is that of deliberation with potential saboteurs. The current model violates the monotonicity condition (axiom 3) in Austen-Smith and Feddersen (2003b), which requires commonality on how the uncertain state influences induced preferences over the policy alternatives. Here, with positive probability there will be a pair of agents having induced preferences that respond to information about the state in opposite directions. In the jury setting this possibility may be unlikely – requiring that agents fear some jurors prefer convicting the innocent and acquitting the guilty – but in other collective choice settings it can be quite likely. One recent example involves corporate governance following the Sarbanes-Oxley Act which attempts to increase diversity on corporate boards. As the net of potential members enlarges, so does the potential for conflicting preferences. We may worry that once board members face severe uncertainty about the motivations of other board members the effectiveness of deliberative decision-making may suffer. Another example involves a bureaucratic agency, consisting of some political appointees and some career bureaucrats, that make policy recommendations. In these last two examples the probability of preference heterogeneity may be a policy lever that institution designers can control.

In section 2 we present the basic model. In section 3 we focus first on truthful equilibria, stating necessary and sufficient conditions for their existence and then we characterize a particularly simple type of non-truthful equilibria. Section 4 considers the asymptotic properties of these equilibria proving that in the limit the full information majority rule decision is reached. In section 5 we highlight the potential trade-off between group size and informational efficiency. In section 6 we conclude with a brief discussion. The appendix presents an application of the revelation principal to this model showing the relationship between the set of truthful equilibria with a binary message space and the set of separating equilibria in games with a larger message space.

2 The Model

This departure from common knowledge of common values seems consistent with Sunstein’s weak principal of universalism, "exemplified by the notion of a common good, and made possible by practical reason. The republican commitment to universalism, or agreement as a regulative ideal, takes the form of a belief in the possibility of setting at least some normative disputes with substantively right answers." (p. 1541). The key phrase here is the belief in the possibility, as opposed to a belief in the certainty.

6
We consider a set of agents $N = \{1, 2, ..., n\}$ ($n$ odd) that must make a binary group decision, choosing a policy $p \in P = \{a, b\}$. Each agent has a binary preference type $\theta_i \in \Theta = \{-1, 1\}$ and there is an unknown state of the world $x \in X = \{a, b\}$. The common prior probability over the state is $\Pr(x = a) = \pi \geq \frac{1}{2}$. Agent $i$ has a Von Neumann-Morgenstern utility function that depends on the policy $p \in P$ the state $x \in X$ and her type $\theta_i \in \Theta$. This utility function is of the form

$$u_i(p; x, \theta_i) = \theta_i \cdot 1_{\{p=x\}}$$

where $1_{\{p=x\}}$ is the indicator function taking value 1 if $p = x$ and 0 otherwise. So type $\theta_1 = 1$ agents want to match $x$ and $p$ while type $\theta = -1$ agents want $x$ and $p$ unmatched. The potential for agents of both types captures a stark case of potentially opposed preferences. To capture situations in which agents are uncertain about the preferences of the other members of the deliberative body, we assume that only agent $i$ knows her type. In addition the agents do not observe $x$, but instead each agent receives an informative private signal $s_i \in \{a, b\}$ about $x$ with $\Pr(s_i = x) = g > \frac{1}{2}$. These signals are assumed to be conditionally (on $x$) independent. Thus, each agent observes a private signal $(\theta_i, s_i) \in \{-1, 1\} \times \{a, b\}$. Finally, we model the generation of preference types in the following manner. With probability $c \in [0, 1]$ the population has common values and with probability $1 - c$ the individual preference types are i.i.d draws. Specifically, $\Pr(\{\theta_i = \theta_j = 1 \forall i, j \in N\}) = xc$, $\Pr(\{\theta_i = \theta_j = -1 \forall i, j \in N\}) = (1 - z)c$ and with probability $1 - c$, $\theta_i$’s are generated from the binomial distribution with $\Pr(\theta_i = 1) = z$. So with either $c = 1$ or $z = 0, 1$ the model involves common-values. Without loss of generality we assume that $z \geq \frac{1}{2}$. We assume that the generation of agent types and the state $x$ are independent. For $c$ close to 1 and/or $z$ close to 1 this is a model in which it is highly likely that the population has common values, but agents know that there is a chance that some members of the population have dissimilar preferences.

The above describes a lottery over the state space $\Omega := X \times \Theta^n \times \{a, b\}^n$. The first $n + 1$ dimensions represent payoff relevant information and the last $n$ dimensions represent the imperfect signals that agents learn. We model deliberation as a simple two period Bayesian game of common knowledge with this state space. In period 0 nature selects the agent types, state and signals –with each agent observing her own preference type and signal $(\theta_i, s_i)$. In period 1 agents simultaneously send public messages, $m_i \in \{a, b\}$. By $m = (m_1, ..., m_n)$ we denote the profile of messages. In period 2 each agent having observed $m$ casts a simultaneous binary vote, $v_i \in \{a, b\}$. We assume that the collective decision is made with majority rule. So that $p = a$ is chosen iff $|i : v_i = a| > |i : v_i = b|$.8

A symmetric perfect Bayesian equilibrium in weakly undominated voting is a speech function $m(\theta_i, s_i) : \{-1, 1\} \times \{a, b\} \to \{a, b\}$, a voting rule $v_i(\theta_i, s_i, m) : \{-1, 1\} \times \{a, b\}^{n+1} \to \{a, b\}$ and a belief system

8Since the environments considered here satisfy the conditions in Gerardi and Levy, characterization of equilibrium outcomes for majority rule is equivalent ot characterization for all rules other than majority rule if one is willing to give up weak dominance in voting.
satisfying the requirements that the strategies are sequentially rational, the beliefs satisfy Bayes’ rule when it applies, and the voting rule does not select the agent’s second ranked alternative (given her beliefs about \( x \) conditional on \( m \)).

3 Equilibria

3.1 Truthful equilibria

In this setting, if \( x \in X \) and \( \theta \in \Theta^n \) were commonly known weakly undominated voting would yield the policy

\[
p^* = \begin{cases} 
  a & \text{if } x = a \text{ and } |i: \theta_i = 1| > \frac{n+1}{2} \text{ or } x = b \text{ and } |i: \theta_i = 1| < \frac{n+1}{2} \\
  b & \text{otherwise.} 
\end{cases}
\]

(2)

If instead only the preference types \( \theta \) and the private signals \( s \) were public knowledge undominated voting would yield the policy

\[
p^+ = \begin{cases} 
  a & \text{if } \mu > \frac{1}{2} \text{ and } |i: \theta_i = 1| > \frac{n+1}{2} \text{ or } \mu < \frac{1}{2} \text{ and } |i: \theta_i = 1| < \frac{n+1}{2} \\
  b & \text{if } \mu < \frac{1}{2} \text{ and } |i: \theta_i = 1| > \frac{n+1}{2} \text{ and } \mu > \frac{1}{2} \text{ or } |i: \theta_i = 1| < \frac{n+1}{2} \\
  a & \text{if } \mu = \frac{1}{2} 
\end{cases}
\]

(3)

where

\[
\mu = \frac{\pi g^a(s)(1-g)^b(s)}{\pi g^a(s)(1-g)^b(s) + (1-\pi)(1-g)^a(s)g^b(s)}
\]

(4)

and \( a(s) \) is the number of private signals with value \( a \) and \( b(s) \) is the number of private signals with value \( b \).

We now investigate whether there is a Perfect Bayesian equilibrium in weakly undominated voting strategies which implements the decision rule \( p^+(\theta, s) \). We focus on the potential for truthful equilibria. In the appendix we show that a truthful equilibrium will exist in this game if and only if a fully-revealing equilibrium exists when the message space has at least two elements. Thus, the focus on truthful equilibria and binary message spaces is not restrictive. Suppose agents are truthful, using the message function \( m(\theta_i, s_i) = s_i \). Let \( a(m) \) denote the number of messages that say \( a \) is the state and \( b(m) \) the number that say \( b \) is the state. Based on the public messages \( m = \omega^2 \), consistency requires that agent posteriors on \( x \) satisfy Bayes’ rule

\[
pr(x = a | a(m), b(m)) = \frac{\pi g^a(m)(1-g)^b(m)}{\pi g^a(m)(1-g)^b(m) + (1-\pi)g^b(m)(1-g)^a(m)}.
\]

(5)

Weak dominance requires that the individual voting rule satisfies,
\[ v_i(\theta_i, s_i, m) = \begin{cases} 
  a & \text{if } pr(x = a \mid m) > \frac{1}{2} \text{ and } \theta_i = 1 \\
  a & \text{if } pr(x = a \mid m) < \frac{1}{2} \text{ and } \theta_i = -1 \\
  b & \text{otherwise.} 
\end{cases} \] (6)

We ignore the knife-edged parameterizations in which \( pr(x = a \mid m) = \frac{1}{2} \) is possible (with truthful messages).

In a truthful equilibrium, all agents have the same information at the time that they cast ballots. Because they have opposing preferences, agents of types \( \theta_i = 1 \) and \( \theta_i = -1 \) will respond to \( pr(x = a \mid a(m), b(m)) \) in different manners. The former (latter) will prefer \( a \) when \( pr(x = a \mid a(m), b(m)) \) is high (low). Given that \( k \) of the \( n - 1 \) others are announcing \( m_1 = a \) a unilateral deviation by agent \( i \) in the message stage will effect voting behavior iff

\[ pr(x = a \mid k, n - k) < \frac{1}{2} < pr(x = a \mid k + 1, n - k - 1). \] (7)

If this condition is not true for any \( k < n \) then the messages will not effect voting and a truthful equilibrium trivially exists. In a truthful equilibrium, it must be the case that agent \( i \) would prefer to send a truthful message. In order for this to be the case the incentive to provide the agents with the same type as agent \( i \) the correct information when (7) is satisfied must dominate the incentive to provide agents with the other type incorrect information when (7) is satisfied. This will be the case if given \( \theta_i \) the probability that at least \( \frac{n+1}{2} \) agents have type \( \theta_i \) exceeds the probability that at least \( \frac{n+1}{2} \) agents have type not equal to \( \theta_i \). Since the types and signals are independent, an agent’s assessment of the probability that she is in the majority preference type is independent of her private signal \( s_i \). Specifically, agent \( i \) is willing to truthfully reveal \( s_i \) if either (7) is not satisfied for any \( k \) or

\[ pr(\mid j : \theta_j = \theta_i \mid > \frac{n + 1}{2} \mid \theta_i) > pr(\mid j : \theta_j \neq \theta_i \mid > \frac{n + 1}{2} \mid \theta_i). \] (8)

The inequality in (7) is satisfied for some \( k \) if

\[ pr(x = a \mid 0, n) < \frac{1}{2} < pr(x = a \mid n, 0) \] (9)

which is equivalent to

\[ \left( \frac{1-g}{g} \right)^n < \frac{\pi}{1-\pi} < \left( \frac{g}{1-g} \right)^n. \] (10)

Given the probability model, Bayes’ rule yields

\[ pr(\theta_j = \theta_i \mid \theta_i) = \begin{cases} 
  \frac{cz + (1-c)z^2}{cz + (1-c)z^2} & \text{if } \theta_i = 1 \\
  \frac{c(1-z) + (1-c)(1-z)^2}{c(1-z) + (1-c)(1-z)} & \text{if } \theta_i = -1. 
\end{cases} \] (11)
Since \( z \geq \frac{1}{2} \geq 1 - z \) this implies that if \( z < 1 \) the conditional probability that a decisive coalition of other participants has type \( \theta_j = \theta_i \)

\[
\eta(c, z, n) := \min_{\theta_i \in \{-1, 1\}} \left\{ pr(|j : \theta_j = \theta_i| \geq \frac{n + 1}{2} | \theta_i) \right\}
\]

\[
= cz + (1 - c) \sum_{j = \frac{n+1}{2}}^{n-1} \binom{n-1}{j} (1 - z)^j z^{n-j} \frac{cz + (1 - c)z}{cz + (1 - c)z} = c + (1 - c) \sum_{j = \frac{n+1}{2}}^{n-1} \binom{n-1}{j} (1 - z)^j z^{n-1-j}. \tag{12}
\]

and if \( z = 1 \) then \( \eta(c, z, n) = 1 \). Similarly the conditional probability that a decisive coalition has type \( \theta_j \neq \theta_i \) is

\[
\mu(c, z, n) := (1 - c) \sum_{j = \frac{n+1}{2}}^{n-1} \binom{n-1}{j} z^j (1 - z)^{n-1-j}.
\]

When (10) is satisfied the existence of truthful equilibria hinges on \( \eta \geq \mu \). To see this note that in a truthful equilibrium if the number of agents in \( N \setminus i \) with type \( \theta = 1 \) is exactly \( \frac{n-1}{2} \) then \( i \)'s expected utility from truthfulness about \( s_i \) or dishonesty about \( s_i \) followed by optimal voting is the same. In the remaining cases where at least \( \frac{n+1}{2} \) of the remaining \( n - 1 \) others have the same type, \( i \) can only affect the outcome through her message. When there are more \( \theta_j = \theta_i \) types truthfulness is best and when there are more \( \theta_j \neq \theta_i \) types dishonesty is best.

Clearly for \( c \geq \frac{1}{2} \) we have \( \eta \geq \frac{1}{2} \) so \( \eta \geq \mu \) (since \( \eta + \mu \leq 1 \)). With \( c < \frac{1}{2} \), the answer also depends on \( z \) and \( n \). Collecting these observations yields necessary and sufficient conditions for the existence of a truthful PBE in weakly undominated voting.

**Proposition 1** A truthful PBE in weakly undominated voting exists iff (10) is not satisfied or \( \eta(c, z, n) \geq \mu(c, z, n) \).

This characterization can be tightened if we consider whether a parameterization \( (\pi, c, z) \) will have truthful PBE in weakly undominated voting for arbitrary \( n \).

**Proposition 2** A truthful PBE in weakly undominated strategies exists for arbitrary \( n \) iff at least one of the following conditions is true:

1. \( c \geq \frac{1}{2} \)
2. \( z = \frac{1}{2} \).
3. \( z = 1 \)

10
Proof: First assume that $c \geq \frac{1}{2}$. In this case $\eta(c, z, n) \geq \frac{1}{2}$ and by proposition 1 a truthful PBE in weakly undominated strategies exists for any $n$. Second assume that $z = \frac{1}{2}$. This implies that for any $n$, $\eta(c, z, n) = \frac{1}{2} + \frac{c}{2} \geq \frac{1}{2}$ and by proposition 1 a truthful PBE in weakly undominated strategies exists for any $n$ and $c$. Third assume that $z = 1$ in this case $\eta(c, z, n) = 1$. Now suppose that neither 1 nor 2 nor 3 hold. For $n$ sufficiently large condition (10) is satisfied since $\left(\frac{1 - g}{g}\right)^n \rightarrow 0$ and $\left(\frac{g - 1}{1 - g}\right)^n \rightarrow \infty$. Thus existence of truthful equilibria for large $n$ requires that $\eta(c, z, n) - \mu(c, z, n) \geq 0$ is eventually true. For $z \in (\frac{1}{2}, 1)$, the term

$$\sum_{j=\frac{n+1}{2}}^{n-1} (n-1)_j (1 - z)^j z^{n-1-j}$$

converges to 0 as $n \rightarrow \infty$ so that $\eta(c, z, n) \rightarrow c$. Since $c < \frac{1}{2}$ we have $\eta(c, z, n) \rightarrow c < \frac{1}{2}$. Alternatively, in this case $\mu(c, z, n) \rightarrow (1 - c) > \frac{1}{2}$ so by proposition 1 there is not a truthful PBE in weakly undominated strategies for sufficiently large $n$. \[\blacksquare\]

The intuition is straightforward. In order for a truthful equilibrium to exist it must be the case that either messages do not effect voting, as is the case if (10) is not satisfied, or for any profile of private information an agent with this information will believe that she is more likely to have the same preference type as a majority of the voters and thus want to communicate her private information to the majority.

We can now consider the implications of proposition 2 for robustness of truthful equilibria in a model with common knowledge of common values. Consider a fixed triple $(n, \pi, g)$. Common knowledge of common values requires either $c = 1$ or $z = 1$. Starting with the case of $c = 1$ and $z > \frac{1}{2}$ if we perturb the game (by varying $z, c$ by the amounts $-\varepsilon_z, -\varepsilon_c$) then proposition 2 says that truthful equilibria will exist for any population size as long as $c - \varepsilon_c \geq \frac{1}{2}$. However, if we start with a game with $c = 0$ and $z = 1$ (common knowledge of common values) and then consider a perturbation with $c + \varepsilon_c < \frac{1}{2}$ proposition 2 tells us that this perturbation will not have truthful equilibria for every $n$. But since by taking $\varepsilon$ sufficiently small we can make the probability of any measurable event arbitrarily close under the original and perturbed games (for a fixed $n$) in a very real sense a model with $c < \frac{1}{2}$ and $z$ close to 1 is a perturbation of the common knowledge of common values assumption. Figure 1 represents the space of possible games (in terms of $c$ and $z$) depicting the set on which truthful equilibria exist for all $n$. This set is not relatively open indicating that truthful equilibria for all $n$ is not a robust finding, However there exist relatively open sets of models that have common knowledge of common values indicating that under some types of perturbations the property of existence of truthful equilibria for all $n$ is robust.

[Figure 1 about here]

The conclusions of this robustness analysis are restated in the following result.
Corollary 1 (Robustness): Departures from common values that involve $c$ close to 1 have no impact on the existence of truthful equilibria, but departures that have $c < \frac{1}{2}$ and $z$ close to 1, have a large impact on large population properties of the equilibrium set.

3.2 Non truthful equilibria

When neither of the sufficient conditions in proposition 1 are satisfied then in any PBE in weakly undominated voting some agents are not correctly revealing their signal $s_i$. In the appendix we shown that truthful equilibria exist in this game iff separating equilibria exist in a game with at least two possible messages. We now characterize a particularly simple equilibrium when $z > \frac{1}{2}$. Recall that $z \geq \frac{1}{2}$ and at $z = \frac{1}{2}$ a truthful equilibrium exists for any population size. Suppose that agents with $\theta_i = 1$ truthfully announce $m_i = s_i$ and agents with $\theta_i = -1$ announce $m_i \neq s_i$. So if $x = a$ then agent $j$ will announce message $a$ in the events that (1) agent $j$ has type $\theta = 1$ (because it is common values with this type or non-common values and this is her draw) and she observes $a$, or (2) agent $j$ has type $\theta = -1$ (because it is common values with this type or non-common values and this is her draw) and she observes $b$. Since agent $i$ observes her type at least one possible common values scenario (everyone has type opposite of the realized $\theta_i$) can be ruled out. Accordingly, the probability that an agent sends message $m_j = a$ given that the state is $x = a$ and $\theta_i = 1$ is

$\rho(a \mid a, 1) = \frac{czg + (1 - c)(zg + (1 - z)(1 - g))z}{cz + (1 - c)z} = czg + (1 - c)(zg + (1 - z)(1 - g)). \tag{14}$

Note that $\rho(b \mid b, 1) = \rho(a \mid a, 1)$. Similarly, the probability that an agent sends message $m_j = a$ given that the state is $x = b$ and $\theta_i = 1$ is

$\rho(a \mid b, 1) = c(1 - g) + (1 - c)(z(1 - g) + (1 - z)g). \tag{15}$

Note that $\rho(b \mid a, 1) = \rho(a \mid b, 1)$. The probability that an agent sends message $m_j = a$ given that the state is $x = a$ and $\theta_i = -1$ is

$\rho(a \mid a, -1) = c(1 - g) + (1 - c)(zg + (1 - z)(1 - g)). \tag{16}$

Note that $\rho(b \mid b, -1) = \rho(a \mid a, -1)$. The probability that an agent sends message $m_j = a$ given that the state is $x = b$ and $\theta_i = -1$ is

$\rho(a \mid b, -1) = c(1 - g) + (1 - c)(z(1 - g) + (1 - z)g). \tag{17}$

Note that $\rho(b \mid a, -1) = \rho(a \mid b, -1)$.
Using these terms we can express the posterior probability of \( x = a \) given the public messages \( m \) (and the summary statistics \( a(m) \) and \( b(m) \)) of the remaining \( n - 1 \) agents and agent \( i \)'s private information \((\theta_i, s_i)\) as

\[
pr(x = a \mid a(m), b(m), \theta_i, s_i) = \frac{\pi \phi(s_i) \rho(a \mid a, \theta_i)^{a(m)} \rho(b \mid a, \theta_i)^{b(m)}}{\pi \phi(s_i) \rho(a \mid a, \theta_i)^{a(m)} \rho(b \mid a, \theta_i)^{b(m)} + (1 - \pi)(1 - \phi(s_i)) \rho(a \mid b, \theta_i)^{a(m)} \rho(b \mid b, \theta_i)^{b(m)}}.
\]

where \( \phi(s_i) \) is \( g \) if \( s_i = a \) and \( 1 - g \) if \( s_i = b \). Agent \( i \)'s message is payoff relevant in the event that for some \( k < n - 1 \), and pair \((s_i, \theta_i)\) either.

\[
pr(x = a \mid k, n - 1 - k, \theta_i, s_i) < \frac{1}{2} < pr(x = a \mid k + 1, n - k, \theta_i, s_i)
\]

or

\[
pr(x = a \mid k, n - 1 - k, \theta_i, s_i) > \frac{1}{2} > pr(x = a \mid k + 1, n - k, \theta_i, s_i).
\]

For a triple \((k, s_i, \theta_i)\) in which one of the above conditions is satisfied agent \( j \)'s message will influence agent \( i \)'s voting. In contrast to the case of truthful messages, here it need not be the case that higher \( a(m) \) means \( x = a \) is more likely. Specifically, we need to worry about the possibility of (20) holding for some values of \( k \) and (21) holding for other values of \( k \). However, as long as \( \rho(a \mid a, \theta_i) > \rho(b \mid a, \theta_i) \) for both types of \( \theta_i \) this will not be the case. For \( \theta_i = 1 \) the relevant terms are

\[
\rho(a \mid a, 1) = cg + (1 - c)(zg + (1 - z)(1 - g))
\]

and

\[
\rho(b \mid a, 1) = c(1 - g) + (1 - c)(z(1 - g) + (1 - z)g)
\]

with the former larger (weakly) since \( g > (1 - g) \) and \( z > (1 - z) \). For \( \theta_i = -1 \) the relevant terms are

\[
\rho(a \mid a, -1) = c(1 - g) + (1 - c)(zg + (1 - z)(1 - g))
\]

and

\[
\rho(b \mid a, -1) = c(1 - g) + (1 - c)(z(1 - g) + (1 - z)g).
\]

Again the former is larger (weakly) since \( g > (1 - g) \) and \( z > (1 - z) \).

\footnote{The argument is simple. The difference, \( zg + (1 - z)(1 - g) - [z(1 - g) + (1 - z)g] = 4gz - 2z - 2g + 1 \) is increasing in \( z \) and \( g \) when \( z \geq \frac{1}{2} \) and \( g \geq \frac{1}{2} \) and at \( g = z = \frac{1}{2} \) the difference is 0.}
Thus if players $N \setminus j$ are using the conjectured strategy profile, regardless of $j$’s private information, if $\theta_j = 1$ ($\theta_j = -1$) and $j$’s weakly undominated best response following $a(m)$ and $b(m)$ messages of $a$ and $b$ is to vote for $a$ ($b$) then following $a'(m) > a(m)$ messages of $a$ agent $j$’s weakly undominated best response will be to vote for $a$ ($b$). And similarly increasing the number of messages for $b$ will make a $\theta_j = 1$ ($\theta_j = -1$) type more satisfied with voting for $b$ ($a$). Since the likely majority preference type is $\theta_i = 1$ (as $\theta > \frac{1}{2}$) this monotonicity implies that agent $i$’s best response is to announce $a$ if she would prefer that $a$ is selected and announce $b$ if she would prefer that $b$ is selected. We now examine the form of $i$’s induced preferences over messages.

In Austen-Smith and Feddersen 2003a the relevant condition is whether $a$ or $b$ is desirable conditional on being pivotal (message pivotal). However, here agents of type $\theta_i = 1$ can only be dissatisfied with the outcome that results from being truthful (when everyone else plays the conjectured equilibrium) if it is the case that their message was irrelevant (not message pivotal). Alternatively, agents of type $\theta_i = -1$ can only be dissatisfied with the outcome that results from lying (when everyone else plays the conjectured equilibrium) if they are not message pivotal.

To see this, consider an agent $i$ with $s_i = a$ and $\theta_i = 1$. The conjectured equilibrium calls for $m_i = a$. When would such a message turn out to be suboptimal for agent $i$? This can occur only if following the messages $m$ agent $i$ prefers policy $b$ (because there are $k$ other messages of $a$ and $pr(x = a | k, n-k-1, \theta_i, a) < \frac{1}{2}$) but $a$ passes and would not have passed had $m_i = b$ been announced (thus $i$’s message was pivotal). However, since $s_i = a$ if given the public messages $i$ believes that $b$ is more likely than $a$ than every other agent will form the same conclusion based on the public messages and their own private signal. This is true because

$$pr(x = a | k, n-k-1, \theta_i, a) > pr(x = a | k, n-k-1, \theta_i, b).$$  \hfill (25)

Given this $i$’s message could not have been pivotal. The only way that agent $i$ with $s_i = a$ and $\theta_i = 1$ is dissatisfied with a decision of $a$ is if both the majority type is $\theta = -1$ and a change of $m_i$ would not have affected the outcome. Thus when $\theta_i = 1$, $m_i(a) = a$ is a best response message. Now suppose that agent $i$ has $s_i = b$ and $\theta_i = 1$. When would a message of $b$ be suboptimal? As before this requires that $k$ messages of $a$ were sent and $pr(x = a | k, n-k-1, \theta_i, b) > \frac{1}{2}$. But given (26) this means that everyone will believe that $a$ is more likely than $b$ and a different message from $b$ would not have changed the outcome. Accordingly, if $\theta_i = 1$ then a truthful message is a best response.

Now consider agent $i$ with $s_i = a$ and $\theta_i = -1$. When would $i$ be dissatisfied with her message of $b$? This would require that she wound up preferring $a$ and her message caused $b$ to be chosen. Again since she has a signal of $a$ and believes that $b$ is more likely conditional on the messages everyone must form the same conclusion and thus her message could not have been influential. Finally for an agent $i$ with $s_i = b$ and $\theta_i = -1$. When would $i$ be dissatisfied with her message of $a$? This would require that she wound
up preferring b and her message caused a to be chosen. Again since she has a signal of b and believes that a is more likely conditional on the messages everyone must form the same conclusion and thus her message could not have been influential. Accordingly, if \( \theta_i = -1 \) then \( m_i(s_i) \neq s_i \) is a best response.

In an equilibrium of this type it is believed that more agents are sending truthful messages then none truthful messages (since the probability that an individual is of type \( \theta_i = 1 \) is greater than \( \frac{1}{2} \)). Given this, in forming beliefs about the state, more messages of a increase the posterior probability that \( x = a \). Again since it is expected that more agents will be of type \( \theta_i = 1 \) this means that more messages of a increase the likelihood (weakly) that a is passed. In deciding what message to send agents first determine if they expect to be in the majority or minority in terms of preference types \( \theta_i \). If they expect to be in the majority (here \( \theta_i = 1 \)) then they truthfully reveal their private information about \( x \). However, if they expect to be in the minority (\( \theta_i = -1 \)) they lie sending the opposite message. This behavior is rational for the agent because if their message has any effect on the final voting it will be in the natural way: saying \( a \) increases the likelihood that \( a \) is enacted and if an agent decides that based on hearing the messages she prefers a policy other than the one she advocated it must be the case that either a majority of the individuals will also wind up supporting the agent’s new preferred policy or the agent’s message was inconsequential.

Assume participants use this message strategy, let \( m_p \in \{a, b\} \) denote the message that obtains a plurality and assume that participants \( N \setminus j \) use the voting rule

\[
v_i(\theta_i, s_i, m) = \begin{cases} 
a & \text{if } pr(x = a \mid a(m), b(m), \theta_i, s_i) > \frac{1}{2} \text{ and } \theta_i = 1 \\
 or < \frac{1}{2} \text{ and } \theta_i = -1 
\end{cases} \tag{26}
\]

b otherwise.

We consider the choice of a participant, \( j \), that observed signal \( s_j \) and has type \( \theta_j = 1(-1) \). If \( pr(x = m_p \mid a(m), b(m), \theta_i, s_i \neq m_p) \geq \frac{1}{2} \) then every type 1 participant will vote for \( v = m_p \) and every type -1 participant will vote for \( v \neq m_p \). In this case being pivotal is uninformative about \( x \) and \( j \)’s best response in the voting stage is to use the specified strategy-voting for \( v = (\neq)m_p \) if \( \theta_j = 1(-1) \). If \( pr(x = m_p \mid a(m), b(m), \theta_i, s_i = m_p) < \frac{1}{2} \) then every type 1 participant will vote for \( v \neq m_p \) and every type -1 participant will vote for \( v = m_p \). In this case being pivotal is uninformative about \( x \) and \( j \)’s best response in the voting stage is to use the specified strategy-voting for \( v = (\neq)m_p \) if \( \theta_j = 1(-1) \). If \( pr(x = m_p \mid a(m), b(m), \theta_i, s_i = m_p) \geq \frac{1}{2} \) \( \geq pr(x = m_p \mid a(m), b(m), \theta_i, s_i \neq m_p) \) then according to the strategy above all \( \{\theta_i = 1, s_i = m_p\} \) and \( \{\theta_i = -1, s_i \neq m_p\} \) participants vote for \( m_p \), all \( \{\theta_i = 1, s_i \neq m_p\} \) and \( \{\theta_i = -1, s_i = m_p\} \) participants vote for \( v \neq m_p \). In this case \( j \) is pivotal iff both sets have the same number of participants. Accordingly, the probability of \( x = a \), conditional on being pivotal and \( s_j \) corresponds to \( pr(x = a \mid a(m) = \frac{n-1}{2}, b(m) = \frac{n-1}{2}, \theta_i, s_i) \). This means that \( j \) can only be pivotal in the case that \( pr(x = m_p \mid a(m) = \frac{n-1}{2}, b(m) = \frac{n-1}{2}, \theta_i, s_i = m_p) \geq \frac{1}{2} \geq pr(x = m_p \mid a(m) = \frac{n-1}{2}, b(m) = \frac{n-1}{2}, \theta_i, s_i \neq m_p) \). But in this case the specified strategy is a best response for \( j \).
We have thus established the following result.

**Proposition 3** If $z > \frac{1}{2}$ (recall that $g > \frac{1}{2}$ is assumed) then there exists a PBE in weakly undominated voting with message mapping

$$m_i(s_i, \theta_i) = \begin{cases} 
  s_i & \text{if } \theta_i = 1 \\
  \{a, b\} \setminus s_i & \text{if } \theta_i = -1.
\end{cases}$$

\section{Efficiency}

Since a truthful equilibrium with weakly undominated voting strategies implements the decision rule (3) it fully aggregates the private information. We now consider what happens as the population size tends to infinity showing that the probability that the decision corresponds to that of (2) approaches one. Since individual signals, $s_i$ are conditionally independent the probability that the majority of public messages are correct tends to one. This result is a version of the statistical Condorcet Jury Theorem (Berend and Paroush 1998). Given this the probability that individuals in society know $x$ converges to 1, and thus under weakly undominated voting the majority decision will correspond to the majority preferred policy (with $x$ known) with probability approaching one.\(^\text{10}\) We now formalize this argument.

Let $\Gamma(n, c, z, g, \pi)$ denote a deliberation game with parameters $n, c, z, g, \pi$ satisfying one of the conditions in proposition 2. Consider the triangular array of independent games, with $n$ (odd) tending to infinity. Let the random variable $x^T(\Gamma(n, c, z, g, \pi))$ denote the realized state in game $\Gamma(n, c, z, g, \pi)$. Let the random variable $\theta^T(\Gamma(n, c, z, g, \pi))$ denote the preference type of the majority of voters in the game. Let the random variable $m^T(\Gamma(n, c, z, g, \pi))$ denote the message that is sent by the majority of senders in the truthful PBE in weakly undominated voting to the game. Let the random variable $o^T(\Gamma(n, c, z, g, \pi))$ denote the state that is more likely under $\text{pr}(x \mid a(m), b(m))$ characterized by (5). Let the random variable $p^T(\Gamma(n, c, z, g, \pi))$ denote the policy that is chosen in the truthful PBE in weakly undominated voting to the game. Let the random variable $h^T(\Gamma(n, c, z, g, \pi))$ denote the majority rule core decision in the problem in which $x$ and $\theta$ is known (chosen by the rule in (2)). Thus

$$h^T(\Gamma(n, c, z, g, \pi)) = \begin{cases} 
  a & \text{if } x^T(\Gamma(n, c, z, g, \pi)) = a \text{ and } \theta^T(\Gamma(n, c, z, g, \pi)) = 1 \\
  b & \text{if } x^T(\Gamma(n, c, z, g, \pi)) = b \text{ and } \theta^T(\Gamma(n, c, z, g, \pi)) = -1 \\
  b & \text{otherwise.}
\end{cases} \quad (27)$$

**Proposition 4** If $c, z$ satisfy the conditions of proposition 2 ($c \geq \frac{1}{2}$ or $z = \frac{1}{2}$) then $\text{pr}[p^T(\Gamma(n, c, z, g, \pi)) = h^T(\Gamma(n, c, z, g, \pi))] \to 1.$

\(^{10}\)Some authors (Miller 1986, Ladha 1992) have used the term full information majority rule (FIMR) outcome to describe the choice of the policy that is majority preferred.
Proof: Assume the hypothesis. For each \( n \) in the selected PBE to \( \Gamma(n, c, z, g, \pi) \) each message is an i.i.d draw that corresponds to \( x \) with probability \( g > \frac{1}{2} \). By the statistical Condorcet Jury Theorem for independent draws

\[
\Pr[m^T(\Gamma(n, c, z, g, \pi)) = x^T(\Gamma(n, c, z, g, \pi))] \rightarrow 1. \tag{28}
\]

Since the optimal decision rule \( p^+ \) in (3) will do at least as well as a rule that concludes that \( x^T(\Gamma(n, c, z, g, \pi)) = m^T(\Gamma(n, c, z, g, \pi)) \) we have

\[
\Pr[o^T(\Gamma(n, c, z, g, \pi)) = x^T(\Gamma(n, c, z, g, \pi))] \rightarrow 1. \tag{29}
\]

Since weakly undominated voting selects \( p^T(\Gamma(n, c, z, g, \pi)) = o^T(\Gamma(n, c, z, g, \pi)) \) iff \( \theta^T(\Gamma(n, c, z, g, \pi)) = 1 \) we have

\[
\Pr[p^T(\Gamma(n, c, z, g, \pi)) = h^T(\Gamma(n, c, z, g, \pi))] \rightarrow 1. \tag{30}
\]

We now consider the asymptotics of sequences of non-truthful PBE establishing that the equilibria characterized in Proposition 3 also fully aggregate information in the limit. Let \( z, g > \frac{1}{2} \) and let \( x^L(\Gamma(n, c, z, g, \pi)), \theta^L(\Gamma(n, c, z, g, \pi)), m^L(\Gamma(n, c, z, g, \pi)), p^L(\Gamma(n, c, z, g, \pi)), h^L(\Gamma(n, c, z, g, \pi)) \) represent the random variables for equilibria characterized in proposition 3. So the superscript \( L \) captures the fact that some participants lie.

**Proposition 5** If \( z, g \) satisfy the conditions of proposition 3 then \( \Pr[p^L(\Gamma(n, c, z, g, \pi)) = h^L(\Gamma(n, c, z, g, \pi))] \rightarrow 1. \)

**Proof:** Assume the hypothesis. For each \( n \) in the selected PBE to \( \Gamma(n, c, z, g, \pi) \) the probability that an individual message corresponds to the state is

\[
q = zg + (1 - z)(1 - g). \tag{31}
\]

Since the games form a triangular array of independent draws, we have

\[
\Pr[m^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi))] = \\
\sum_{j = \frac{n+1}{2}}^{n} \binom{n}{j} q^j (1 - q)^{n-j}. \tag{32}
\]
Because \( z, g > \frac{1}{2} \) we have \( q > \frac{1}{2} \) implying that

\[
pr[m^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi))] \to 1.
\] (33)

Since the rule that individual voters use (27) depends on the posterior (19) that uses \( m^L(\Gamma(n, c, z, g, \pi) \) and additional information individual voters will do no worse at choosing the action they want than if the based their decision on just \( m^L(\Gamma(n, c, z, g, \pi) \). We consider sequences of two conditional probabilities. The previous argument implies that

\[
pr[p^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi)) | \theta^L(\Gamma(n, c, z, g, \pi)) = 1] \geq
pr[m^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi))] \] (34)

and thus

\[
pr[p^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi)) | \theta^L(\Gamma(n, c, z, g, \pi)) = 1] \to 1. \] (35)

Similarly,

\[
1 - pr[p^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi)) | \theta^L(\Gamma(n, c, z, g, \pi)) = -1] \geq
pr[m^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi))] \] (36)

and thus

\[
pr[p^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi)) | \theta^L(\Gamma(n, c, z, g, \pi)) = -1] \to 0. \] (37)

But these two statements mean that

\[
pr[p^L(\Gamma(n, c, z, g, \pi)) = h^L(\Gamma(n, c, z, g, \pi))] \to 1. \] (38)

Proposition 4 demonstrates that with truthful messages, full-information majority rule outcomes are chosen with probability approaching one even though the problem does not involve common values. Proposition 5 demonstrates that even when there are not equilibria with truthful messages there are sequences of equilibria in which full information majority rule outcomes are chosen with probability approaching one.
From the perspective of proposition 5 a key feature of the model is that agents have common knowledge about which type is more likely to be held by a majority of voters. If instead there was first a lottery over whether \( z \) took the value \( \frac{1}{4} \) or \( \frac{3}{4} \) and agents did not learn the realization of this lottery, then we could have \( pr[m^L(\Gamma(n, c, z, g, \pi)) = x^L(\Gamma(n, c, z, g, \pi))] \) bounded away from 1. In this case the public messages may not aggregate information about \( x \). Nonetheless, proposition 5 demonstrates that full informational efficiency for large populations does not depend on full revelation. This finding should not be too surprising, given that in the common values setting jury models often achieve large population efficiency (Condorcet Jury theorems of the second type) even though no information transmission occurs.

5 Medium sized groups

Propositions 4 and 5 demonstrate that full-revelation is not necessary for large population efficiency. We now consider the desirability of subtracting members from a deliberative body. For convenience in this section we assume that \( \pi = \frac{1}{2} \) to allow the lower bounds on efficiency used in the proofs of propositions 4 and 5 to be tight. Under the conditions of proposition 2 (\( c \geq \frac{1}{2} \) or \( z = \frac{1}{2} \)) it is not difficult to see that the probability that the majority rule core is reached increases with the group size. This is true since \( pr[m^T(\Gamma(n, c, z, g, \pi)) = x^T(\Gamma(n, c, z, g, \pi))] \) is increasing in \( n \). However, this need not hold in a parameterization that violates the conditions of proposition 2. We construct a simple example with \( c < \frac{1}{2} \) and \( z > \frac{1}{2} \) that demonstrates the pathology, at \( n = 3 \) a truthful equilibria exists but at \( n = 5 \) it does not.

To rule out the trivial cases we assume that (10) is satisfied. Proposition 1 tells us that a truthful equilibria exists with \( n = 3 \) if

\[
c + (1 - c)(1 - z)^2 \geq (1 - c)z^2.
\]

In order to have

\[
c + (1 - c)(1 - z)^2 = (1 - c)z^2
\]

we need

\[
c = \frac{z^2 - (z + 1)^2}{z^2 - (z + 1)^2 + 1}
\]

So for \( z = \frac{3}{4} \) and \( c = \frac{1}{3} \) (which solve this equation) there is a truthful equilibria with \( n = 3 \). However at \( n = 5 \) the relevant necessary condition is

\[
c + (1 - c) \sum_{j=3}^{4} \binom{4}{j} (1 - z)^j z^{4-j} \geq (1 - c) \sum_{j=3}^{4} \binom{4}{j} z^j (1 - z)^{4-j}.
\]
But for $z = \frac{3}{4}$ and $c = \frac{1}{3}$ the left hand side is

$$c + (1 - c) \left[4(1 - z)^3 z^1 + (1 - z)^4 \right] = \frac{47}{128}$$

and the right hand side is

$$(1 - c) \left[4(1 - z)^1 z^3 + z^4 \right] = \frac{63}{128}$$

In this case a truthful equilibria exists with $n = 3$ but not with $n \geq 5$. More generally, when truthful equilibria exist for some but not all population sizes, we have a critical population size at which adding participants is unattractive.

**Proposition 6** If the conditions of proposition 2 are not satisfied then either truthful equilibria do not exist for any $n$ or there exists an $n^*$ such that at $n < n^*$ a truthful equilibria exists and at $n > n^*$ no truthful equilibria exist.

**Proof:** Pick $z, c, \pi$ that don’t satisfy the conditions of proposition 2. By the proof of proposition 2 this means that $\eta(c, z, n) - \mu(c, z, n)$ is eventually bounded below 0. This means that for some $n^*$ if $n > n^*$ then there is not a truthful equilibria. Now either this is true for $n^* = 1$ in which case there are not truthful equilibria for any $n$, or there is a maximal number $n'$ such that $\eta(c, z, n') - \mu(c, z, n') \geq 0$. Since $\eta(c, z, n') - \mu(c, z, n')$ is monotone in $n$, letting $n^* = n'$ establishes the critical $n^*$.

Whether the probability that the majority rule core is reached is higher at $n^*$ or $n^* + 2$ hinges on the parameters. With the symmetry afforded by $\pi = \frac{1}{2}$, in the truthful equilibria agent induced preferences over voting depend only on whether a majority of public messages are $a$ or $b$. The probability that the majority rule core is reached at $n^*$ is

$$pr[m^T(\Gamma(n^*, c, z, g, \pi)) = x^T(\Gamma(n^*, c, z, g, \pi))].$$

This term is

$$\sum_{j=\frac{n^*+1}{2}}^{n^*} \binom{n^*}{j} g^j(1-g)^{n^*-j}. \tag{43}$$

In the non-truthful equilibria, since not all private information about $s$ is public, when the number of $a$ messages is close to the number of $b$ messages a participant that has $s_i$ different than the modal message $m^p$ may still believe that her private signal is correct. This is true because $g \geq q = zg + (1 - z)(1 - g)$,
where the latter measures the quality of public messages as perceived by the participants. However, all participants whose private signals match the modal public message, will have the same preferences over the choices as they did based just on their private signal. This group is a majority. Accordingly, at least a majority of participants will cast votes that correspond to their message. Accordingly, the probability that the non-truthful equilibrium selects the majority rule core corresponds to the probability that a majority of participants send messages that correspond to the state. This term is:

\[
\sum_{j=n^*+2}^{n^*+2} \binom{n^*+2}{j} q^j (1-q)^{n^*+2-j}
\]

(44)

The ordering of (46) and (47) depends on how much smaller is \( q \) than \( g \). As an example where the addition of two people lowers the probability that the majority rule decision is reached we return to the case of \( z = \frac{3}{4} \) and \( c = \frac{1}{3} \). In this case (46) is

\[
[2(1-g)g + (1-g)^2]
\]

(45)

while (47) is

\[
\left[6(1-(\frac{1}{2}g + \frac{1}{4}))^2(\frac{1}{2}g + \frac{1}{4})^2 + 4(1-(\frac{1}{2}g + \frac{1}{4}))^3(\frac{1}{2}g + \frac{1}{4}) + (1-(\frac{1}{2}g + \frac{1}{4}))^4\right].
\]

(46)

At \( g = .6 \) the former is 0.64 and the latter is approximately 0.61 indicating that the smaller group (using a truthful equilibria) performs better than the slightly larger group (for which no truthful equilibrium exists). This example leads to the following conclusion.

**Proposition 7** There are parameterizations which do not satisfy the conditions of proposition 2 for which excluding randomly selected participants increases the likelihood that the deliberative body selects the full-information majority rule core policy.

An immediate question is how long it takes for the statistical effect to swamp the equilibrium effect? In otherwords, while the proposition says that sometimes group performance is better at \( n^* \) then \( n^* + 2 \), the applicability of the result is limited if group performance at \( n^* + 4 \) is better than at \( n^* \). Figure 2 depicts the probability that the best equilibrium selects the majority rule core as a function of group size \( n \) for plausible values of the parameters. By the time \( n \) gets to \( n^* \) the truthful equilibria tends to do quite well. However once \( n \) exceeds \( n^* \) the probability of a core can be dramatically lower and the curve is not very steep.

[Insert Figure 2 here]

If it is finite and non-zero the value \( n^* \) solves the program
\[
n^*(c, z) = \min\{n : \eta(c, z, n) - \mu(c, z, n) \geq 0\}
\]

The inequality is of the form

\[
\eta(c, z, n) - \mu(c, z, n) = c + (1 - c) \left[ \sum_{j=\frac{n+1}{2}}^{n-1} \binom{n-1}{j} \left[ (1 - z)^j z^{n-1-j} - z^j (1 - z)^{n-1-j} \right] \right]
\]

(48)

Since this difference is increasing in \(c\) and decreasing in \(n\) if \(c' > c\) then \(n^*(c', z) \geq n^*(c, z)\). Moreover, since the difference is decreasing in \(z\) if \(z' > z\) then \(n^*(c, z') \leq n^*(c, z)\). This leads to the following conclusion

**Proposition 8** The critical value, \(n^*(c, z)\), is increasing in \(c\) and decreasing in \(z\).

Figure 3 plots the function \(n^*(c, z)\). Note that \(n^*\) does not depend on \(g\). Several conclusion are worth emphasizing. For values of \(z\) greater than 0.6, \(n^*\) is less than 50 and if \(c < 0.3\) then \(n^*\) is less than 50. Moreover, for most parameterizations \(n^*\) is less than 25. So for problems in which preference heterogeneity is relatively likely \(n^*\) is fairly small (smaller than the capacity of New-England town halls).

We have seen that deliberation at \(n^* + 2\) can fair worse than deliberation at \(n^*\) but of course this is not always the case. Figure 4 demonstrates the pervasiveness of the finding that smaller deliberative bodies can be better. Interestingly, the prevalence of environments in which group performance is non-monotonic in \(n\) is higher for hard problems in which individual signals about the state are of low quality (\(g\) closer to \(\frac{1}{2}\)).

**6 Discussion**

The analysis offers some good news for deliberative democrats. It is not necessary that everyone know \textit{a priori} that their is commonality of interests for strategic participants to reveal their private information. Individuals believing that the potential for saboteurs (or just participants with very different interests) is non-trivial may still be willing to reveal their private information. Truthful deliberation is possible as long as each participant believes that more than half of the other participants will have the same preferences as her. As such, heterogeneity is not a problem if participants are confident that most others share their tastes. This finding suggests that the “egocentrism bias” (Kruger and Burrus 2003; Windshitl, Kruger, Simms 2003) of overestimating the likelihood that that others are similar to oneself, may foster effective information sharing.
in settings where equilibria (with unbiased beliefs) preclude truthful behavior. Second, even when some types of participants will believe that they are in a minority, so truthful deliberation is not supportable, in large enough settings their exist equilibria which reach decisions that are nearly certain to select the full information majority rule outcome. This finding should be tempered by the observation that without any communication, full information majority rule decisions may also be reached in the limit. However, deliberation need not be an effective means of aggregating information. In particular, in deliberative settings in which agents do not have strong a priori beliefs that their is commonality of values participants may purposely deceive each other and participants’ beliefs may not be very responsive to what others know and say.

Most interestingly, more participants need not be better. In some cases truthful equilibria exist for a small deliberative body and adding an additional few individuals can destroy the truthful equilibria resulting in less efficient decisions. In addition the number of additional agents needed for the Condorcetian statistical effect to offset the loss of effective information transmission can be quite large. Loosely stated, we find that for small deliberative bodies, truthful deliberation is easiest to support, but that in very large bodies efficient decisions are reached even without truthful discourse. For medium sized bodies removal of enough participants to restore truthful equilibria can be desirable. A promising avenue for future work is the consideration of alternative institutional arrangements which foster information sharing in problems of choice with private beliefs and values.

\[11\] This finding may be used to develop an evolutionary explanation for the persistence of this bias.
We now establish a version of the revelation principal equating the existence of truthful equilibria in the binary message game to separating equilibria in games with a larger message space. Let $\Gamma(n, c, z, g, \pi)$ denote a deliberation game with parameters $n, c, z, g, \pi$ and let $\Gamma^\Pi(n, c, z, g, \pi)$ denote a modified deliberation game with parameters $n, c, z, g, \pi$ and a larger message space, $\Pi$, which contains at least 2 messages.

**Proposition 9** A truthful PBE in weakly undominated voting exists in $\Gamma(n, c, z, g, \pi)$ iff a separating PBE in weakly undominated voting exists in $\Gamma^\Pi(n, c, z, g, \pi)$.

**Proof:** ($\Rightarrow$) Assume that $\Gamma^\Pi(n, c, z, g, \pi)$ has a separating PBE in weakly undominated voting. Since the equilibrium is separating, consistency of beliefs requires that on the path agents believe that $s = (s_1, \ldots, s_n)$ is public information prior to voting. Now since the message mapping must be invertible, and all types are possible, any agent can send a message that leads the others to believe that she has a signal different than $s_i$. Accordingly, since a separating equilibrium exists it must be the case that either (10) is satisfied or $\eta(c, z, n) \geq \frac{1}{2}$. But by proposition 1 this means that a truthful equilibrium in weakly undominated strategies exists in $\Gamma(n, c, z, g, \pi)$.

($\Rightarrow$) Now suppose that a truthful equilibrium in weakly undominated strategies exists in $\Gamma(n, c, z, g, \pi)$. Since weakly undominated voting requires that voting decisions are based only on posterior assessments of $\Pr(x = a \mid m)$ we can construct a separating equilibrium in $\Gamma^\Pi(n, c, z, g, \pi)$ from the truthful equilibrium in $\Gamma(n, c, z, g, \pi)$. Consider an arbitrary partition of $\Pi = \{A, B\}$. Since $\Pi$ contains at least 2 messages such a partition exists. Consider the mapping $\psi_i : \{-1, 1\} \times \{a, b\} \to \Pi$ such that $\psi_i(\theta_i, a) \in A$ and $\psi_i(\theta_i, b) \in B$. Now let

$$\Pr(x = a \mid \psi_1, \ldots, \psi_n) = \Pr(x = a \mid A(m), B(m))$$

where $A(m) = \{i : m_i \in A\}$ and $B(m) = \{i : m_i \in B\}$. Finally, let beliefs in $\Gamma^\Pi(n, c, z, g, \pi)$ be measurable with respect to the variables $A(m)$ and $B(m)$. This requires that any deviation from the strategy $\psi_i(\theta_i, b)$ will only be desirable if a deviation from the truthful strategy was desirable in the truthful equilibrium to $\Gamma(n, c, z, g, \pi)$. Since truthful messages are a best response in $\Gamma(n, c, z, g, \pi)$ and only the inference about $A(m)$ and $B(m)$ is payoff relevant under weakly undominated voting it must be the case that $\psi_i(\cdot, \cdot)$ is a best response in $\Gamma^\Pi(n, c, z, g, \pi)$. ■
References


Fig. 1: Space of Environments

- Common Knowledge of Common Values
- Truthful equilibria exist for all $n$

- Not Robust
- Robust
Figure 2

Probability of Selecting a Majority Rule Core as a Function of \( n \), where \( z=0.59, c=0.315, n^*=21, g \) varies from 0.6 to 0.9
n* does better than n*+2

n*+2 does better than n*

n*+2 does better than n*