A Mathematical Model For Optimal Decisions In A Representative Democracy

Malik Magdon-Ismail
RPI
magdon@cs.rpi.edu

Lirong Xia
RPI
xial@cs.rpi.edu

Abstract

Direct democracy is a special case of an ensemble of classifiers, where every person (classifier) votes on every issue. This fails when the average voter competence (classifier accuracy) falls below 50%, which can happen in noisy settings where voters have only limited information, or when there are multiple topics and the average voter competence may not be high enough for some topics. Representative democracy, where voters choose representatives to vote, can be an elixir in both these situations. Representative democracy is a specific way to improve the ensemble of classifiers. We introduce a mathematical model for studying representative democracy, in particular understanding the parameters of a representative democracy that gives maximum decision making capability. Our main result states that under general and natural conditions,

1. Representative democracy can make the correct decisions simultaneously for multiple noisy issues.
2. When the cost of voting is fixed, the optimal representative democracy requires that representatives are elected from constant sized groups: the number of representatives should be linear in the number of voters.
3. When the cost and benefit of voting are both polynomial, the optimal group size is close to linear in the number of voters.

This work sets the mathematical foundation for studying the quality-quantity tradeoff in a representative democracy-type ensemble (fewer highly qualified representatives versus more less qualified representatives).

1 Introduction

Suppose a voter-population of size n must vote in a referendum to make an important binary decision to optimize some objective, e.g. social welfare growth over 10 years. A typical solution is *direct democracy* which decides based on a majority vote, the so called "wisdom of the crowd." Direct democracy works when, the crowd does indeed have wisdom. In reality, the voters cannot directly observe which decision is correct. Instead, they form beliefs using perceived information, which can be inaccurate, misinterpretable or even manipulated. For example, suppose that each voter's chance to vote for the correct decision, her *competence*, is i.i.d. generated uniformly over $[0, 0.99]$. Now, the majority among many voters makes the wrong decision with near certainty (Fey, 2003).

This example highlights a critical flaw of direct democracy, where voters participate in decision-making irrespective of their competence. Direct democracy is particularly problematic in high noise situations, where there is often a close tie in people's beliefs between two choices, as in the previous example. Unfortunately, such close ties are common in real-life high-stakes scenarios. For example, in the 2016 United

[0]Rensselaer Polytechnic Institute, $\{$magdon,xial$\}$@cs.rpi.edu

Kingdom European Union membership referendum, 51.89% voted for leave and 48.11% voted for remain. In the 2016 US Presidential Election, 46.1% voted for Trump and 48.2% voted for Clinton.[1] How can democracy cope with high-stakes noisy issues where the average voter competence may drop below 0.5, especially if there is misinformation?

One promising rescue is *representative democracy*, where voters form groups (or districts). Each group chooses a representative, and representatives decide via a majority vote. The tradeoff is that there are fewer representatives than base-voters, but, in return, each representative is (hopefully) better informed, being the "wisest" from its group, or at least having a higher competence than the average in its group members. Continuing the example above, let us now use representative democracy, where people are divided into households (e.g. 5 people per group), and let each group choose the member with highest competence as its representative. Then, with high probability the representative's competence is strictly larger than 0.5. Now, with enough representatives, a majority vote will now make the correct decision with near certainty (Fey, 2003).

Many countries and organizations adopt a mixture of direct democracy and representative democracy. In the US, voters in each state vote on multiple referenda (direct democracy). In addition, voters elect members of the congress (representative democracy) to make decisions for everyone by voting on bills. Representative democracy can have a *fixed number of representatives* regardless of the population (e.g. the US senate has 100 members, two from each state), or a *fixed group size*, where the number of representatives is proportional to the population (between 1789 and 1913, the US House increases from 65 to 435 members based on the nation's population growth[2]). While it is widely accepted that representative democracy is efficient due to lower operational cost and has better turnout than direct democracy, there is still a large debate on the following fundamental question: *which type of democracy makes better decisions?*

We are not aware of quantitative answer to this question, nor a mathematical framework for analyzing representative democracy w.r.t. its capability to make correct decisions. This is in sharp contrast to direct democracy, which has been mathematically analyzed in depth to provide a justification of the "wisdom of the crowd," which dates back to the *Condorcet Jury Theorem* in the 18th century (Condorcet, 1785). Roughly speaking, the Jury Theorem states that a large group of voters are likely to make a correct decision by majority voting, which "*lays, among other things, the foundations of the ideology of the democratic regime*" (Paroush, 1998). Direct democracy is just a representative democracy where each group has one voter. Thus, a mathematical characterization of optimal representative democracies would also highlight the subcases where direct democracy is best. The goal of this paper is to establish rigorous mathematical foundations of representative democracy, and provide quantitative answers to the following key questions.

Key questions.

1. What is the optimal number of representatives for representative democracy?
2. How should one optimally divide the voters into groups, with each group electing one representative? Specifically, what is the optimal size distribution for the groups?

We will answer these questions in a general setting, where the groups satisfy a weak form of homogeneity. At a high level, the representative election process can be described by a function μ that takes a group as input and outputs its representative. By homogeneous, we mean that this representative election function is the same in every group. A concrete example of this paradigm is when each group runs the same type of election process on its members who are independent and drawn from some underlying voter-distribution.

In our analysis, we consider two cases. The first is when there is a fixed cost for the voting. In this case, the goal of the representative democracy is to maximize the chances of making the correct decision.

[1] These examples are only used to show real situations where close ties exist. We do not know if direct democracy would succeed or fail in these cases since we do not know what the "correct" outcome is.

[2] After 1913 the number is fixed to 435.

This is case is relevant to making extremely important decisions where the operational cost of voting is not considered a valid tradeoff for correctness. The second case is when the cost of voting increases with the number of representatives, in which case one must balance the cost with the benefit that accrues to all n voters.

1.1 Our Contributions

We provide a novel mathematical model of representative democracy w.r.t. its ability to make correct decisions. With our model, we obtain characterizations on optimal number of representatives in representative democracy, and our main messages are the following.

1. When the cost of voting is fixed, fixed group size is optimal.
2. When the cost and benefit of voting are both polynomial, $O(\log n)$ representatives are optimal, where n is the number of voters.

In our basic model, there is a single binary issue to be decided. n voters are divided into L groups, each group chooses a representative by a *representative selection function*, and the representatives will use majority voting to make a binary decision on the issue. We assume that each voter's type is characterized by her *competence*, which is her probability to vote for the correct decision on the issue, and is generated i.i.d. from a distribution F. Let $\mathrm{Ben}(n) \in \mathbb{R}$ denote the benefit of making the correct decision compared to making the wrong decision for n voters, and let $\mathrm{Cost}(L) \in \mathbb{R}$ denote the operational cost for L representatives to vote.

We reduce the representative selection function to a *group competence function* $\mu : \mathbb{N} \to [0, 1]$, which maps each group size to the expected competence of the representative. We then extend the Condorcet Jury Theorem to representative democracy by characterizing group competence functions that would choose representatives to make the correct decision with probability 1 as $n \to \infty$. Our main theorems are surprising characterizations of the optimal group size when the cost of voting is fixed.

Theorems 2, 3 and 4 (optimal representative democracy for single issue with fixed cost of voting, informally put). Under natural and mild conditions on the group competence function μ and when $\mathrm{Cost}(L)$ is a constant:

1. (Homogeneous groups) The optimal group size $K^*(n)$ is at most a constant independent of n.
2. (Inhomogeneous groups) The optimal number of representatives L^* is linear in n.
3. The optimal group size distribution is nearly homogeneous.

These results hold, independently of the specific details of the representative selection process. Let us highlight why the result is unexpected in a concrete context. Suppose voter competence is drawn from some continuous density on $[0, 1]$ and a group elects the most competent of its voters as representative (very optimistic since it cannot get better than that). Then by choosing larger and larger groups, the competence of the representative approaches 1. The price paid is that there are fewer of these ultra-smart representatives, but there will still be many of them as $n \to \infty$. Indeed, one might posit that some optimal tradeoff exists whereby the group size tends to infinity but at a slower rate than n so that the representative gets smarter and smarter *and* there are more and more representatives. Our theorem establishes the contrary. The optimal group size will never exceed some constant (the constant's value may depend on the specific parameters of the selection process).

To prove our results, we use novel combinations of combinatorial bounds. In addition, some of our results may be of independent interest, for example, Lemma 7 answers an open question on the probability for majority voting to be correct when the average competence of voters is exactly 0.5 where there are no general results for the non-asymptotic behavior (Fey, 2003) and the asymptotic behavior is only conjectured (Owen et al., 1989, Lemma 5).

We then consider the case of polynomial cost and polynomial benefit, that is, there exist constants $q_1 > 0$ and $q_2 > 0$ such that $\text{Cost}(L) = L^{q_1}$ and $\text{Cost}(n) = n^{q_2}$. A special case is linear cost and linear gain, where $q_1 = q_2 = 1$. It turns out that the cost of voting has a significant impact on the optimal group size.

Theorems 5 (optimal representative democracy for single issue with polynomial cost and polynomial benefit, informally put). Under natural and mild conditions on the group competence function μ and suppose there exist constant $q_1 > 0$ and $q_2 > 0$ such that $\frac{\text{Cost}(L)}{\text{Ben}(n)} = \Theta(\frac{L^{q_1}}{n^{q_2}})$:

1. The optimal homogeneous group size is $\Omega(n/\log n)$.
2. When $\mu(K)$ converges to 1 at a polynomial rate, the optimal homogeneous group size is $\Theta(n/\log n)$
3. If $\mu(K)$ is upper bounded away from 1, then the optimal group size is $\Theta(n)$.

Finally, we extend our analysis to the situation where the representative will have to vote on $d > 1$ possibly correlated issues. This is the case with the US Senate and House members who after election may have to cast a vote on multiple occasions on different issues (about once per week). Now, the group competence function μ must be extended to a mapping from group size K to a 2^d-dimensional multinomial distribution. We extend the Condorcet Jury Theorem to this setting, and our main theorem states that the optimal group size is increased by $\Theta(\ln d)$.

Theorem 7 (optimal representative democracy for multiple issues with fixed cost, informally put). Under natural and mild conditions on the group competence function, the optimal group size is $O(\ln d)$ and the optimal number of representatives is $\Omega(n/\ln d)$ for $d \geq 2$ issues.

With multiple issues, there is no obvious way to elect representatives to achieve consistency in *every issue*. For example, when issues are independent, if a group picks its representative according to its favorite (or random) issue, the results can be disastrous. We introduce a natural representative selection process for multiple issues, the MAX SUM process, which chooses the group member with the maximum sum of competences across all issues. This mimics a small group (e.g. a household) choosing its most "informed" member. We show that under natural assumptions and for independent issues, MAX SUM is consistent on *all issues*.

1.2 Related Work and Discussions

There have been large voices in the US in favor of increasing the number of House representatives, as nowadays each of them represents roughly 700K people, and such number used to be as low as 33K. The arguments are mostly from the perspective of representation rights and the intention of framers of the Constitution and the Bill of Rights (Flynn, 2012; Bartlett, 2014; Humphreys, 2016). More information and activities can be found at `http://www.thirty-thousand.org`. Our work provide a mathematical foundation to analyze the rationale behind this move.

While mathematical models of representative democracy are not new, e.g. (Besley and Coate, 1997; Auriol and Gary-Bobo, 2012), we are not aware of previous work that quantitatively characterizes optimal representative democracy w.r.t. its ability to make correct decisions. Our work is related to the literature on extensions of The Condorcet Jury Theorem to heterogeneous agents, where voters may have different competences (Nitzan and Paroush, 1980; Grofman et al., 1983; Nitzan and Paroush, 1984). More previous work along this vein is in one of the three subareas: (1) understanding the conditions for consistency of majority voting (Owen et al., 1989; Paroush, 1998; Kanazawa, 1998; Fey, 2003; Sapir, 2005), (2) studying optimal population size to maximize the correctness of majority voting (Feld and Grofman, 1984; Miller, 1986; Gradstein and Nitzan, 1987; Paroush and Karotkin, 1989; Berend and Paroush, 1998; Ben-Yashar and Paroush, 2000; Mukhopadhaya, 2003; Karotkin and Paroush, 2003; Berend and Sapir, 2005, 2007; Stone and Kagotani, 2013; Ben-Yashar and Zahavi, 2011), and (3) incentivizing voters to increase their competence (Nitzan and Paroush, 1980; Karotkin and Paroush, 1995; Ben-Yashar and Paroush, 2003). See Nitzan and Paroush (2017) for a recent survey, including a nice overview of extensions of the Condorcet Jury theorem to dependent voters and strategic voters.

Our work on single issue is related to the first subarea (consistency) and the second subarea (optimal size). The key differences are, first, in our work, the competence of the representatives are computed endogenously as a result of partitioning and representative selection, while the competence of voters are given as input in the first subarea above. Second, in our work, increasing the group size will reduce the number of groups, thus may reduce the correctness of majority voting, even though each representative has a higher competence, while in the second subarea above there is one group of voters with variable size, who directly vote on the issue. In other words, our framework allows for a quantitative analysis of the quality vs. quantity tradeoff in representative democracy. Our work on multiple issues is significantly different from previous work.

Our work is related but different to two recent works on liquid democracy (a.k.a. delegative democracy). Cohensius et al. (2017) studied proxy voting, where a set of proxies are selected either randomly or voluntarily, and each voter chooses a proxy to vote for her. The authors compared the accuracy of proxy voting theoretically and experimentally, and identify conditions for proxy voting to be superior than direct democracy. Kahng et al. (2018) studied whether delegation mechanisms on networks can improve direct democracy w.r.t. the probability to reveal the ground truth. The main difference between our work and the two paper is the dynamic of decision-making process, and therefore, the main questions are different. In our work (representative democracy), voters are divided into multiple groups, and each group chooses a representative to vote for the group members. In the work of Cohensius et al. (2017), proxies are chosen first, and are then weighted by their popularity. In the work of Kahng et al. (2018), each voter can delegate her vote and all votes delegated to her, to another voter with higher competence. We note that operational cost of voting was also not considered in (Cohensius et al., 2017) and (Kahng et al., 2018).

Our work is also related to recent research in computational social choice that uses statistics to make better decisions (Conitzer and Sandholm, 2005; Caragiannis et al., 2016; Xia and Conitzer, 2011; Young, 1988; Procaccia et al., 2012; Pivato, 2013; Elkind and Shah, 2014; Azari Soufiani et al., 2014; Xia, 2016). These works focus on direct democracy, not representative democracy.

Lastly, our results on multiple issues is related to multi-issue voting (Koriyama et al., 2013; Skowron, 2015; Lang and Xia, 2016; Conitzer et al., 2017). The main difference is that the consideration in multi-issue voting is often fairness and representativeness, rather than making correct decisions.

2 Modeling Representative Democracy

In this section we propose a mathematical model for representative democracy for one issue. We will extend it to multiple issues in Section 4. As in the Condorcet Jury Theorem, we assume that voters' ability to vote for the correct decision is drawn i.i.d. from a distribution F. The voters are divided into $L \geq 1$ groups. For any group ℓ with K voters, whose competences are $\{q_{\ell,1}, \dots, q_{\ell,K}\}$, a deterministic or randomized representative selection process V_ℓ chooses a member q_ℓ. The chosen representative casts a vote, and majority voting succeeds if strictly more than half of representatives vote 1. The process for a group to choose a representative to vote is summarized below.

$$F \xrightarrow{\text{generate voters}} \{q_{\ell,1}, \dots, q_{\ell,K}\} \xrightarrow{\text{elect representative}} q_\ell \xrightarrow{\text{cast vote}} x_\ell$$

Formally, our basic model is defined as follows.

Definition 1. *A representative democracy for one issue is composed of the following component.*

- **Issue.** *Suppose there is one issue to decide, whose outcome is* 1 *(correct), and* 0 *incorrect.*

• **Partition function** $\vec{K}$. *For any number of voters* n*,* $\vec{K}$ *denote a partition function that divides* n *voters into* $L(n)$ *groups, that is,* $\vec{K}(n) = (K_1, \ldots, K_{L(n)}) > 0$ *and* $\sum_{i=1}^{L(n)} K_i = n$*.*

• **Distribution of competence** F. *We assume that each voter's* type *is characterized by her* competence*, which is the probability for her to vote for* 1*. Each voter's competence is i.i.d. generated from a distribution* F *over* $[0, 1]$.

• **Representative selection process** V_ℓ. *Suppose group* ℓ *has* K *users whose competences are* $q_{\ell,1}, \ldots, q_{\ell,K}$ *respectively. The group uses a (randomized) process* $V_\ell(q_{\ell,1}, \ldots, q_{\ell,K})$ *to select a representative* q_ℓ*, whose vote is represented by a random variable* x_ℓ*.*

• **Voting by the representatives.** *The chosen representatives will vote for* 1 *with probability equivalent to her competence, and vote for* 0 *otherwise. The majority rule is used to decide the outcome based on representatives' votes. We assume that a strict majority of votes is necessary to make the correct decision. That is, when there is a tie, the outcome is* 0 *(incorrect).*

When each group has exactly 1 member, we obtain the direct democracy, otherwise we have the representative democracy as in the following example.

Example 1 (Uniform Voters with Uniform Process)**.** Suppose there are $L \geq 1$ groups, each group has $K \geq 1$ members. Each voter's competence is i.i.d. generated from a uniform distribution $F = \text{Uniform}[a, b]$. For all groups, V_ℓ chooses a member uniformly at random. It is not hard to see that majority voting by the representatives is correct with probability no more than 0.5 when $a + b < 1$. ■

We note that the vote of group ℓ's representative, i.e. the random variable x_ℓ, is characterized entirely by its expectation, which contains all information needed in the analysis in this paper. Therefore, we will simplify the representative selection process to a single *group competence function* μ, which specifies the expected competence of the representative as a function of the group size K.

Definition 2. *For one issue, a representative democracy is a* selection function $\mu : \mathbb{N} \mapsto [0, 1]$.

For example, the selection function for the uniform process in Example 1 can be represented by $\mu_{\text{U}}(K) = (a + b)/2$ for all K. We will see that the selection function significantly simplifies the process in the following two examples. In the next example, the group chooses its most-informed member—the one with the highest competence—as the representative.

Example 2 (Max Process)**.** As in Example 1, suppose $F = \text{Uniform}[a, b]$. Each group now chooses the member with the highest competence as the representative. So $q_\ell = \max\{q_{\ell,1}, \ldots, q_{\ell,K}\}$. In other words, q_ℓ is the K-th order statistic of K uniformly distributed random variables, which means that $\frac{q_\ell - a}{b - a}$ is a Beta random variable with PDF $\text{Beta}(K, 1)$ Gentle (2009), whose mean is $\frac{K}{K+1}$. We call this representative democracy the MAX process, denoted by $\mu_{\max}$. Therefore,

$$\mu_{\max} = \mathbb{E}[x_\ell] = \frac{K}{K+1} \times (b - a) + a = \frac{1}{K+1} a + \frac{K}{K+1} b$$

■

The MAX process serves as an upper bound on group competence functions, as no other process can select a voter with a higher competence. Often the competence is not directly observable as in Example 2. However, when the group size is not too large, for example a group is a household, then it is natural to assume that the family members are able to choose the max-informed representative. Nevertheless, this process might be noisy, which is captured in the next example.

Example 3 (Noisy-Max Process)**.** As in Example 1, suppose $F = \text{Uniform}[a, b]$. Given a K-dimensional vector $\vec{p} = (p_1, \ldots, p_K)$ such that $p_1 \cdots \geq p_K \geq 0$ and $\sum_{i=1}^K p_i = 1$. For any $i \leq K$, suppose the group chooses the member with i-th highest competence as the representative with probability p_i. Let $q_{\ell,i}$ denote the i-th highest competence. It follows that $\frac{q_{\ell,i}-a}{b-a}$ is a Beta random variable with PDF $\text{Beta}(K+1-i, i)$ Gentle (2009), whose mean is $\frac{K+1-i}{K+1}$. We call this representative democracy the NOISY-MAX process, denoted by $\mu_{\vec{p}}$. Therefore,

$$\mu_{\vec{p}}(K) = \sum_{i=1}^{K} p_i \left(\frac{i}{K+1} a + \frac{K+1-i}{K+1} b \right) = (\vec{p} \cdot \vec{k}_+) a + (\vec{p} \cdot \vec{k}_-) b,$$

where $\vec{k}_+ = (\frac{1}{K+1}, \frac{2}{K+1}, \ldots, \frac{K}{K+1})$ and $\vec{k}_- = (\frac{K}{K+1}, \frac{K-1}{K+1}, \ldots, \frac{1}{K+1})$. ■

It is easy to verify that the group competence function for MAX process is monotonically increasing in K and is between a and b, while the group competence function for UNIFORM process outputs the same value for all K.

3 Optimal Representative Democracy for One Issue

In this section, we focus on the characterizing optimal representative democracy for one issue.

3.1 Consistent Representative Democracy

We first extend the classical Condorcet Jury Theorem to representative democracy. Let us first formally define consistency, the main desired property of a representative democracy. As the number of voters increases, i.e. asymptotically in n, it should be possible to choose a partition of n voters into L groups, with potentially different number of voters in each group, such that with probability 1 the majority representatives vote for 1.

Given $\vec{K} = (K_1, \ldots, K_{L(n)})$, we let $S_{n,\vec{K},\mu}$ denote the random variable that represents the fraction of 1's in $L(n)$ independent Bernoulli random variables with success probabilities $(\mu([\vec{K}(n)]_1), \ldots, \mu([\vec{K}(n)]_{L(n)}))$, where for any $i \leq L(n)$, $[\vec{K}(n)]_i = K_i$ is the i-th component of $\vec{K}(n)$. In other words, $S_{n,\vec{K},\mu}$ is $\frac{1}{L(n)}$ of the Poisson trail that represents the representatives' votes.

Definition 3. *Given a partition function $\vec{K}$ and a group competence function μ, for any n, we let $R_n(\vec{K}, \mu)$ denote the probability for majority voting by representatives according to $\vec{K}$ and μ to succeed. That is, $R_n(\vec{K}, \mu) = \mathbb{P}(S_{n,\vec{K},\mu} > \frac{1}{2})$.*

The subscript n in R_n is sometimes omitted when causing no confusion.

Definition 4. *A representative democracy with group competence function μ is* consistent *if there exists a partition function $\vec{K}(n)$ for which the majority voting of representatives succeeds with probability 1 as $n \to \infty$, that is,* $\lim_{n\to\infty} R_n(\vec{K}, \mu) = 1$.

Theorem 1. *For each distribution F, the representative democracy with group competence function μ is consistent if and only if there exists $K_* \in \mathbb{N}$ such that $\mu(K_*) > 0.5$.*

Proof. $\Rightarrow$: suppose for the sake of contradiction, $\mu(K_*) \leq 0.5$ but there exists $\vec{K}$ such that $\lim_{n\to\infty} R_n(\vec{K}, \mu) = 1$. It follows that for any n, $R_n(\vec{K}, \mu)$ is no more than the average value of $L(n)$ Bernoulli trials, each succeeds with probability 0.5. However, the probability for the latter to be strictly larger than 0.5 is no more than 0.5, because its PDF is symmetric and its mean is 0.5. This leads to a contradiction .$\Leftarrow$: we choose the partition function such that all but one group have K_* members. It follows that as $n \to \infty$ the average competence is at least $\frac{n/K_*-1}{n/K_*+1}\mu(K_*) > \frac{1}{2}$. This means that $R_n(\vec{K}, \mu) = 1$ as n goes to infinity (Fey, 2003). $\square$

Costly voting. We now formally define the benefit of correct decision, cost of voting, and the social welfare for optimization.

Definition 5. *For any n and L, let $Ben(n) \in \mathbb{R}$ denote the benefit of making the correct decision and let $Cost(L)$ denote the monetary cost of maintaining L representatives. Given a partition $\vec{K} = (K_1, \ldots, K_{L(n)}) \in \mathbb{N}^{L(n)}$, the social welfare of $\vec{K}$ is the expected benefit minus the cost of voting, that is,*

$$SW(\vec{K}) = Ben(n)R_n(\vec{K}, \mu) - Cost(L) = Ben(n)\left(R_n(\vec{K}, \mu) - \frac{Cost(L)}{Ben(n)}\right)$$

In the following subsections we will characterize optimal $\vec{K}$ that maximizes SW$(\vec{K})$ for different cases.

3.2 Optimal Group Size for Single Issue: Fixed Cost of Voting

In this subsection we focus on the setting where the cost of voting is fixed regardless of the number of representatives. Since the cost of voting is fixed, the goal is to find the optimal partition $\vec{K}$ that maximizes $R_n(\vec{K}, \mu)$.

Before we formally present our main results, let us first discuss the effect of increasing the size of groups at a high level. The effect can be seen as a tradeoff between quality (competence of each representative) vs. quantity (the total number of representatives), and we will see that it comes down to a form of mean vs. variance tradeoff.

The quality vs. quantity tradeoff. Suppose n is fixed and we are deciding to use group size of K_1 or K_2, where $K_1 < K_2$ and $\mu(K_2) \geq \mu(K_1) > 0.5$, and for the purpose of presentation, suppose that both divides n. That is, $n = K_1L_1 = K_2L_2$. Let $\vec{K}_1 = (\underbrace{K_1 \ldots, K_1}_{L_1})$ and $\vec{K}_2 = (\underbrace{K_2 \ldots, K_2}_{L_2})$.

When there are K_1 members in each group, each representative has competence $\mu(K_1) < \mu(K_2)$. On the other hand, the number of representatives is $L_1 > L_2$. Therefore, we have $\mathbb{E}(S_{n,\vec{K}_1,\mu}) = \mu(K_1) \leq \mu(K_2) = \mathbb{E}(S_{n,\vec{K}_2,\mu})$, while the variance of $S_{n,\vec{K}_2,\mu}$, which is $\mu(K_1)(1 - \mu(K_1))/L_1$ can potentially be smaller than the variance of $S_{n,\vec{K}_2,\mu}$.[3] Therefore, the optimal group size K^* minimizes the left tail probability of $S < 0.5$. This can roughly be seen as a mean vs. variance tradeoff, as illustrated using the distribution of $S_{n,\vec{K},\mu}$ in Figure 1, where we set $n = 2000$, $F =$ Uniform$[0.44, 0.55]$, the MAX process μ_{max} is used, and we compare the distribution of three group sizes: $K = 2$ (low mean, low variance), $K^* = 8$ (optimal, middle), and $K = 20$ (high mean, high variance).

Naturally, our next goal is to identify an optimal number of representatives (groups), denoted by $L^*(n)$. We will first focus on the specific partition functions where almost all groups have the same size, with the exception for the last group, whose size is allowed to be larger than other groups. This is a natural setting in practice because each representative is supposed to represent equal number of voters. Later in this section we will show how to extend our study to general partition functions.

Definition 6 (Homogeneous groups)**.** *Given n and a group size K, in the homogeneous setting, $L = \lfloor n/K \rfloor$ representatives are selected using partition function $\vec{H}_K = (\underbrace{K, \ldots, K}_{L-1}, n - (L-1)K)$.*

[3]It is possible for $S_{n,\vec{K}_2,\mu}$ to have a smaller variance, if $\mu(K)$ grows faster than linear.

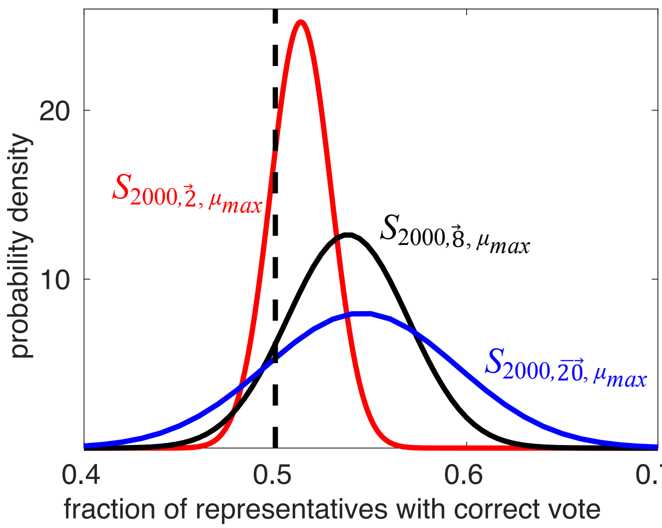

Figure 1: The mean vs. variance tradeoff in choosing different group sizes.

Next, we prove a general result: among all homogeneous group size settings, the optimal group size is bounded by a constant, provided that some value of K can achieve consistency and the group competence function μ is polynomially bounded away from 1 as K goes to infinity. Let $K^*_{\text{hom}}(n)$ denote the optimal homogeneous group size that maximizes $\text{SW}(\vec{H}_K)$, where $\text{Cost}(L)$ is a constant. In other words, $K^*_{\text{hom}}(n)$ maximizes $R_n(\vec{H}_{K^*_{\text{hom}}(n)}, \mu)$, the probability for the majority of representatives to vote for the correct decision.

Theorem 2 (Optimal homogeneous group size)**.** *Suppose the Cost*(L) *is a constant, there exists* $K_* \in \mathbb{N}$ *such that* $\mu(K_*) \geq \frac{1}{2} + \epsilon$ *for* $0 < \epsilon < \frac{1}{2}$*, and for all* $K \in \mathbb{N}$*,* $\mu(K) \leq 1 - A/K^\alpha$ *for constants* $A > 0$ *(w.l.o.g.* $A \leq 1$*) and* $\alpha \geq 0$*. Then,* $K^*_{hom}(n) \leq cK_*$*, where*

$$c = \frac{-4}{\ln(1-4\epsilon^2)} \left(\ln \frac{32}{9\epsilon^2 A} + \alpha \ln \frac{4\alpha K_*}{e} - \alpha \ln |\ln(1-4\epsilon^2)| \right).$$

The intuition is that as K increases, we experience diminishing returns with respect to μ because μ is bounded away from 1. On the other hand, there is a loss due to decreasing $L = \lfloor n/K \rfloor$, the number of representatives. We find this constant bound surprising because one may expect that the best tradeoff is achieved when K is a function of n, but our theorem proves that it is a constant.

Proof. (of Theorem 2) We first prove a series of lemmas as building blocks. W.l.o.g. suppose each of the first $L-1$ groups has exactly K members. We first observe an elementary bound that allows us to ignore the last group whose number of representatives is unknown,

$$\underbrace{\mathbb{P}\left[\sum_{\ell=1}^{L-1} x_\ell \geq \left\lceil \frac{L+1}{2} \right\rceil\right]}_{R_-(L,\mu(K))} \leq R_n(\vec{H}_K, \mu) \leq \underbrace{\mathbb{P}\left[\sum_{\ell=1}^{L-1} x_\ell \geq \left\lceil \frac{L-1}{2} \right\rceil\right]}_{R_+(L,\mu(K))}.$$

The functions R_- and R_+ are Binomial upper tail probabilities,

$$R_-(L,p) = \sum_{\ell=\lceil \frac{L+1}{2} \rceil}^{L-1} \binom{L-1}{\ell} p^\ell (1-p)^{L-1-\ell}.$$
$$R_+(L,p) = \sum_{\ell=\lceil \frac{L-1}{2} \rceil}^{L-1} \binom{L-1}{\ell} p^\ell (1-p)^{L-1-\ell}.$$

For any n, K, μ, and $L = \lfloor n/K \rfloor$, we have $R_-(L, \mu(K)) \le R_n(\vec{H}_K, \mu) \le R_+(L, \mu(K))$. The following five lemmas are properties of $R_-(L,p)$ and $R_+(L,p)$. So far, we have not yet invoked any properties of the group competence function μ.

Lemma 1 (Monotonicity of $R(L,p)$)**.** *For fixed L, $R_-(L,p)$ and $R_+(L,p)$ are increasing in p. For fixed $p > \frac{1}{2}$, $R_-(L,p)$ and $R_+(L,p)$ are increasing in L.*

Lemma 2. *For $p \le \frac{1}{2}$, $R_+(L,p) \le \frac{3}{4}$ (the maximum is attained for $L = 3, p = \frac{1}{2}$).*

We will prove bounds on $R_-(L,p)$ and $R_+(L,p)$, for which we will need bounds on binomial probabilities and central binomial coefficients. Specifically,

Lemma 3 (Binomial Tail Inequality)**.** *Given $p > \frac{1}{2}$, L and $k \le \lceil L/2 \rceil$,*

$$\binom{L}{k} p^k (1-p)^{L-k} \le \sum_{\ell=0}^{k} \binom{L}{\ell} p^\ell (1-p)^{L-\ell} \le \frac{p}{2p-1} \binom{L}{k} p^k (1-p)^{L-k}.$$

Proof. The lower bound is just the last term in the sum. The upper bound comes from the observation that each term in the sum is at least a factor $\mu/(1-\mu)$ bigger than the previous term (because $k \le \lceil L/2 \rceil$). $\square$

Lemma 4 (Near-Central binomial coefficient bound)**.** *For $L > 1$,*

$$\frac{3}{4} \cdot \frac{4^{L/2}}{\sqrt{\pi L}} \le \binom{L}{\lceil \frac{1}{2}(L-1) \rceil} \le 2 \cdot \frac{4^{L/2}}{\sqrt{\pi L}}.$$

Lemma 5 (Bounding $R_-(L,p)$ and $R_+(L,p)$)**.** *For $\frac{1}{2} < p < 1$,*

$$1 - \left(\frac{2}{(2p-1)}\right) \cdot \frac{(4p(1-p))^{L/2}}{\sqrt{\pi L}} \le R_-(L,p) \le R_+(L,p) \le 1 - \frac{3}{8p} \cdot \frac{(4p(1-p))^{L/2}}{\sqrt{\pi L}}.$$

The proofs are relegated to the appendix.

We are ready to prove the theorem. Let c be defined as in the statement of the theorem. We may assume $n > cK_*$ otherwise the theorem automatically holds. Further, if $L* = 1$, then there is just one group and any $K > K_*$ will also have just one group and be equivalent. Therefore, we may assume $L_* \ge 2$. Now suppose $K > cK_*$. Define $\mu_K = \mu(K)$ and $L_K = \lfloor n/K \rfloor$, $\mu_* = \mu(K_*)$ and $L_* = \lfloor n/K_* \rfloor$. Observe that $L_K \le L_*$. We show that $R(L_*, \mu_*) \ge R(L_K, \mu_K)$ which means that K cannot be better than K_* for a homogeneous partition of n, proving the theorem.

If $\mu_K \le \frac{1}{2}$ then $R_+(L_K, \mu_K) \le \frac{3}{4}$ (Lemma 2). We show that $R_-(L_*, \mu_*) > \frac{3}{4}$. Indeed, since $n > cK_*$, we have $n/K_* > c \ge \frac{-4}{\ln(1-4\epsilon^2)} \cdot \ln \frac{32}{9\epsilon^2 A}$, and so

$$L_* = \left\lfloor \frac{n}{K_*} \right\rfloor \ge \frac{n}{2K_*} > \frac{-2}{\ln(1-4\epsilon^2)} \cdot \ln \frac{32}{9\epsilon^2 A} \qquad \Longrightarrow \qquad \frac{(1-4\epsilon^2)^{L_*/2}}{\epsilon\sqrt{\pi L_*}} < \frac{9\epsilon A}{32\sqrt{\pi L_*}} < \frac{1}{4},$$

where in the last inequality we used $A \leq 1$ and $L_* \geq 2$. Now, using the bound for R_- from Lemma 5, we conclude that $R_-(L_*, \mu_*) > \frac{3}{4}$, which proves K cannot be optimal. Therefore, we may assume that $\mu_K > \frac{1}{2}$. Also, if $\epsilon = \frac{1}{2}$ then the representatives always vote for the correct decision and the theorem automatically holds, so we may assume $\epsilon < \frac{1}{2}$. Using Lemma 5,

$$\begin{aligned}
& R_n(\vec{H}_{K_*}, \mu) - R_n(\vec{H}_K, \mu) \geq R_-(L_*, \mu_*) - R_+(L_K, \mu_K)] \\
\geq \quad & \frac{3}{8}\frac{1}{\mu_K} \cdot \frac{(4\mu_K(1-\mu_K))^{L_K/2}}{\sqrt{\pi L_K}} - \left(\frac{2}{2\mu_* - 1}\right) \cdot \frac{(4\mu_*(1-\mu_*))^{L_*/2}}{\sqrt{\pi L_*}} \\
\overset{(L_* \geq L_K)}{\geq} \quad & \frac{3}{8}\frac{1}{\mu_K} \cdot \frac{(4\mu_K(1-\mu_K))^{L_K/2}}{\sqrt{\pi L_*}} - \left(\frac{2}{2\mu_* - 1}\right) \cdot \frac{(4\mu_*(1-\mu_*))^{L_*/2}}{\sqrt{\pi L_*}} \\
\geq \quad & (\text{positive}) \cdot \left[1 - C\left(\frac{(4\mu_*(1-\mu_*))^{L_*/L_K}}{4\mu_K(1-\mu_K)}\right)^{L_K/2}\right],
\end{aligned}$$

where $C = \frac{16}{3} \cdot \frac{\mu_K}{2\mu_* - 1} = \frac{8\mu_K}{3\epsilon}$. Note that $C \leq \frac{8}{3\epsilon}$ (because $\mu_K < 1$) and $C > 1$ (because $\mu_K > \frac{1}{2}$ and $\epsilon < \frac{1}{2}$). We prove the term in square parentheses is positive. Observe that

$$\frac{L_*}{L_K} = \frac{\lfloor n/K_* \rfloor}{\lfloor n/K \rfloor} \geq \frac{\lfloor n/K_* \rfloor}{n/K} = \frac{K}{K_*}\frac{\lfloor n/K_* \rfloor}{n/K_*} \geq \frac{K}{2K_*}.$$

We used $\lfloor x \rfloor / x \geq \frac{1}{2}$ when $x \geq 1$. Because $\mu_K > \frac{1}{2}$ and $1 - \mu_K \geq A/K^\alpha$, we have $\mu_K(1-\mu_K) \geq A/2K^\alpha$. Also recall that $\mu_* \geq \frac{1}{2} + \epsilon$, which means that $\mu_*(1-\mu_*) \leq (\frac{1}{2}+\epsilon)(\frac{1}{2}-\epsilon)$, and $K \leq n$. Therefore,

$$C\left(\frac{(4\mu_*(1-\mu_*))^{L_*/L_K}}{4\mu_K(1-\mu_K)}\right)^{L_K/2} \quad \leq \quad C\left(\frac{K^\alpha(1-4\epsilon^2)^{K/2K_*}}{2A}\right)^{L_K/2}$$

We show that the RHS is at most 1, or equivalently its logarithm is at most zero, concluding the proof. Taking the logarithm of the RHS, we get:

$$\begin{aligned}
& L_K\left(\frac{K}{4K_*}\ln(1-4\epsilon^2) + \frac{\alpha}{2}\ln K - \frac{1}{2}\ln 2A\right) + \ln C \\
\overset{(L_K \geq 1, C > 1)}{\leq} \quad & L_K\left(\frac{K}{4K_*}\ln(1-4\epsilon^2) + \frac{\alpha}{2}\ln K - \frac{1}{2}\ln 2A + \ln C\right) \\
\leq \quad & L_K\left(\frac{K}{4K_*}\ln(1-4\epsilon^2) + \frac{\alpha}{2}\left(\ln\frac{-4\alpha K_*/e}{\ln(1-4\epsilon^2)} - \frac{\ln(1-4\epsilon^2)}{4\alpha K_*}K\right) - \frac{1}{2}\ln 2A + \ln\frac{8}{3\epsilon}\right).
\end{aligned}$$

The last step follows by using the fact for any $z > 0$, $\ln x \leq \ln(z/e) + x/z$, which holds because for any $y = x/z > 0$, $\ln y - y + 1$ is maximized at $y = 1$. In the last step, we set $z = -4\alpha K_*/\ln(1-4\epsilon^2)$. Collecting terms, we have

$$\begin{aligned}
& L_K\left(\frac{K}{8K_*}\ln(1-4\epsilon^2) + \frac{\alpha}{2}\ln\frac{-4\alpha K_*/e}{\ln(1-4\epsilon^2)} - \frac{1}{2}\ln 2A + \ln\frac{8}{3\epsilon}\right) \\
= \quad & \frac{L_K \ln(1-4\epsilon^2)}{8}\left(\frac{K}{K_*} - c\right) \leq 0
\end{aligned}$$

where the last step follows because $K/K_* > c$ and $\ln(1-4\epsilon^2) < 0$. $\square$

As a corollary, we will show that Theorem 2 can be applied to any F with continuous density function, which means that in such cases the optimal number of groups with homogeneous size is $\Omega(n)$. To this end, we prove a lemma stating that any NOISY-MAX group competence function (Example 3) is polynomially bounded away from 1.

Lemma 6. *For any* NOISY-MAX *group competence function* $\mu_{\vec{p}}$ *and any* F *with a continuous density* f *on* $[0,1]$*, we have* $\mu_{\vec{p}}(K) \leq 1 - (\frac{1}{2\max(f)})/K$*. Moreover, the max representative democracy* $\mu_{\max}$ *is concave, which means that* $\mu_{\max}(K) + \mu_{\max}(K+2) \leq 2\mu_{\max}(K+1)$ *for all* $K \geq 1$*.*

Proof. It suffices to prove the inequality for $\mu_{\max}$ because expectation of any other order statistics is no more than it. Since f is continuous on the compact set $[0,1]$, it attains a maximum $B = \max(f)$. We note that since f is a density on $[0,1]$, $B \geq 1$. Let F be the CDF for f. We have

$$\mu_{\max}(K) \;=\; K\int_0^1 s\, f(S)F(s)^{K-1}ds = \int_0^1 s(F(s)^K)'ds = sF(s)^K\big|_0^1 - \int_0^1 F(s)^K ds$$

Since $f(s) \leq B$ is continuous, $F(s)$ is differentiable and $F'(s) \leq B$. Therefore, using Taylor's theorem at $s = 1$, $F(s) = 1 - (1-s)F'(t)$ for some $t \in [s,1]$, and since $F'(t) \leq B$, we have $F(s) \geq \max(0, 1 - B(1-s))$. Therefore,

$$\begin{aligned}\mu_{\max}(K) \;&\leq\; 1 - \int_0^1 ds\; \max(0, 1-B(1-s))^K = 1 - \int_{1-1/B}^1 ds\; (1 - B(1-s))^K \\ &=\; 1 - \frac{1}{B(K+1)} \overset{(K+1\leq 2K)}{\leq} 1 - (\tfrac{1}{2B})/K.\end{aligned}$$

Concavity of $\mu_{\max}$ holds because for any $K \geq 1$, we have

$$\begin{aligned}\mu_{\max}(K) + \mu_{\max}(K+2) - 2\mu_{\max}(K+1) &= -\int_0^1 \left[F(s)^K + F(s)^{K+2} - 2F(s)^{K+1}\right] ds \\ &= -\int_0^1 F(s)^K(1-F(s))^2 ds \leq 0\end{aligned}$$

□

Combining Theorem 2 and Lemma 6, and notice that $\mu_{\max}$ provides an upper bound on the competence of the representative chosen by *any* representative selection process μ, we have the following surprising corollary.

Corollary 1 (Constant group size)**.** *Suppose* F *has a continuous density function on* $[0,1]$*. For any* μ *such that* $\mu(K_*) > 0.5$ *for some* K_**. The optimal number of representatives for homogeneous groups is at least* $\frac{n}{cK_*}$*, where* c *is the constant defined in Theorem 2.*

Therefore, the representative democracy with fixed group size makes better choices than the representative democracy with fixed number of representatives. More discussions can be found in the next subsection.

One limitation of Theorem 2 is that it only holds for homogeneous group size. Next, we will extend the theorem to inhomogeneous groups, allowing groups to have different sizes, at the cost of further requiring that the group competence function $\mu(K)$ to be concave. Let $L^*(n)$ denote the optimal number of groups for n voters.

Theorem 3 (Optimal number of representatives for general group sizes)**.** *Suppose Cost(L) is a constant, there exists $K_* \in \mathbb{N}$ such that $\mu(K_*) \geq \frac{1}{2} + \epsilon$ for $0 < \epsilon < \frac{1}{2}$, μ is concave, and $\mu(K) \leq 1 - A/K^\alpha$ for constant $A < 1$ and $\alpha \geq 0$. Then, for any n, $L^*(n) \geq \left\lfloor \frac{n}{K_*} \right\rfloor / c$, where*

$$c = \frac{-4}{\ln(1-4\epsilon^2)} \left(\ln \frac{32}{9\epsilon^2 A} + \alpha \ln \frac{4\alpha K_*}{e} - \alpha \ln |\ln(1-4\epsilon^2)| \right).$$

Proof. We show how to modify the proof for Theorem 2. Let $R(\vec{K}_*, \mu)$ denote the probability that the majority vote among the L_* representatives is correct under partition $\vec{K}_*$. First, we may assume that $\lfloor n/K_* \rfloor > c$, because otherwise the RHS in the bound is less than 1 and the theorem automatically holds.

Our key technique is to allow group size to be non-integers by extending μ to a piece-wise linear function $\bar{\mu} : \mathbb{R}_{\geq 0} \to [0, 1]$. Then, we show that for any given number of groups L, $R(\vec{K}, \mu)$ is maximized when $\vec{K}$ partitions n into L equal-size groups, each with $\frac{n}{L}$ voters. This is guaranteed by concavity of μ and a result by Fey (2003).

More precisely, let $\bar{\mu} : \mathbb{R}_{\geq 0} \to [0, 1]$ be the piecewise linear function that interpolates μ at the integers. It follows that $\bar{\mu}$ is concave and bounded above by $1 - A/k^\alpha$ for all $k \geq 1$. For any $L \in \mathbb{N}$, we let $k = n/L$ and let $\bar{R}(L, p)$ denote the probability to obtain a majority of successes in L Bernoulli trials, where each trial succeeds with probability p. The following Lemma gives an upper bound on the probability for majority voting to succeed, when the average competence is at most 0.5. This answers an open question in (Owen et al., 1989, Lemma 5) and (Fey, 2003).

Lemma 7 (Poisson-Binomial majority)**.** *A Poisson-Binomial random variable $X = x_1 + \cdots + x_n$ is a sum of n independent Bernoulli trials x_i with respective (possibly different) success-probabilities p_i. Let q be the average probability, $nq = \sum_i p_i$ and suppose $q \leq \frac{1}{2}$. Then, the probability to get a strict majority of successful trials is bounded by a constant. Specifically,*

1. *If $n = 2K$ is even:*

$$\mathbb{P}[X > n/2] \leq 1 - B, \quad \text{where} \quad B = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i} q^i (1-q)^{n-i} \leq \frac{1}{2} \tag{1}$$

2. *If $n = 2K+1$ is odd and $n \geq 1/(1-2q)$ then (2) holds. Otherwise,*

$$\mathbb{P}[X > n/2] \leq \sqrt{1 - B}, \quad \text{where} \quad B = \sum_{i=0}^{n} \binom{2n}{i} q^i (1-q)^{2n-i} \leq \frac{1}{2}$$

The bounds are in terms of a standard Binomial with probability q equal to the average probability in the ensemble. The bound for even n is tight (set all probabilities to q). The boundary case which we alluded to earlier is for odd $n < 1/(1-2q)$. We conjecture that the bound for odd n can be improved (for $q = \frac{1}{2}$, we believe $\mathbb{P}[\text{majority}] \leq 1/\sqrt{e}$, which is better than $1/\sqrt{2}$). It is interesting that the probability of a majority can depend so much on whether the number of voters is even or odd, and this is true even asymptotically (for $q = \frac{1}{2}$, set K of the p_i to 0 and $K+1$ of the p_i to $\frac{2K+1}{2K+2}$, then $\mathbb{P}[\text{majority}] = (\frac{2K+1}{2K+2})^{K+1} \to 1/\sqrt{e}$). The proof is relegated to the appendix.

Continuing with the proof of Theorem 3, consider a partition $\vec{K}(n)$ with $|\vec{K}(n)| = L$ and define the average competence q of the representatives by $Lq = \sum_{i=1}^{L} \mu([K(n)]_i)$. If $q \leq \frac{1}{2}$, then by Lemma 7, $R(\vec{K}, \mu) < \frac{1}{\sqrt{2}}$. Using the bound for R_- in Lemma 5, with $\lfloor n/K_* \rfloor > \frac{-4}{\ln(1-4\epsilon^2)}(\frac{\alpha}{2} \ln n + \ln \frac{4\sqrt{2}}{3\epsilon\sqrt{A}})$, we have $R(\vec{K}_*, \mu) > \frac{1}{\sqrt{2}}$ and so $\vec{K}(n)$ cannot be optimal. Therefore, we may assume that $q > \frac{1}{2}$.

By concavity of $\bar{\mu}$, for any $\vec{K}(n)$ with $|\vec{K}(n)| = L$, we have $\bar{\mu}(k) = \bar{\mu}(n/L) \geq \sum_{i=1}^{L} \frac{1}{L} \mu([K(n)]_i) = q > \frac{1}{2}$. By (Fey, 2003), since $q > \frac{1}{2}$, the probability of a majority is maximized when all the probabilities are the same as the average value q. And by monotonicity, $\bar{R}(L, \cdot)$ can only increase in going from q to

$\bar{\mu}(k)$. Therefore, $\bar{R}(L, \bar{\mu}(k)) \geq R(\vec{K}, \mu)$. What we have established is that if you fix the number of groups, then we may assume that the probability to get a majority cannot be better than the outcome where all group sizes are homogeneous (possibly non-integer), equal to n/L. This gives us an upper bound on the quality of the number of groups L.

The rest of the proof is to show that if L is smaller than the claim of the theorem, then this upper bound on the quality is less than the choosing L_* equal to the bound in the Theorem, therefore proving that L is not optimal. This part of the proof proceeds in the same way as the proof of Theorem 2, by replacing K by k and replacing $\mu(K)$ by $\bar{\mu}(k)$. Lemma 1–5 still hold because they do not depend on k. We note that the key steps involving K in the proof of Theorem 2 are $\mu(K) \geq 0.5$ and $1 - \mu(K) \geq A/K^{\alpha}$, which hold for k because we focus on L such that $\mu(n/L) > 0.5$ and $\bar{\mu}(k) \leq 1 - A/k^{\alpha}$. □

Next, we further extend Theorem 3 by showing that the optimal partitioning function $\vec{K}$ must be nearly homogeneous. This property comes at the cost of requiring that μ is log-concave and $1 - \mu$ is log-convex. Log-concavity is a weaker assumption than concavity. Let us consider an example.

Example 4 (Log-concavity and log-convexity of $\mu_{\max}$)**.** For uniformly distributed voters $F = \text{Uniform}[0, 1]$ as in Example 1, the MAX group competence function $\mu_{\max}(K) = K/(K+1)$ is log-concave (it is also concave). Log-convexity of $1 - \mu_{\max}$ holds because $1 - \mu_{\max}(K) = 1/(K+1)$ and $(K+2)^2 > (K+1)(K+3)$, therefore $(1 - \mu_{\max}(K))(1 - \mu_{\max}(K+2)) > (1 - \mu_{\max}(K+1))^2$. ■

Theorem 4 (Near-Homogeneity of group sizes)**.** *Suppose Cost(L) is a constant, the group competence function $\mu(K)$ is log-concave and non-decreasing, and further that $1 - \mu(K)$ is log-convex. Given L groups, there is an optimal partition $(K_1, \ldots, K_L)$ of n into the L groups with no two groups differing in size by more than 1. That is,* $\max_i K_i - \min_i K_i \leq 1$.

Proof. Suppose we have a partition of n voters into L groups with sizes $K_1 \leq K_2 \leq \cdots \leq K_L$, and suppose that $K_L - K_1 \geq 2$. Let $\mu_1 = \mu(K_1)$ and $\mu_L = \mu(K_L)$. Let the groups $K_2, \ldots, K_{L-1}$ yield an arbitrary ensemble A of $L - 2$ representatives which is fixed in this proof. Define the functions

$$\begin{aligned} f(k) &= \mathbb{P}[k \text{ successes in } A] \\ Q(k) &= \mathbb{P}[\text{at least } k \text{ successes in } A] = f(k) + f(k+1) + \cdots + f(L-2). \end{aligned}$$

Let $M = \lceil (L+1)/2 \rceil$ be the majoirty threshold for the L representatives and define P as the probability that a majority of successes is obtained from all the representatives. Conditioning on the votes of representatives 1 and L, we have,

$$\begin{aligned} P &= \mu_1\mu_L Q(M-2) + (\mu_1(1-\mu_L) + \mu_L(1-\mu_1))Q(M-1) + (1-\mu_1)(1-\mu_L)Q(M) \\ &= \mu_1\mu_L(Q(M-2) + Q(M) - 2Q(M-1)) + (\mu_1+\mu_L)(Q(M-1) - Q(M)) + Q(M) \\ &= \mu_1\mu_L(f(M-2) - f(M-1)) + (\mu_1+\mu_L)f(M-1) + Q(M) \\ &= \mu_1\mu_L f(M-2) + (\mu_1 + \mu_L - \mu_1\mu_L)f(M-1) + Q(M) \\ &= \mu_1\mu_L f(M-2) + (1 - (1-\mu_1)(1-\mu_L))f(M-1) + Q(M) \end{aligned}$$

We now consider the partition of the n voters obtained by keeping the ensemble A fixed, increasing K_1 to $K_1 + 1$ and decreasing K_L to $K_L - 1$. Let P' be the probability of a majority for the new representatives. The only difference with P is that $\mu_1 \to \mu(K_1+1)$ and $\mu_L \to \mu(K_L-1)$. By log-concavity of μ, $\mu(K_1+1)\mu(K_L-1) \geq \mu_1\mu_L$. By log-convexity of $1-\mu$, $(1-\mu(K_1+1))(1-\mu(K_L-1)) \leq (1-\mu_1)(1-\mu_L)$. Therefore $P' \geq P$. As long as the difference in largest and smallest sizes is at least 2, we can continue merging without decreasing P, proving the theorem. □

Theorem 4 also applies to non-constant cost functions because the number of groups is fixed.

3.2.1 Numerical Example of Optimal Group Size

As an application of our results, we consider the uniform voters with MAX representative selection process (Example 2) applied to US House. Below, we show how the optimal homogeneous group size K^* and the minimum group size required to achieve consistency K_* depend on the upper bound on a voter's probability of being correct, b. We fix a, the lower bound, to 0.45, so $F = \text{Uniform}[0.45, b]$. When b is small, as voters

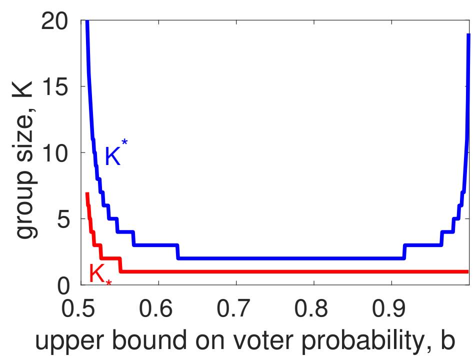

Figure 2: Optimal group size.

get wiser (b goes up), K_* and K^* are decreasing. At some point, even the direct-democracy ($K_* = 1$) is consistent. For very large b, the optimal group size starts to increase due to the $\mu(1-\mu)$ term. In general, the optimal group size is less than about 5, the size of the typical household. The optimal representative democracy is obtained when each household elects a head to vote on its behalf.

The optimal group size is useful to know, but for practical purposes, there may not be a significant difference between different values of K for a large n. Let us take the US House as an example, which has 435 representatives. Suppose the voting population is about $n = 235$ million, and that voter competence is uniformly distributed from 0.45 to 0.52 (the average competence is slightly less than 0.5).

	$K = 1$	$K_* = 3$	$K^* = 9$	$5 \times$ House	$2 \times$ House	House
Success rate	0%	100%	100%	97%	88%	80%

In this simple setting, direct democracy ($K = 1$) would be wrong, and the current size of House is far from optimal, 20% less accurate than what is achievable. Doubling congress gets you to 88% and multiplying by 5 pretty much gets you to optimal. A House that is 20 times larger (each member representing 35K citizen) would be essentially indistinguishable from optimal. This suggests that a much larger House is needed for noisy issues like the one in this example. Also note that if group sizes are about 5 (the size of a household) then we have near-perfection results.

3.3 Optimal Group Size for Single Issue: Polynomial Cost of Voting

In this subsection, we focus on the setting where the cost of voting and the benefit of correct decision are both polynomial.

Theorem 5 (Optimal homogeneous group size, polynomial cost and polynomial benefit)**.** *Suppose* $\frac{Cost(L)}{Ben(n)} = \Theta(\frac{L^{q_1}}{n^{q_2}})$ *for constants* $q_1 > 0$ *and* $q_2 > 0$*, there exists* $K_* \in \mathbb{N}$ *such that* $\mu(K_*) > \frac{1}{2}$*,* μ *is non-decreasing, and for all* $K \in \mathbb{N}$*,* $\mu(K) \leq 1 - A/K^\alpha$ *for constants* $A > 0$ *and* $\alpha \geq 0$*. Then, the optimal group size* $K^*_{hom}(n) = \Omega(n/\log n)$*.*

Moreover, we have:

(i) If $\lim_{K\to\infty} \mu(K) < 1$*, then* $K^*_{hom}(n) = \Theta(n/\log n)$*.*

(ii) If there exists $B, \beta > 0$ *such that* $\mu(K) \geq 1 - B/K^\beta$*, then* $K^*_{hom}(n) = \Theta(n)$*.*

Proof. We first prove that the optimal number of groups L^*_{Hom} is $O(\log n)$. Let K' be the smallest number such that $\mu(K') > 0.5$. For any $K > K'$ and $L = \lfloor n/K \rfloor$, it follows from Lemma 5 that

$$R_n(\vec{H}_K, \mu) - \frac{\text{Cost}(L)}{\text{Ben}(n)} = 1 - Q(L, n),$$

where $Q(L,n) = \Theta(\frac{(4\mu(K)(1-\mu(K)))^{L/2}}{\sqrt{L}} + \frac{L^{q_1}}{n^{q_2}})$. This is because $\mu(K) > \mu(K') > 0.5$.

Therefore, $L^*_{\text{Hom}} = \arg\min_L Q(L,n)$. For any $L \geq n/K'$, we let $p(L) = 4\mu(\lceil n/L \rceil)(1 - \mu(\lceil n/L \rceil))$, which means that $p(L)$ is non-decreasing in L. Let $c_1, c_2 > 0$ be such that for all $L \geq n/K'$,

$$c_1(\frac{p(L)^{L/2}}{\sqrt{L}} + \frac{L^{q_1}}{n^{q_2}}) \leq Q(L,n) \leq c_2(\frac{p(L)^{L/2}}{\sqrt{L}} + \frac{L^{q_1}}{n^{q_2}})$$

Let $p' = 4\mu(K')(1 - \mu(K'))$. We have $p' < 1$. Let $c = -2q_2/\log p'$, which means that $(p')^{\frac{c}{2}\log n} = \frac{1}{n^{q_2}}$. When $n \geq 2^{1/c}$, we have

$$Q(c\log n, n) \leq c_2(\frac{(p')^{\frac{c}{2}\log n}}{\sqrt{c\log n}} + \frac{(c\log n)^{q_1}}{n^{q_2}}) \leq c_2(\frac{1}{n^{q_2}\sqrt{c\log n}} + \frac{(c\log n)^{q_1}}{n^{q_2}}) \leq 2c_2\frac{(c\log n)^{q_1}}{n^{q_2}}$$

For any $L \geq c\sqrt[q_1]{\frac{2c_2}{c_1}}\log n$, we have $Q(L,n) > c_1 L^{q_1}/n^{q_2} = 2c_2\frac{(c\log n)^{q_1}}{n^{q_2}} \geq Q(c\log n, n)$. It follows that $\arg\min_L Q(L,n) \leq \frac{c_2 c}{c_1}\log n = O(\log n)$.

To prove (i), let $b = \lim_{K\to\infty}\mu(K) < 1$. For any $L < -\frac{q_2}{\log(4b(1-b))}\log n$, we have

$$Q(L,n) > c_1\frac{p(L)^{L/2}}{\sqrt{L}} > c_1\frac{(4b(1-b))^{L/2}}{\sqrt{L}} > c_1\frac{(4b(1-b))^{-\frac{q_2}{2\log(4b(1-b))}\log n}}{\sqrt{L}} = \Theta(\frac{1}{n^{q_2/2}\sqrt{\log n}})$$

It follows that when n is large enough, for any $L < -\frac{q_2}{\log(4b(1-b))}\log n$, we have $Q(L,n) > 2c_2\frac{(c\log n)^{q_1}}{n^{q_2}} = Q(c\log n, n)$. This means that $K^*_{\text{hom}}(n) = \Omega(n/\log n)$.

To prove (ii), let $\bar{L} = 2q_2/\beta$, it is not hard to check that $Q(\bar{L}, n) = O((\frac{4B}{(n/\bar{L})^\beta})^{\bar{L}/2}/\sqrt{\bar{L}} + \bar{L}^{q_1}/n^{q_2}) = O(1/n^{q_2})$. Therefore, for any L larger than some constant, we have $Q(L,n) > L^{q_1}/n^{q_2} > Q(\bar{L}, n)$, which means that $K^*_{\text{hom}}(n) = \Theta(n)$. □

4 Optimal Representative Democracy for Multiple Issues

We now extend our setting to $d \geq 2$ issues, $I_1, \ldots, I_d$ with fixed cost of voting. Each voter's type is now represented by a 2^d-dimensional *competence vector* $\vec{c}$, which is a distribution over $\{0,1\}^d$. A voter's competence vector represents her probability to cast a combination of votes over the d issues. Let Δ_{2^d} denote the 2^d-dimensional simplex, which is the set of all possible competence vectors. We assume that each voter's competence vector is generated i.i.d. from a distribution F over Δ_{2^d}.

Example 5. For two issues I_1 and I_2, let F be the uniform distribution over two competence vectors $\vec{c}_1, \vec{c}_2$, as illustrated below, where $(0,1)$ means that the voter is wrong on issue 1 but correct on issue 2.

F	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$\frac{1}{2}@\vec{c}_1$	2/9	4/9	1/9	2/9
$\frac{1}{2}@\vec{c}_2$	1/9	1/9	4/9	3/9

The two issues are independent in $\vec{c}_1$: issue I_1 (respectively, I_2) takes 1 with probability $\frac{1}{3}$ (respectively, $\frac{2}{3}$), independent of the other issue. The two issues are correlated in $\vec{c}_2$. ■

Suppose K voters $v_{\ell,1}, \ldots, v_{\ell,K}$ choose a representative r_ℓ with competence vector $\vec{c}_\ell$. As with one issue, the representative can be summarized by $\mathbb{E}(\vec{c}_\ell) \in \Delta_{2^d}$. Thus, we can succinctly describe a d-issue representative selection process as a group competence function.

Definition 7. *A representative selection process over d issues is a* group competence function $\rho : \mathbb{N} \mapsto \Delta_{2^d}$.[4] *The multi-issue competence function ρ induces d single-issue* marginal group competence functions $\mu_1, \mu_2, \ldots, \mu_d$*, where μ_i is the marginal distribution for issue I_i. For any $K \in \mathbb{N}$, let $\rho(K) = (p_{\alpha_1,\alpha_2,\ldots,\alpha_d} : \forall i, \alpha_i \in \{0,1\})$. Then, $\mu_i(K) = \sum_{\alpha:\alpha_i=1} p_{\alpha_1,\alpha_2,\ldots,\alpha_d}$.*

Example 6. Continuing Example 5, suppose for a group of two voters independently sampled from F, the representative selection process is to choose the voter with maximum expected number of correct votes. Then, the representative will be a $\vec{c}_2$ voter with probability $\frac{3}{4}$ and a $\vec{c}_1$ voter with probability $\frac{1}{4}$. This group competence function, called MAX-SUM and denoted by ρ_{ms}, will output

$$\rho_{\text{ms}}(2) = \frac{1}{4}\vec{c}_1 + \frac{3}{4}\vec{c}_2 = \left(\frac{5}{36}, \frac{7}{36}, \frac{13}{36}, \frac{11}{36}\right)$$

And the marginal group competence functions are $\mu_1(2) = \frac{13}{36} + \frac{11}{36} = \frac{2}{3}$, $\mu_2(2) = \frac{7}{36} + \frac{11}{36} = \frac{1}{2}$. ■

As with a single issue, when there are d issues, given a partition function $\vec{K}$ and a group competence function ρ, we let $S^d_{n,\vec{K},\rho}$ denote the 2^d-dimensional random variable that is the average of $L(n) = |\vec{K}(n)|$ independent random variables $(\rho([\vec{K}(n)]_1), \ldots, \rho([\vec{K}(n)]_{L(n)}))$, where each $\rho([\vec{K}(n)]_i)$ represent the random vote on d issues by the representative of group i. We let $R^d_n(\vec{K}, \rho) = \mathbb{P}[S^d_{n,\vec{K},\rho} > \frac{1}{2} \cdot \vec{1}]$ denote the probability that majority voting is correct on all d issues.

4.1 Consistent Representative Democracy for Multiple Issues

Our next theorem extends the Condorcet Jury Theorem to representative democracy with multiple issues. It states that when the marginal group competence functions μ_i's are monotonic, the representative democracy is consistent if and only for each issue, there exists a group size for which marginal group competence of that issue is strictly larger than 0.5.

Theorem 6. *For $d \geq 2$, let ρ be a d-issue group competence function with monotonic marginals μ_i.*[5] *Then, ρ is consistent w.r.t. every issue if and only if every marginal is consistent, i.e. for all $i \in \{1, \ldots, d\}$, there exists K_i such that $\mu_i(K_i) > \frac{1}{2}$.*

The proof uses the union bound and is similar to the proof of Theorem 1. The full proof can be found in the appendix.

4.2 Optimal Group Size for Multiple Issues

We now prove the analog of Theorem 2 for multiple issues. Namely an upper bound on the optimal homogeneous group size for any consistent multi-issue representative democracy. We recall that $K^*_{\text{hom}}(n)$ is the optimal partition function for n voters.

[4]We use ρ to distinguish from single-issue group competence function μ.

[5]Technically, all we need is that $\mu_i(K)$ can be lower-bounded by an increasing function of K, and that this lower bound exceeds 0.5 for some K_i.

Theorem 7. *Let ρ be a d-issue group competence function for which there exists $K_* \in \mathbb{N}$ such that for all $i \leq d$, $\mu_i(K_*) \geq \frac{1}{2} + \epsilon$ for constant $0 < \epsilon < \frac{1}{2}$, and $\mu_i(K) \leq 1 - A/K^\alpha$ for constant $A < 1$ and $\alpha \geq 0$. Then, $K^*_{hom}(n) \leq (c - \frac{8}{\ln(1-4\epsilon^2)} \ln d)K_*$, where*

$$c = \frac{-4}{\ln(1-4\epsilon^2)} \left(\ln \frac{32}{9\epsilon^2 A} + \alpha \ln \frac{4\alpha K_*}{e} - \alpha \ln |\ln(1-4\epsilon^2)| \right).$$

Proof. Given any distribution γ over $\{0,1\}^d$ and any $L \geq 1$, we let $Q_-(L,\gamma)$ (respectively, $Q_+(L,\gamma)$) denote the probability that for every one of the d issues, the majority of voters vote for 1, where there are L independent voters, the first $L-1$ vote according to γ, and the last voter always vote for 0 (respectively 1) for all issues. We first extend Lemma 5 to multiple issues.

Lemma 8. *For any multi-issue representative democracy with group competence function ρ and given n, K, let the marginals be μ_i and let $q = \min_i \mu_i(K)$ be the minimum marginal. Then,*

$$1 - d\left(\frac{2}{(2q-1)}\right) \cdot \frac{(4q(1-q))^{L/2}}{\sqrt{\pi L}} \leq Q_-(L,\rho(K)) \leq Q_+(L,\rho(K)) \leq 1 - \frac{3}{8q} \cdot \frac{(4q(1-q))^{L/2}}{\sqrt{\pi L}}$$

We can now mimic the analysis in the proof of Theorem 2 to prove an upper bound on the optimal group size in a multi-issue representative democracy. We give only the main steps, omitting some details. Let $L_* = \lfloor n/K_* \rfloor$, $q_* = \min_i \mu_i(K_*)$, $L_K = \lfloor n/K \rfloor$, and $q_K = \min_i \mu_i(K)$. Using Lemma 8, $q_* \geq \frac{1}{2} + \epsilon$ and $q_k \leq 1 - A/K^\alpha$ (which implies $q(1-q) \geq A/2K^\alpha$), we get

$$R^d_n(\vec{H}_{K_*},\rho) - R^d_n(\vec{H}_K,\rho) \geq Q_-(L_*,\rho(K_*)) - Q_+(L_K,\rho(K)) \geq (\text{pos}) \left[1 - dC \left(\frac{K^\alpha(1-4\epsilon^2)^{K/2K_*}}{2A} \right)^{L_K/2} \right]$$

Here $C = 8\mu_K/3\epsilon \leq 8/3\epsilon$. It now remains to prove that the expression in square parentheses is positive, for which it suffices to show that the logarithm of the second term is at most 0. Taking the logarithm of the second term,

$$\begin{aligned}
& L_K \left(\frac{K}{4K_*} \ln(1-4\epsilon^2) + \frac{\alpha}{2} \ln K - \frac{1}{2} \ln 2A \right) + \ln(dC) \\
\overset{(L_K \geq 1, dC > 1)}{\leq} \quad & L_K \left(\frac{K}{4K_*} \ln(1-4\epsilon^2) + \frac{\alpha}{2} \ln K - \frac{1}{2} \ln 2A + \ln(dC) \right) \\
\leq \quad & L_K \left(\frac{K}{4K_*} \ln(1-4\epsilon^2) + \frac{\alpha}{2} \left(\ln \frac{-4\alpha K_*/e}{\ln(1-4\epsilon^2)} - \frac{\ln(1-4\epsilon^2)}{4\alpha K_*} K \right) - \frac{1}{2} \ln 2A + \ln \frac{8}{3\epsilon} + \ln d \right) \\
= \quad & \frac{L_K \ln(1-4\epsilon^2)}{8} \left(\frac{K}{K_*} - c + \frac{8 \ln d}{\ln(1-4\epsilon^2)} \right) \leq 0
\end{aligned}$$

The last step follows from the choice of c. $\square$

4.3 Realizing a Consistent Representative Democracy

In Theorem 7, the group competency function ρ is given as part of the input. In this section we focus on representative selection processes that would lead to a consistent representative democracy. We first start with an example on F where no representative democracy is consistent.

Example 7. Let F be the uniform distribution over two competence vectors: $\vec{c}_1$ from Example 5 and $\vec{c}_2 = (\frac{2}{9}, \frac{1}{9}, \frac{4}{9}, \frac{2}{9})$ for $(0,0), (0,1), (1,0), (1,1)$, respectively. The two issues are independent in both $\vec{c}_1$ ($\frac{1}{3}$ for I_1

and $\frac{2}{3}$ for I_2) and $\vec{c}_2$ ($\frac{2}{3}$ for I_1 and $\frac{1}{3}$ for I_2). We note that for each type of voter, the marginal competences of two issues always sum up to 1.

Let ρ denote the group competence function of any representative selection process that chooses a group member as the representative. Then, the marginal competences of the chosen representative must sum up to 1. Therefore, for any K, we must have $\mu_1(K) + \mu_2(K) = 1$. Therefore, for any partition function $\vec{K}$ and any n, either the average marginal competence for I_1 is no more than 0.5, or the average marginal competence for I_2 is no more than 0.5, which means that the majority vote of I_1 or I_2 will not be correct with probability 1 as $n \to \infty$ (Lemma 7). ■

We see in Example 7 that when the issues are correlated, sometimes nothing can be done to get multi-issue consistency. A natural selection process is for a group to choose the voter having the maximum "total competence" over all the issues. This is the process used in Example 5.

Definition 8 (Max-Sum Process)**.** *Given a group of K voters whose votes on d issues are represented by random variables $v_1, \ldots, v_K$, the* MAX-SUM *representative selection process chooses a voter i with maximum $\mathbb{E}[v_i] \cdot \vec{1}$. Let ρ_{ms} denote its group competence function.*

ρ_{ms} naturally extends the MAX group competence function $\mu_{\max}$ defined in Example 2. We will prove that ρ_{ms} works well when the d issues are independent, formally defined below.

Definition 9 (Independent Issues)**.** *In this setting, each voter's competence vector is uniquely characterized by a vector of d numbers $(p_1, \ldots, p_d)$, where p_i is the probability that the voter's vote for issue i is correct, independent of the other issues. We further assume that F is a product distribution $f_1(p_1)f_2(p_2)\cdots f_d(p_d)$, and each $f_j(\cdot)$ has support on $(\frac{1}{2}, 1]$.*[6]

Theorem 8. *The max sum representative democracy is consistent for independent issues.*

Proof. Based on the distributions $f_i(\cdot)$, let M_i be the maximum possible value attainable by p_i ($M_i \geq \frac{1}{2} + \epsilon$) and let $M = \sum_i M_i$ be the maximum possible value for a voter's sum of probabilities $\sum_{i=1}^d p_i$. The basic idea in the proof is to show that for sufficiently large K, the representative's $\sum_i p_i$ approaches M, which means that each probability must approach M_i (all above $\frac{1}{2} + \epsilon$), and once that happens, the majority vote among many representatives will get all issues correct.

For a voter, let $s = \sum_i p_i$. First, let us prove that the probability for s to be close to its maximum possible value M is large. Recall that $M_i \geq \frac{1}{2} + \epsilon$. Let $P_i = \mathbb{P}[p_i \geq M_i - \frac{1}{2d}\epsilon] > 0$ (because f_i has support on $(\frac{1}{2}, 1]$). Let $P = P_1 \times \cdots \times P_d$. Then, $\{p_1 \geq M_1 - \frac{1}{2d}\epsilon$ and $p_2 \geq M_2 - \frac{1}{2d}\epsilon$ and $\ldots$ and $p_d \geq M_d - \frac{1}{2d}\epsilon\}$ implies $\{s \geq M - \frac{1}{2}\epsilon\}$ and we have $\mathbb{P}[s \geq M - \frac{1}{2}\epsilon] \geq P_1 \times \cdots \times P_d = P > 0$.

Let us now consider K independent voters and the sums of probabilities $s_1, \ldots, s_K$. The representative r is picked as the voter with maximum sum, and we have

$$\mathbb{P}[\max_\ell s_\ell \geq M - \tfrac{1}{2}\epsilon] \;=\; 1 - \prod_\ell \mathbb{P}[s_\ell < M - \tfrac{1}{2}\epsilon] = 1 - \mathbb{P}[s < M - \tfrac{1}{2}\epsilon]^K \geq 1 - (1-P)^K.$$

Observe that $s \geq M - \frac{1}{2}\epsilon$ implies $p_i \geq M_i - \frac{1}{2}\epsilon$ for every issue I_i. Therefore, for the representative r,

$$\mathbb{P}[p_i \geq M_i - \tfrac{1}{2}\epsilon] \geq \mathbb{P}[s(r) \geq M - \tfrac{1}{2}\epsilon] \geq 1 - (1-P)^K.$$

Since $M_i \geq \frac{1}{2} + \epsilon$, $p_i \geq M_i - \frac{1}{2}\epsilon$ implies $p_i \geq \frac{1}{2} + \frac{1}{2}\epsilon$, and so $\mathbb{P}[p_i \geq \frac{1}{2} + \frac{1}{2}\epsilon] \geq 1 - (1-P)^K$.

The marginal single-issue competence function for issue I_i is just $\mathbb{E}[p_i]$,

$$\mu_i(K) = \mathbb{E}[p_i] \geq \mathbb{P}[p_i \geq \tfrac{1}{2} + \tfrac{1}{2}\epsilon] \times (\tfrac{1}{2} + \tfrac{1}{2}\epsilon) \geq (1 - (1-P)^K) \times (\tfrac{1}{2} + \tfrac{1}{2}\epsilon)$$

We thus have a lower bound for each marginal density which is monotonically increasing in K. Further, by setting $K > \log(\frac{\epsilon}{1+\epsilon}) / \log(1-P)$, we find that $\mu_i(K) > \frac{1}{2}$. Therefore the marginal single-issue group competence functions are monotonic and consistent. By Theorem 8, the multiple-issue group competence function is consistent. □

[6] $M_i = \sup_p\{p : f_i(p) > 0\} \geq \frac{1}{2} + \epsilon$ and f_i has nonzero support at M_i, that is $\mathbb{P}[p_j \geq M_i - \delta] > 0$ for all $\delta > 0$.

5 Summary and Future Work

We set the mathematical foundation for studying the quality-quantity tradeoff in a representative democracy by introducing a mathematical framework for studying representative democracy, and show that under general and natural conditions, the optimal group size is constant when the cost of voting is a constant, and is $\Omega(n/\log n)$ when the cost and benefit are both polynomial.

There are many open questions and future directions under our framework. Can we extend our results to inhomogeneous representative selection processes, e.g. different states use different processes to choose representatives? Does diversity in population help make better decisions? More generally, it would be interesting and important to consider similar extensions as done for the Condorcet Jury Theorem, for example to inhomogeneous agents and strategic agents.

References

Emmanuelle Auriol and Robert J. Gary-Bobo. 2012. On the optimal number of representatives. *Public Choice* 153, 3–4 (2012), 419–445.

Hossein Azari Soufiani, David C. Parkes, and Lirong Xia. 2014. Statistical Decision Theory Approaches to Social Choice. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. Montreal, Quebec, Canada.

Bruce Bartlett. 2014. Enlarging the House of Representatives. `https://economix.blogs.nytimes.com/2014/01/07/enlarging-the-house-of-representatives/`. (2014).

Ruth Ben-Yashar and Jacob Paroush. 2000. A nonasymptotic Condorcet jury theorem. *Social Choice and Welfare* 17, 2 (2000), 189–199.

Ruth Ben-Yashar and Jacob Paroush. 2003. Investment in Human Capital in Team Members Who Are Involved in Collective Decision Making. *Journal of Public Economic Theory* 5, 3 (2003), 527—539.

Ruth Ben-Yashar and Mor Zahavi. 2011. The Condorcet jury theorem and extension of the franchise with rationally ignorant voters. *Public Choice* 148, 3 (2011), 435–443.

Daniel Berend and Jacob Paroush. 1998. When is Condorcet's Jury Theorem valid? *Social Choice and Welfare* 15, 4 (1998), 481–488.

Daniel Berend and Luba Sapir. 2005. Monotonicity in Condorcet Jury Theorem. *Social Choice and Welfare* 24 (2005), 83–92.

Daniel Berend and Luba Sapir. 2007. Monotonicity in Condorcet's Jury Theorem with dependent voters. *Social Choice and Welfare* 28, 3 (2007), 507–528.

Timothy Besley and Stephen Coate. 1997. An Economic Model of Representative Democracy. *The Quarterly Journal of Economics* 122, 1 (1997), 85–114.

Ioannis Caragiannis, Ariel D. Procaccia, and Nisarg Shah. 2016. When Do Noisy Votes Reveal the Truth? *ACM Transactions on Economics and Computation* 4, 3 (2016), Article No. 15.

Gal Cohensius, Shie Mannor, Reshef Meir, Eli Meirom, and Ariel Orda. 2017. Proxy Voting for Better Outcomes. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. 858–866.

Marquis de Condorcet. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: L'Imprimerie Royale.

Vincent Conitzer, Rupert Freeman, and Nisarg Shah. 2017. Fair Public Decision Making. In *Proceedings of the 18th ACM Conference on Economics and Computation*.

Vincent Conitzer and Tuomas Sandholm. 2005. Common Voting Rules as Maximum Likelihood Estimators. In *Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (UAI)*. Edinburgh, UK, 145–152.

Edith Elkind and Nisarg Shah. 2014. Electing the Most Probable Without Eliminating the Irrational: Voting Over Intransitive Domains. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*. 182–191.

Scott L. Feld and Bernard Grofman. 1984. The accuracy of group majority decisions in groups with added members. *Public Choice* 42, 3 (1984), 273–285.

Mark Fey. 2003. A note on the Condorcet Jury Theorem with supermajority voting rules. *Social Choice and Welfare* 20, 1 (2003), 27–32.

Brian Flynn. 2012. What's wrong with Congress? It's not big enough. `https://www.cnn.com/2012/03/09/opinion/flynn-expand-congress/index.html`. (2012).

James E. Gentle. 2009. *Computational Statistics*. Springer.

Mark Gradstein and Shmuel Nitzan. 1987. Organizational decision-making quality and the severity of the free-riding problem. *Economics Letters* 23, 4 (1987), 335–339.

Bernard Grofman, Guillermo Owen, and Scott L. Feld. 1983. Thirteen theorems in search of the truth. *Theory and Decision* 15, 3 (1983), 261–278.

Wassily Hoeffding. 1956. On the Distribution of the Number of Successes in Independent Trials. *The Annals of Mathematical Statistics* 27, 3 (1956), 713–721.

Keith Humphreys. 2016. Why we might want to grow the House of Representatives by 250 more seats. `https://www.washingtonpost.com/news/wonk/wp/2016/10/03/why-we-might-want-to-grow-the-house-of-representatives-by-250-more-seats/?utm_term=.2948bef5e749`. (2016).

Anson Kahng, Simon Mackenzie, and Ariel D. Procaccia. 2018. Liquid Democracy: An Algorithmic Perspective.. In *Proc. 32nd AAAI Conference on Artificial Intelligence*.

Satoshi Kanazawa. 1998. A brief note on a further refinement of the Condorcet Jury Theorem for heterogeneous groups. *Mathematical Social Sciences* 35, 1 (1998), 69–73.

Drora Karotkin and Jacob Paroush. 1995. Incentive schemes for investment in human capital by members of a team of decision makers. *Labour Economics* 2, 1 (1995), 41—51.

Drora Karotkin and Jacob Paroush. 2003. Optimum committee size: Quality-versus-quantity dilemma. *Social Choice and Welfare* 20 (2003), 429–441.

Yukio Koriyama, Antonin Macé, Rafael Treibich, and Jean-François Laslier. 2013. Optimal Apportionment. *Journal of Political Economy* 121, 3 (2013), 584–608.

Jérôme Lang and Lirong Xia. 2016. Voting in Combinatorial Domains. In *Handbook of Computational Social Choice*, Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel Procaccia (Eds.). Cambridge University Press, Chapter 9.

Nicholas R. Miller. 1986. Information, Electorates, and Democracy: Some Extensions and Interpretations of the Condorcet Jury Theorem. In *Information Pooling and Group Decision Making*, Grofman B. and Owen G. (Eds.). JAI Press, 173—192.

Kaushik Mukhopadhaya. 2003. Jury Size and the Free Rider Problem. *Journal of Law, Economics, and Organization* 19, 1 (2003), 24–44.

Shmuel Nitzan and Jacob Paroush. 1980. Investment in Human Capital and Social Self Protection under Uncertainty. *International Economic Review* 21, 3 (1980), 547–557.

Shmuel Nitzan and Jacob Paroush. 1984. The significance of independent decisions in uncertain dichotomous choice situations. *Theory and Decision* 17, 1 (1984), 47–60.

Shmuel Nitzan and Jacob Paroush. 2017. Collective Decision Making and Jury Theorems. In *The Oxford Handbook of Law and Economics: Volume 1: Methodology and Concepts*, Francesco Parisi (Ed.). Oxford University Press.

Guillermo Owen, Bernard Grofman, and Scott L. Feld. 1989. Proving a distribution-free generalization of the Condorcet Jury Theorem. *Mathematical Social Sciences* 17, 1 (1989), 1–16.

Jacob Paroush. 1998. Stay away from fair coins: A Condorcet jury theorem. *Social Choice and Welfare* 15, 1 (1998), 15–20.

Jacob Paroush and D. Karotkin. 1989. Robustness of Optimal Majority Rules Over Teams with Changing Size. *Social Choice and Welfare* 6, 2 (1989), 127–138.

Marcus Pivato. 2013. Voting rules as statistical estimators. *Social Choice and Welfare* 40, 2 (2013), 581–630.

Ariel D. Procaccia, Sashank J. Reddi, and Nisarg Shah. 2012. A Maximum Likelihood Approach For Selecting Sets of Alternatives.. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*.

Luba Sapir. 2005. Generalized means of jurors' competencies and marginal changes of jury's size. *Mathematical Social Sciences* 50, 1 (2005), 83–101.

Piotr Skowron. 2015. What Do We Elect Committees For? A Voting Committee Model for Multi-Winner Rules. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*. 1141–1147.

Peter Stone and Koji Kagotani. 2013. Optimal Committee Performance: Size versus Diversity. draft. (2013).

Lirong Xia. 2016. Bayesian estimators as voting rules. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence*. 785–794.

Lirong Xia and Vincent Conitzer. 2011. A Maximum Likelihood Approach towards Aggregating Partial Orders. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*. Barcelona, Catalonia, Spain, 446–451.

H. Peyton Young. 1988. Condorcet's Theory of Voting. *American Political Science Review* 82 (1988), 1231–1244.

6 Appendix: Proofs

6.1 Proof of Lemma 4

Lemma 4 [NEAR-CENTRAL BINOMIAL COEFFICIENT BOUND] *For* $L > 1$,

$$\frac{3}{4} \cdot \frac{4^{L/2}}{\sqrt{\pi L}} \le \binom{L}{\lceil \frac{1}{2}(L-1) \rceil} \le 2 \cdot \frac{4^{L/2}}{\sqrt{\pi L}}.$$

Proof. Consider L odd/even and use the well known bound on the central binomial coefficient $4^k/\sqrt{2\pi k} \le \binom{2k}{k} \le 4^k/\sqrt{\pi k}$. $\square$

6.2 Proof of Lemma 5

Lemma 5 [BOUNDING $R_-(L,p)$ AND $R_+(L,p)$] *For* $\frac{1}{2} < p < 1$,

$$1 - \left(\frac{2}{(2p-1)}\right) \cdot \frac{(4p(1-p))^{L/2}}{\sqrt{\pi L}} \le R_-(L,p) \le R_+(L,p) \le 1 - \frac{3}{8p} \cdot \frac{(4p(1-p))^{L/2}}{\sqrt{\pi L}}.$$

Proof.

$$\begin{aligned}
R_-(L,p) =& 1 - \sum_{\ell=0}^{\lfloor \frac{L-3}{2} \rfloor} \binom{L-1}{\ell} p^{\ell}(1-p)^{L-1-\ell} \\
\ge& 1 - \frac{p}{2p-1} \binom{L-1}{\lfloor \frac{L-3}{2} \rfloor} p^{\lfloor \frac{L-3}{2} \rfloor}(1-p)^{L-1-\lfloor \frac{L-3}{2} \rfloor} && (Lemma\ 3) \\
\ge& 1 - \frac{1}{(2p-1)} \binom{L}{\lceil \frac{L-1}{2} \rceil} p^{\lfloor \frac{L-1}{2} \rfloor}(1-p)^{L-\lfloor \frac{L-1}{2} \rfloor} \\
\ge& 1 - \left(\frac{2}{2p-1}\right) \cdot \frac{4^{L/2}}{\sqrt{\pi L}} p^{\lfloor \frac{L-1}{2} \rfloor}(1-p)^{L-\lfloor \frac{L-1}{2} \rfloor} && (Lemma\ 4) \\
\ge& 1 - \left(\frac{2}{(2p-1)}\right) \cdot \frac{(4p(1-p))^{L/2}}{\sqrt{\pi L}}
\end{aligned}$$

The second inequality is because $\binom{L-1}{\lfloor \frac{L-3}{2} \rfloor} < \binom{L}{\lceil \frac{L-1}{2} \rceil}$.

$$\begin{aligned}
R_+(L,p) =& 1 - \sum_{\ell=0}^{\lfloor \frac{L-1}{2} \rfloor} \binom{L-1}{\ell} p^{\ell}(1-p)^{L-1-\ell} \\
\le& 1 - \binom{L-1}{\lfloor \frac{L-1}{2} \rfloor} p^{\lfloor \frac{L-1}{2} \rfloor}(1-p)^{L-1-\lfloor \frac{L-1}{2} \rfloor} && (Lemma\ 3) \\
\le& 1 - \frac{1}{2p} \binom{L}{\lceil \frac{L-1}{2} \rceil} p^{\lfloor \frac{L+1}{2} \rfloor}(1-p)^{L-\lfloor \frac{L+1}{2} \rfloor} \\
\le& 1 - \frac{1}{2p}\frac{3}{4} \cdot \frac{4^{L/2}}{\sqrt{\pi L}} p^{\lfloor \frac{L+1}{2} \rfloor}(1-p)^{L-\lfloor \frac{L+1}{2} \rfloor} && (Lemma\ 4) \\
\le& 1 - \frac{3}{8} \cdot \frac{1}{p} \cdot \frac{(4p(1-p))^{L/2}}{\sqrt{\pi L}}
\end{aligned}$$

The second inequality is because $\binom{L-1}{\lfloor \frac{L-1}{2} \rfloor} \geq \binom{L}{\lceil \frac{L-1}{2} \rceil}/2$. $\square$

6.3 Proof of Lemma 7

Lemma 7 [POISSON-BINOMIAL MAJORITY*A Poisson-Binomial random variable $X = x_1 + \cdots + x_n$ is a sum of n independent Bernoulli trials x_i with respective (possibly different) success-probabilities p_i. Let q be the average probability, $nq = \sum_i p_i$ and suppose $q \leq \frac{1}{2}$. Then, the probability to get a strict majority of successful trials is bounded by a constant. Specifically,*

1. *If $n = 2K$ is even:*

$$\mathbb{P}[X > n/2] \leq 1 - B, \quad \textit{where} \quad B = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i} q^i (1-q)^{n-i} \leq \frac{1}{2} \tag{2}$$

2. *If $n = 2K + 1$ is odd and $n \geq 1/(1-2q)$ then (2) holds. Otherwise,*

$$\mathbb{P}[X > n/2] \leq \sqrt{1-B}, \quad \textit{where} \quad B = \sum_{i=0}^{n} \binom{2n}{i} q^i (1-q)^{2n-i} \leq \frac{1}{2}$$

Proof. For even n, we can use (Hoeffding, 1956, Theorem 4, equation (26)), because $K = n/2 \geq nq$. Thus, $\mathbb{P}[X \leq n/2] \geq \sum_{\ell=0}^{n/2} \binom{n}{\ell} q^\ell (1-q)^{n-\ell}$, from which the bound follows. Suppose n is odd, $n = 2K+1$. If $n \geq 1/(1-2q)$, then $n - 2nq \geq 1$ or $\frac{1}{2}(n-1) \geq nq$. Again, we can use (Hoeffding, 1956, equation (26)) to get $\mathbb{P}[X \leq \frac{1}{2}(n-1)] \geq \sum_{\ell=0}^{n/2} \binom{n}{\ell} q^\ell (1-q)^{n-\ell}$.

When $\frac{1}{2}(n-1) < nq$ we can apply (Hoeffding, 1956, equation (25)), but analyzing $Q(K,q)$ in (Hoeffding, 1956, equation (27)) is tricky. Instead we use an indirect approach. Replicate the ensemble of probabilities to $p_1, \ldots, p_n, p_1, \ldots, p_n$, giving another Poisson-Binomial random variable Z, where $Z = X_1 + X_2$ and X_1 and X_2 are i.i.d. having the same distribution as X. Let $x = \mathbb{P}[X \leq n/2]$. Since $Z \leq n$ implies $X_1 \leq n/2$ or $X_2 \leq n/2$, we have:

$$\mathbb{P}[Z \leq n] \leq \mathbb{P}[X_1 \leq n/2 \text{ or } X_2 \leq n/2] = 2x - x^2.$$

The average probability for Z is $q \leq \frac{1}{2}$, so we can apply (2), $\mathbb{P}[Z \leq N] \geq B$, where $B = \sum_{\ell=0}^{n} \binom{2n}{\ell} q^\ell (1-q)^{2n-\ell}$. Therefore, $2x - x^2 \geq B \implies x^2 - 2x + B \leq 0$. The roots of this quadratic are $1 \pm \sqrt{1-B}$, so

$$(x - (1 + \sqrt{1-B}))(x - (1 - \sqrt{1-B})) \leq 0.$$

Since $x < 1$, the second term must be non-negative, that is $x \geq 1 - \sqrt{1-B}$, or $1 - x \leq \sqrt{1-B}$, from which the bound follows. $\square$

6.4 Proof of Theorem 8

Theorem 8 *For $d \geq 2$, let ρ be a d-issue group competence function with monotonic marginals μ_i.*[7] *Then, ρ is consistent w.r.t. every issue if and only if every marginal is consistent, i.e. for all $i \in \{1, \ldots, d\}$, there exists K_i such that $\mu_i(K_i) > \frac{1}{2}$.*

Proof. $\Leftarrow$: For every $i \leq d$, let K_i be an arbitrary number such that $\mu_i(K_i) \geq \frac{1}{2} + \epsilon_i$, where $\epsilon_i > 0$. Let $K_* = \max_i K_i$ and let $\vec{H}_{K_*}$ denote the partition function with homogeneous group size, that is, at most one group is allowed to have K_* or more members, and all groups have exactly K_* members. Let $L_* = \lfloor n/K_* \rfloor$. For $\ell = 1, 2, \ldots, L_*$, let $\mathbf{x}_{\ell,i}$ be i.i.d. binary random variables with probability at least

[7] Technically, all we need is that $\mu_i(K)$ can be lower-bounded by an increasing function of K, and that this lower bound exceeds 0.5 for some K_i.

$\frac{1}{2} + \epsilon_i$ to take 1. For any i, let $X_i = \sum_{\ell=1}^{L_*} \mathbf{x}_{\ell,i}$. It follows that $\mathbb{E}[X_i] = L_*(\frac{1}{2} + \epsilon_i)$. By the Hoeffding bound, we have that for any i:

$$\mathbb{P}\left[\sum_{\ell=1}^{L_*} \mathbf{x}_{\ell,i} \leq \frac{L_*}{2}\right] = \mathbb{P}\left[X_i \leq \frac{L_*}{2}\right] = \mathbb{P}[X_i - \mathbb{E}[X_i] \leq -L_*\epsilon_i] \leq e^{-2L_*\epsilon_i^2}.$$

Therefore, using the union bound, we have

$$R_n^d(\vec{K}, \rho) \geq 1 - \sum_{i=1}^{d} \mathbb{P}[I_i \text{ is wrong}] \geq 1 - \sum_{i=1}^{d} \mathbb{P}\left[X_i \leq \frac{L_*}{2}\right] \geq 1 - de^{-2\lfloor n/K_* \rfloor \min_i(\epsilon_i^2)},$$

which goes to 1 as $n \to \infty$.

$\Rightarrow$: If for some $i \leq d$, for all K we have $\mu_i(K) \leq 0.5$, then as in the proof of Theorem 1, the probability that the majority voting over issue I_i is bounded above by $1/\sqrt{2}$, so ρ is not consistent. $\square$

6.5 Proof of Lemma 8

Lemma 8 *For any multi-issue representative democracy with group competence function ρ and given n, K, let the marginals be μ_i and let $q = \min_i \mu_i(K)$ be the minimum marginal. Then,*

$$1 - d\left(\frac{2}{(2q-1)}\right) \cdot \frac{(4q(1-q))^{L/2}}{\sqrt{\pi L}} \leq Q_-(L, \rho(K)) \leq Q_+(L, \rho(K)) \leq 1 - \frac{3}{8q} \cdot \frac{(4q(1-q))^{L/2}}{\sqrt{\pi L}}$$

Proof. The probability to get all issues correct is at most the probability to get any one issue correct, therefore, for any K we have

$$Q_+(L, \rho(K)) \leq \min_i R_+(L, \mu_i(K)) \leq R_+(L, \min_i \mu_i(K)) = R_+(L, q),$$

where the last step follows by monotonicity (Lemma 1). The upper bound in the lemma follows after Lemma 5.

For the lower bound, by the union bound, getting one of the issues wrong is at most the sum of the probabilities to get each issue wrong, that is,

$$1 - Q_-(L, \rho(K)) \leq \sum_{i=1}^{d}(1 - R_-(L, \mu_i(K))) \leq \sum_{i=1}^{d}(1 - R_-(L, q)) = d(1 - R_-(L, q)),$$

where, in the penultimate step we used monotonicity (Lemma 1). The lower bound now follows by applying Lemma 5. $\square$