Optimal Committee Performance: Diversity, Bias, and Size

Peter Stone *

Koji Kagotani †

April 6, 2013

Abstract

The Condorcet Jury Theorem (CJT), together with a large and growing literature of ancillary results, suggests two conclusions. First, large committees outperform small committees, other things equal. Second, heterogeneous committees can, under the right circumstances, outperform homogeneous ones, again other things equal. But this literature has done little to bring these two conclusions together. This paper compares the respective contributions of size and difference to optimal committee performance, and draws policy recommendations using these comparisons.

*Peter Stone is the corresponding author and assistant professor in the Department of Political Science, University of Dublin, Trinity College. He can be contacted at pstone@tcd.ie.

†Koji Kagotani is assistant professor in the Department of Political Science, University of Dublin, Trinity College. He can be contacted at kagotank@tcd.ie.

1 Introduction

The Condorcet Jury Theorem (CJT) established in the late eighteenth century that when it comes to decision-making, size matters. Other things equal, a large committee will outperform a small one. But a growing literature suggests that *difference* matters as well as size. Other things equal, a heterogeneous committee, with people representing different backgrounds and perspectives, will outperform a committee of clones. Recent generalizations of the CJT model demonstrate the truth of this claim, although they also note that heterogeneity only produces beneficial effects under moderately restrictive circumstances. This paper uses these generalizations to compare the respective merits of size and difference. It then makes policy recommendations based upon this comparison.

Section 2 reviews existing results on the contributions size and diversity make to decision-making. These results rely upon a generalized form of the CJT model. Section 3 derives expressions for measuring committee success in a collective decision-making environment characterized by diversity. Section 4 employs simulations to generate comparative statics for the respective contributions of size and diversity within this model. This enables comparisons of these contributions under a variety of assumptions. These comparisons are conducted in terms of both expected group utility and the distribution of correct votes. Section 5 concludes by exploring the policy implications of the results.

2 Size and Diversity

In his classic work *Essai sur l'Application de l'Analyse á la Probabilité des Décisions Rendues á la Pluralité des Voix* (1785), Condorcet imagined a committee charged with making some decision using simple majority rule. The decision is dichotomous, with the committee facing a choice between options a and b. One option is unambiguously better than the other, although the identity of the better option is obviously not known in advance. All committee members share a common utility function which is maximized if the correct option is chosen. The committee is comprised of $n = 2k+1$ members (i.e., n is odd), each of whom decides correctly with fixed common probability p, with $\frac{1}{2} < p < 1$. The probability that one committee member votes correctly is independent of the probability that any other committee member votes correctly. Under these conditions, Condorcet demonstrated that the following two results hold: 1) the probability that the committee majority will decide correctly is higher than p; and 2) the probability that the committee majority will decide correctly approaches 1 as the size of the committee increases (Baker 1976). Committee size matters, other things equal.

Relaxing the (admittedly rigid) assumptions underlying the CJT can be accomplished without undermining the central results, so long as it is done with caution. Grofman, Owen, and Feld (1983), for example, showed that majorities still outperform the average individual in a group even if p varies by individual, so long as the distribution of values of p is symmetric. Boland (1989) demonstrated that the distribution of values of p does not matter, so long as the average value of p is sufficiently high. Paroush (1997) has demonstrated the importance of Boland's result, by showing that having the average value of p exceed $\frac{1}{2}$ is not a sufficient condition for the CJT's result under certain distributions of voter competence. Owen, Grofman, and L. Feld (1989), however, demonstrated that an average value of p exceeding $\frac{1}{2}$ is sufficient for the asymptotic result (i.e., for the probability of majority success to approach 1 as committee size increases).

Other factors influence, not the results regarding committee performance, but their application. The committee may, for example, know in advance that one option is ex ante more likely than the other to be correct. Assume, without loss of generality, that option a is correct with probability $\pi \geq \frac{1}{2}$. Then the option that is maximally likely to be correct when all information is taken into account may not be the same as the option that is maximally likely to be correct when only the individual judgments are considered. In particular, the optimal voting rule will have a presumption in favor of a, with b winning only if a sufficiently large supermajority favors it. Alternatively, the committee's utility function could treat differently a wrongful rejection of a and wrongful rejection of b. A jury might be more concerned about convicting the innocent than acquitting the guilty. Suppose the group loses utility e_a by wrongfully rejecting a and e_b by wrongfully rejecting b. (The group receives utility 0 if it chooses correctly.) Then the rule that maximizes the probability of successful choice, in terms of committee member judgments, may not be the rule that maximizes expected utility. The committee might prefer to sacrifice some probability of correctly choosing b in exchange for a smaller increase in the probability of correctly choosing a, because the cost of failing to choose a correctly is sufficiently high. Once again, this translates into a voting rule with a presumption in favor of a.

In practice, then, there is a close link between the ex ante probability that each option is correct and the costs associate with each type of mistake. A committee with $\pi \geq \frac{1}{2}$ and $e_a = e_b$ will maximize expected utility with a certain supermajority voting rule that favors a. For every such value of π, there exists a committee with unequal values of e_a and e_b and $\pi = \frac{1}{2}$ that optimizes performance using the same voting rule. As a result, the literature has tended to treat these two cases together (e.g. Nitzan and Paroush 1984, 1994; Ben-Yashar, Koh, and Nitzan 2009).[1]

Much less has been said regarding what (if any) contribution difference makes in the CJT environment. One possible way of capturing this contribution is to assume that homogeneity leads to violations of the independence assumption. Given this assumption, heterogeneity improves decision-making by decreasing the average correlation level between individual judgments. This in turn results in better committee decision-making (Ladha 1992). Another approach is to compare homogeneous and heterogeneous committees for a fixed value of p. This is the approach taken by (Kanazawa 1998) and (Fey 2003). But this approach leaves no room for the marginal contribution of diversity. Other things equal, if a committee adds one more member, it should add one with the highest possible competence level. There is thus

[1] See also Chwe (1999), which demonstrated that a committee whose members have different prior beliefs regarding the correct option is equivalent to a committee whose members have the same prior beliefs but different attitudes towards risk.

no positive advantage to including people of different backgrounds on the Kanazawa/Fey approach, assuming independence is assured.

A third approach is to model the differential contributions made by different groups explicitly. This can be accomplished by introducing the idea of *bias* into the decision. Assume again that option a is correct *ex ante* with probability π. Assume further that the committee can be composed of members from two different groups, X and Y. Members of X select a when a is correct with probability q, and b when b is correct with probability p, with $q > p > \frac{1}{2}$. Members of Y select a when a is correct with probability p, and b when b is correct with probability q. Thus, X-members are biased in favor of a, whereas Y-members have an equal and opposite bias in favor of b. Finally, allow the error costs e_a and e_b to vary. All committee members, regardless of group, receive utility 0 for a correct choice, pay cost e_a if a is wrongfully rejected, and pay cost e_b if b is wrongfully rejected.

Stone (2012) demonstrates that given these assumptions (and assuming the committee votes via majority rule) the optimal committee composition will be a function of p, q, π, e_a, and e_b. The expected utility-maximizing committee will have $k + s^*$ X-members and $k - s^* + 1$ Y-members on it, where s^* is the smallest integral value of s between k and k for which the inequality $\frac{e_a}{e_b}\frac{\pi}{1-\pi} > \left[\frac{p(1-p)}{q(1-q)}\right]^s$ fails to hold. Should the inequality hold for $s = k$, then an all Y-member committee would be optimal. And should it fail to hold for $s = -k$, then an all X-member committee would maximize expected utility. Moreover, this condition requires that $q > 1-p$, a condition ensured whenever $p > \frac{1}{2}$.[2] If $q \leq 1-p$, then heterogeneity never increases, and sometimes decreases, expected committee utility. Heterogeneity thus may, but not necessarily will, make a difference for the better.

Stone's model suggests that committee decision-making can be improved either by taking steps to ensure optimal committee composition or by increasing committee size. Stone's original result, however, does not consider the respective magnitudes of these two approaches. When does size matter a lot, and when does difference matter a lot? Under what conditions

[2]This condition is minimal but not negligible. If it were not satisfied, then an X-member would be more likely to select a when b is correct than when a is correct!

will taking steps to ensure diversity prove more cost-effective than simply involving more people? These are the questions explored in this paper.

3 Comparing Size and Diversity

Committee performance can be assessed either in terms of the probability that the committee decides correctly via majority rule, or in terms of the expected utility of committee decision-making. (The one is equal to the other whenever the costs associated with the two types of error are the same.) We will begin, therefore, by deriving expressions for both of these measures.

Assume a committee with k+s X-members and $k-s+1$ Y-members. Then conditional upon the state of the world (say, a), committee decision making functions as a pair of sets of Bernoulli trials. It is equivalent to making multiple tosses of two coins, one of which has a different bias towards heads than the other. Coin 1 gets tossed $k+s$ times, and lands heads up with probability q, while coin 2 gets tossed $k-s+1$ times, and comes up heads with probability p. Let x be the total number of correct votes by the X-members, y be the number of correct votes by the Y-members, and $z=x+y$ be the number of correct votes in a committee. Then the expected counts of correct votes by different types of members are calculated given the state of the world: $E(x \mid a)=(k+s)q$, $E(x \mid b)=(k+s)p$, $E(y \mid a)=(k+s-1)p$, and so forth. One can use these to calculate the (unconditional) expected counts of correct votes in the committee: $E(z)=(k+s)[\pi q+(1-\pi)p]+(k+s+1)[\pi p+(1-\pi)q]$.

One can use this measure as a building block in order to calculate C_a (or C_b), the probability that the committee decides correctly via majority rule, conditional on a (or b) being correct. Then one can also calculate C, the unconditional probability of a correct committee vote:

$$
\begin{aligned}
C =&\pi C_a + (1-\pi)C_b \\
=&\pi \Sigma_{i=0}^{k} \left[\binom{k+s}{s+i} (1-q)^{s+i} q^{k-i} \left[\Sigma_{j=0}^{i} \binom{k-s+1}{j} (1-p)^{k-s+1-j} p^j \right] \right] \\
&+ (1-\pi)\Sigma_{i=0}^{k} \left[\binom{k+s}{s+i} (p)^{s+i} (1-p)^{k-i} \left[\Sigma_{j=0}^{i} \binom{k-s+1}{j} (q)^{k-s+1-j} (1-q)^j \right] \right].
\end{aligned} \tag{1}
$$

Finally, let u be the utility derived from a committee vote. This vote will yield utility 0 with probability $\pi C_a + (1-\pi)C_b$; utility e_a with probability $\pi(1-C_a)$, and utility e_b with probability $(1-\pi)(1-C_b)$. The expected utility $E(u)$ of a committee vote is thus:

$$e_a \pi(1-C_a) + e_b(1-\pi)(1-C_b). \tag{2}$$

With these preliminaries out of the way, it becomes possible to compare the respective contributions of size and diversity both to C and to $E(u)$. As mentioned before, whenever $e_a = e_b$ the expected utility of the committee is maximized if and only if the probability of correct committee choice is maximized.[3] For this reason, it makes sense to begin by investigating this special case.

As discussed in the previous section, the heterogeneous committee achieves the highest probability of correct committee vote under certain conditions (Stone 2012). This implies that as the parameter s increases beyond a certain point, the probability of correct committee vote decreases. However, the probability of correct committee decision-making C is determined by the multiple parameters n, p, q, s and π, with the effect of each parameter being the function of other four. Even though we can explore the sign of the impact analytically, it is very difficult to examine the relative impact of the committee diversity, bias, and size on its performance. If we can repeat the committee's decision-making a certain number of times (for example, 10,000 times), we can calculate the probability of the correct committee vote given a set of the parameters, which is close to the objective probability of correct

[3] If $e_a = e_b$, then $E(u) = e_a[\pi(1-C_a) + (1-\pi)(1-C_b)]$. This means that we can find a maximum of $E(u)$ by maximizing both C_a and C_b.

collective choice. Hence, we use the Monte Carlo simulation approach to further explore the relationship between the relative impact of each parameter and the optimal committee.

4 Monte Carlo Simulation

Our simulation procedures are as follows. First, the state of the world $\in \{a, b\}$ is drawn from the bernoulli distribution with $\Pr(a) = \pi$. Second, the counts of correct votes in the committee are generating as the sum of a pair of binomial distributions $\mathrm{B}(\mathrm{X}, q) + \mathrm{B}(\mathrm{Y}, p)$ if the state of the world a is chosen, and $\mathrm{B}(\mathrm{X}, p) + \mathrm{B}(\mathrm{Y}, q)$ otherwise. Third, if the majority of the committee ($k+1$ members) vote correctly, this is called committee success. Fifth, we iterate this procedures 10,000 times for a given set of parameters n, p, q, s, and π, and then calculate the ratio of committee success, which is called here the probability of committee success. As a further simplifying assumption, we begin with $\pi = \frac{1}{2}$ and relax it later. We address the three theoretical issues below.

1. **Homogeneity versus heterogeneity.** If $e_a = e_b$ and $\pi = \frac{1}{2}$, then $s^* = 0$. Committee performance is optimized by the presence of k X-members and $k+1$ Y-members. It is also, however, optimized by the presence of $k+1$ X-members and k Y-members. A bare majority from one group, whether it be X or Y, yields the exact same probability of committee success. In comparing the relative impacts of size and diversity, it makes sense to consider the maximum possible contribution that diversity can make for a given size committee. We will therefore compare, for a set of given value of n, the performance of an all-X-member committee with the performance of a committee with a bare majority ($k+1$) of X-members on it. We will call these the *homogeneous* and the *heterogeneous* committees, respectively.

2. **Bias (the relative sizes of q and p).** Committee composition could poten-

tially matter much more when the biases of committee members are strong than when they are weak. We shall consider a high-bias and a low-bias committee. A *high-bias* committee has $q = 0.85$ and $p = 0.51$. A *low-bias* committee has $q = 0.55$ and $p = 0.51$.

3. **Committee size.** Obviously, the higher the value of n, the better the committee will perform. We will take this into account by comparing several values of n. Specifically, we will consider n = 11, 21, 51, 101, 201 and 401. All of these are realistic committee sizes for at least certain classes of decisions.

Figure 1 shows the simulation results for the low-bias committee ($p = 0.51$ and $q = 0.55$) and the high-bias committee ($p = 0.51$ and $q = 0.85$). The committee performance is measured by the probability of correct committee vote. To provide the intuitive understanding of the patterns, we use the degree of homogeneity $\frac{k+s}{n} \times 100$ to examine the committee composition. When we increase s from 1 to $k + 1$, the committee becomes progressively more homogeneous. It changes from approximately 50 percent X-members to 100 percent X-members. The parameter $\pi = 0.5$ means that two states of the world a and b are equal likely to occur.

We observe several results. First, more homogeneity worsens high-bias committee performance, which is consistent with the theoretical arguments by (Stone 2012). The difference is particularly striking for high values of n. For these values, heterogeneous high-bias committees perform nearly perfectly, while homogeneous high-bias committees still have a long way to go before C approaches 1. In contrast, heterogeneity matters little for the performance of the low-bias committee of relatively small size ($n \leq 101$) because the probability of committee success remain the same over different values of the parameters. For the low bias committee of relatively large size ($n > 101$), more homogeneity does constrain the committee performance.

Second, increasing committee size enhances performance regardless of homogeneity level. With $\pi = 0.5$, the most heterogenous committee shows the best performance given a set of the parameters except for the low-bias committee cases with the relatively small number of members, where it makes a negligible difference.

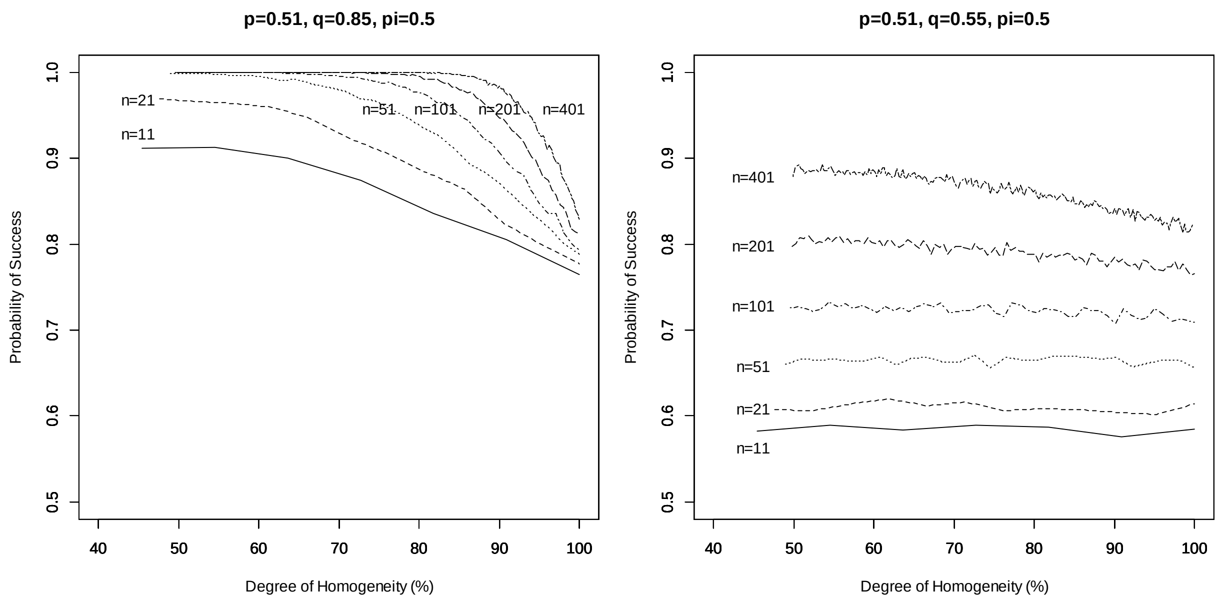

Figure 1: the High-Bias Committee and the Low-Bias Committee

It is not too surprising that the high-bias committees routinely outperform the low-bias committees here, for all values of n. *Ex ante* (i.e., before the state of the world is selected), both X-members and Y-members have an expected competence level of .68 in the high-bias model, but only .53 in the low-bias model. It is therefore difficult to distinguish the effects generated by having a more-or-less competent committee from the effects generated by bias size. The address this problem, we ran simulations for two further models. In both of these models, X- and Y-members have the same /ephex ante competence level, equal to .625. In the modified high-bias model, q = .74 and p = .51, while in the modified low-bias model, q = .65 and p = .6.

The results generated here are similar but not identical. First, once again the heterogeneous committee is clearly best in the presence of high bias while committee composition hardly matters for the low-biased committee. Second, when the *ex ante* probability of success is the same for both committees, with the heterogeneous composition, the probability of success in the low-bias committee is almost the same as that in the high-bias committee. In contrast to the previous results, as the committee composition becomes homogeneous, the low-bias committee clearly begins to outperform the high-bias committee. Third, increasing committee size once again increases the probability of committee success. In the high-bias committee, increasing the size raises the probability of committee success and reduces the relative effect of homogeneity. The heterogeneous committee success rate rapidly approaches 1, and so its comparative advantage lessons as the homogeneous committee gets larger as well. (It has nowhere to go.) On the other hand, the size of the committee matters for the relatively small committee because the probability of committee success is almost one for the committee with more than 100 members.

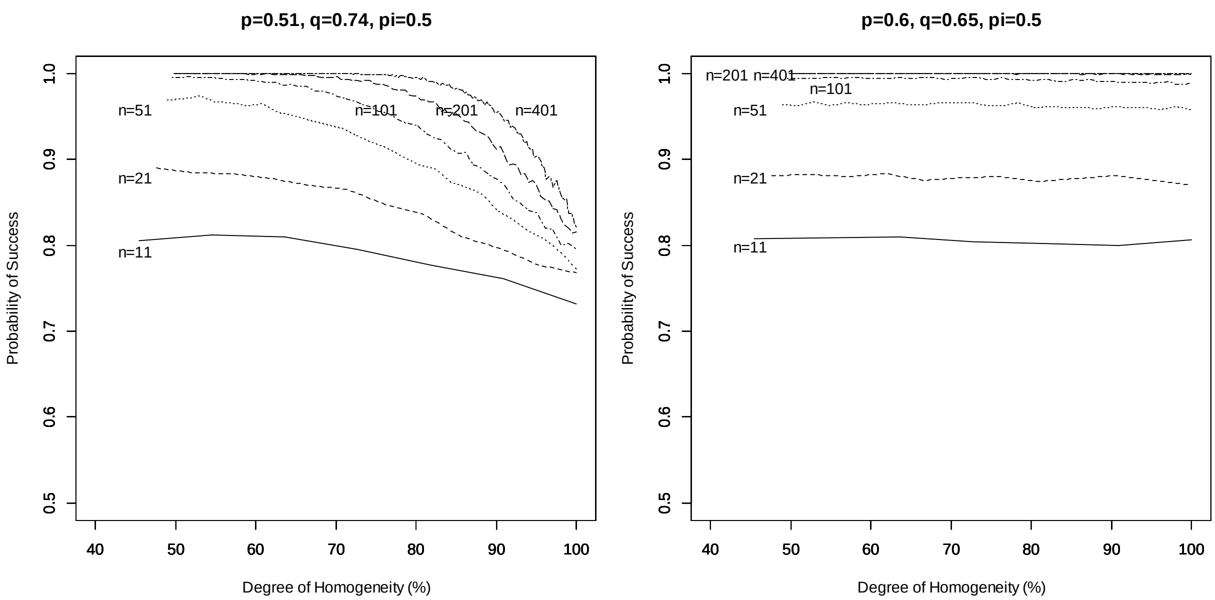

Figure 2: the High-Bias Committee and the Low-Bias Committee

Next we relax our assumption regarding the states of the world and assume $\pi = 0.75$, so that state of the world a is significantly more likely to occur than state of the world b (Figure 3). We employ the same values of p and q as in the first set of simulations and consider the results. First, both the high-bias and the low-bias committees demonstrate interesting results regarding the composition of the committee. The high-bias committee with more than 50 members generates results similar to those previously seen. That is, the best committee is the most heterogeneous and more homogeneity dampens the committee performance. Things are somewhat different for the high-bias committees with 11 and 21 members. The reason for this is that when $\pi = 0.75$, $s^* = 2$. The optimal committee will thus have 5 more X-members than Y-Members on it. This makes a significant difference for small high-bias committees. The optimal 11- and 21-member committees will have 7 and 12 X-members, respectively; put another way, the percentage of these committees comprised of X-members will be 63% and 57%. Maximal heterogeneity is thus no longer optimal heterogeneity, even though homogeneity still does worse.

A similar, but much stronger, result obtains for the low-bias committee. For this case, $s^* = 114$. This means that, for every low-bias committee of less than maximal size, the optimal composition will involve no Y-members at all. Even for the 401-member committee, the optimal committee will have a 314-87 majority in favor of X-members. It will be 78% X-member. The simulation results clearly reflect this, as the homogeneous low-bias committee outperforms any heterogeneous one (albeit marginally), except when n=401. These results are caused by the assumption on the states of the world $\pi = 0.75$. Since the sate of the world a is more likely to occur, the committee can improve its performance by including more X-members who are more likely to detect the states of world a.

5 Conclusion

The simulation results impart several important lessons regarding committee performance in the presence of bias. Most importantly, heterogeneity is more likely to make a difference, for better or for worse, when the committee members have high bias. By "high bias," we mean a large difference between the values of p and q. In the presence of high bias, committee composition will matter much more. The optimal committee will significantly outperform committees with less-than-optimal composition. This will be true regardless of

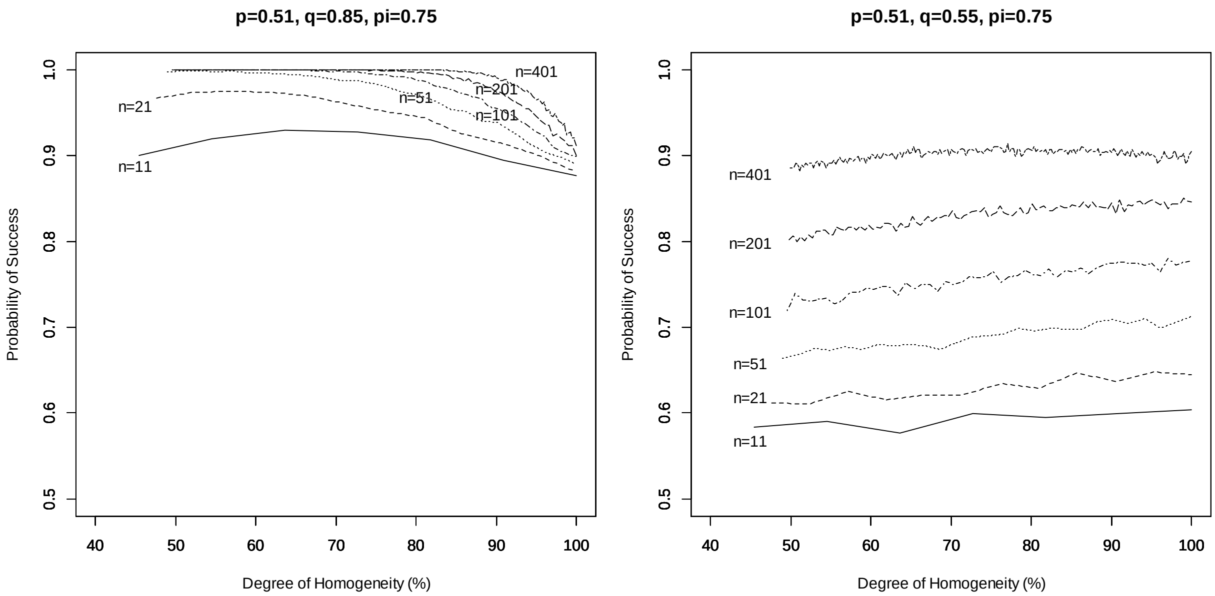

Figure 3: the High-Bias Committee and the Low-Bias Committee

the value of s*. If the optimal committee is homogeneous, then a homogeneous high-bias committee will significantly outperform a heterogeneous one. If the optimal committee is heterogeneous, then the opposite will occur. This difference will tend to shrink as n gets large for one obvious reason. As n gets large, the optimal committee will decide correctly with a probability near one. And so for large values of n, the difference between the best committee's performance (which hovers near 1) and the worst will of necessity grow smaller. Note, however, that even for the largest high-bias committees, poor committee composition will significantly worsen committee performance. A 401-member high-bias committee that is composed well will perform with near-perfection, but a 401-member high-bias committee that is composed poorly will lag signicantly behind.

All of this suggests that, if a policy maker had control over the values of p and q, she should prefer, other things equal, for the values to be close together rather than further appart. The precise manner in which this works remains a subject for further study. Such study could reveal, for example, how much the difference between p and q matters relative to the absolute sizes of the two variables. The present simulation results suggest that both factors will considerably influence committee success.

Normally, however, policy makers have no control over the existing biases within the population. All they control is 1) how many people from that population get selected and 2) which people they select. (Policy makers could, of course, try to find people with low levels of bias, an intuitively obvious idea that our study supports.) If the policy maker has no control over p and q, then the relative attention she gives to committee size versus committee composition should vary according to the size of the bias. In the presence of high bias, committee composition matters a great deal. In the presence of low bias, however, committee composition matters little, and the only way to improve committee performance significantly is to increase n. Given that selection criteria can prove costly, especially if employed on a large scale, this conclusion could matter significantly in policymaking contexts. Policy makers facing decisions in which the population is strongly divided may wish to select smaller committee within which diversity of opinion is ensured. With low bias, policy makers may wish to forego consideration of diversity and simply assemble a large committtee.

References

Baker, Keith M., ed. 1976. *Condorcet: Selected Writings.* Indianapolis: Bobbs-Merrill.

Ben-Yashar, Ruth, Winston T.H. Koh, and Shmuel Nitzan. 2009. "Is Specialization Desirable in Commitee Decision Making?" Department Working Paper 2009-16, Economics Department, Bar Ilan University, Ramat Gan, Israel.

Boland, Philip J. 1989. "Majority Systems and the Condorcet Jury Theorem." *The Statistician* 38(3): 181–189.

Chwe, Michael Suk-Young. 1999. "Minority Voting Rights Can Maximize Majority Welfare." *American Political Science Review* 93(1): 85–97.

Fey, Mark. 2003. "A Note on the Condorcet Jury Theorem with Supermajority Voting Rules." *Social Choice and Welfare* 20: 27–32.

Grofman, Bernard, Guillermo Owen, and ScottL Feld. 1983. "Thirteen Theorems in Search of the Truth." *Theory and Decision* 15: 261–278.

Kanazawa, Satoshi. 1998. "A Brief Note on a Further Refinement of the Condorcet Jury Theorem for Heterogeneous Groups." *Mathematical Social Sciences* 35(1): 69 – 73.

Ladha, Krishna K. 1992. "The Condorcet Jury Theorem, Free Speech, and Correlated Votes." *American Journal of Political Science* 36(3): 617–634.

Nitzan, Shmuel, and Jacob Paroush. 1984. "Are Qualified Majority Rules Special?" *Public Choice* 42(3): 257–272.

Nitzan, Shmuel, and Jacob Paroush. 1994. "A General Theorem and Eight Corollaries in Search of Correct Decision." *Theory and Decision* 17: 211–220.

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

Paroush, Jacob. 1997. "Stay Away from Fair Coins: A Condorcet Jury Theorem." *Social Choice and Welfare* 15: 15–20.

Stone, Peter. 2012. "Introducing Difference into the Condorcet Jury Theorem." Presented at the Annual Meeting of the European Political Science Association, 2012.