

# Arrow's and Gibbard-Satterthwaite's Impossibility Theorems Revisited

## Abstract

Arrow<sup>1</sup> proved that the amalgamation of individual preference orderings into a social ordering was impossible if it met his five normative conditions. Later Gibbard<sup>2</sup> and Satterthwaite<sup>3</sup> proved that any such amalgamation of individual preference orderings that was nonmanipulable, (i.e. wherein there was no advantage to any individual to use strategy to order their preferences insincerely to get a better result for themselves) was impossible or led to the selection of a dictator. According to Satterthwaite, there is a one-to-one correspondence between Arrow's theorem and the Gibbard-Satterthwaite theorem. We will show in this paper that a social welfare function (Arrow's term) or a voting procedure (Satterthwaite's term) in which the strategy is inherent in the voting procedure itself and applies to all choosers or voters can lead to a system in which there is no advantage to individuals to misrepresent their preference orderings or ratings. The voting procedure itself can filter the sincere inputs from the voters in such a way as to maximize the power of each individual vote without, however, doing so such that the ordering of any voter's preferences is misrepresented. Neither would it be advantageous for any voter to vote insincerely. Therefore, Arrow's and Gibbard-Satterthwaite's theorems do not apply, and non-dictatorial voting and social welfare procedures are possible in which each voter or chooser is incentivized to vote sincerely.

## Introduction

In 1951 Arrow postulated five normative conditions that must be satisfied for a social welfare function to be rational. Those conditions were unrestricted domain, positive association of social and individual values, independence of irrelevant alternatives (IIA), citizens' sovereignty and non-dictatorship. Each voter is to order his or her preferences over all alternatives or candidates. For the purpose of this discussion each voter would submit a ballot of the form ABCD which is interpreted as A is preferred to B is preferred to C is preferred to D. Arrow also includes the possibility of indifferences among candidates which we will not consider. Other than IIA Arrow's other conditions are assumed for most voting or social welfare functions. We will limit the discussion without loss of generality to voting procedures.

Arrow's example shows how the simple dropping out of one candidate from the election list can lead to an irrational result. Let the weights assigned to an  $x$  place finish be  $5-x$ . If voter 1 and 2 have ABCD, voter 2 has CDAB, then candidate A is clearly the winner: 10 to 7 to 8 to 5. But if B is eliminated and the places are renumbered  $4-x$ , C and A tie at 7 to 7. This method is known as the Borda count. Arrow thinks this is unacceptable. However, let's consider the result if the original values are kept even though B has dropped out. The result would be 10 for A, 8 for C and 5 for D. A is still the winner. Arrow has chosen the numbering of the Borda scheme to suit his purposes, but there is no rule that says the Borda count must be applied in this manner. If candidate B drops out after the election occurs and there is no revote, then the election results would be the same as they would have been if he hadn't dropped out. This point is significant as it applies to the later discussion. The timing of the drop out is of the essence.

It is well known that some voting systems are manipulable i.e. a voter can improve the outcome of the election for himself by not voting his sincere preferences. The Borda count is certainly one of them. Consider 2 voters with the following profiles and the values 4,3,2,1 for the place finishes: voter 1:

ABCD; voter 2: BCAD. The result is 6 for A, 7 for B, 5 for C and 2 for D so B wins. If voter 1 knows how voter 2 will vote, but voter 2 doesn't know how voter 1 will vote, voter 1 can vote ACDB which gives the outcome: 6 for A, 5 for B, 6 for C and 3 for D or a tie between A and C. In a real world election voters don't know for certain how other voters will vote. They have only polling data which is based on probabilities. Our methods will incorporate probabilities based on polling data instead of certainty regarding other voters. If a candidate dies or drops out after the election occurs but before the results have been announced, then IIA is satisfied at least for the system we will consider. If a candidate drops out or dies before the vote takes place, then voters cast their votes in a different environment. A poll taken after a candidate drops out will give different results than a poll taken before the candidate drops out, and IIA doesn't hold nor should it. Again timing is of the essence.

## **Utilitarian, Approval and Optimal Threshold Voting**

Now consider utilitarian voting. It is similar to range<sup>4</sup> or score voting. Without loss of generality we pick a range between -1 and 1 in which the candidates can rate (rather than rank) the candidates. Let the utility or satisfaction or usefulness for the voter be  $u_{ic}$ ,  $-1 \leq u_{ic} \leq 1$ .  $u_{ic}$  is a real number which gives the utility of the  $i^{\text{th}}$  voter for the  $c^{\text{th}}$  candidate. Now the utilities for each candidate are added up over all voters with the winner being the one with the highest score. The way this method might be manipulable would be if a voter rated a candidate lower or higher compared to another candidate than he would if he voted sincerely. This would presuppose that this particular voter knew how other voters were going to vote or at least had some polling data. However, this kind of strategic voting does not have to be considered insincere either as long as the voter is trying to maximize his expected value of utility for the winning set of candidates. If candidate A is preferred to candidate B, but A is so low in the polls he doesn't have a chance of winning, then it is rational for the voter to rate or rank B higher than A. The voting system we will consider later, Optimal Threshold voting, takes into account that a voter is concerned both about the utility of a candidate for him and the chances that that candidate has enough support in the voting population to get elected. Using both of these elements, probability and utility, should determine what a sincere vote consists of. If there is no polling data available, Optimal Threshold voting would satisfy IIA unconditionally.

Approval voting<sup>5</sup> takes place when each voter assigns a 1 or a 0 to each candidate and the votes are added up over all candidates to determine the winner. Optimal Threshold voting<sup>6</sup> computes a threshold based on an individual voter's utilitarian ratings of the candidates and then assigns a 1 vote to those candidates above threshold and a -1 vote to those candidates below threshold. Optimal Threshold voting is, therefore, a utilitarian/approval hybrid method. It applies not only in cases where one winner is to be determined but in multiwinner elections as well.

We now turn to an analysis of Optimal Threshold voting with reference to the companion paper, "Optimal Threshold for Selection of Candidates in Multi-Winner Elections."

## **Maximizing the Expected Value of the Winning Set**

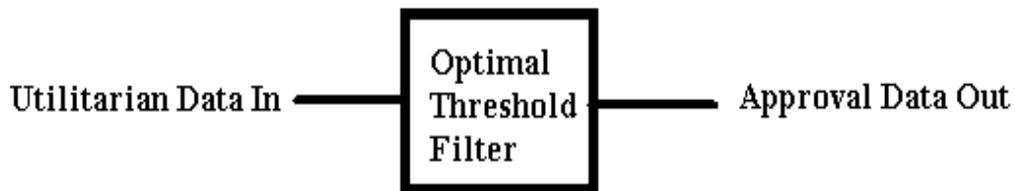
We set a threshold,  $T$ , a real number, in each individual voter's specification of utilities for associated candidates such that  $-1 \leq T \leq 1$ . The winning set is of size  $m$ . We evaluate the expected utility of the winning set of candidates for all possible thresholds and for each particular voter. The optimal threshold is the one that maximizes the expected utility of the winning set. Then each candidate above

threshold is given a 1 vote and each candidate below threshold is given a -1 vote, approval style. We, therefore, seek to maximize

$$E(V_a) = \frac{1}{n_a} \sum_{i=1}^{n_a} p_i u_i$$

where  $V_a$  is a random variable representing the utility of the winning set of candidates for a particular voter,  $E(V_a)$  is the expected value of average utility for the set of candidates above threshold and is unique for each voter,  $n_a$  is the number of candidates above threshold,  $p_i$  is the probability that the  $i^{\text{th}}$  candidate above threshold will be in the winning set and  $u_i$  is the utility of the  $i^{\text{th}}$  candidate. All candidates above threshold will be given an approval style vote of 1 and all those below threshold will be given an approval style vote of -1.

Now suppose that, instead of each individual voter computing his personal optimal threshold, that was done for him by the voting procedure or system itself. The voters would input utilitarian data into the system. The voting procedure would first pass this data through a filter which would compute optimal thresholds for each voter, and finally the output from the filter would be the approval style votes for each candidate from each voter. The model would be similar to the following:



Since the data is processed in an optimal manner for each individual voter, the voters have no incentive to misrepresent their preferences or to vote insincerely. They would either vote sincerely or the optimal threshold filter would process their input in such a way as to give them a suboptimal result. Finally, the votes are tabulated for each candidate and the winning set is proclaimed. It would consist of the  $m$  highest approval style vote getters.

Clearly, Gibbard-Satterthwaite's theorems do not apply. The voters do not have an incentive to vote insincerely and the voting system has not led to a dictator. The strategy has been placed in the processing of the votes rather in each individual voter's hands. The voters themselves are disincentivized from voting insincerely. Also their utilitarian style votes are assumed to remain the same for each candidate regardless of the environment *i.e.* they would give each candidate the same utility rating regardless of whether or not one or more candidates dropped out of the race before or after the election.

Now consider Arrow's example in which one of the candidates dies and how this affects the election results using Optimal Threshold voting. Arrow states: "Suppose that an election is held, with a certain number of candidates in the field, each individual filing his list of preferences, and then one of the candidates dies. Surely the social choice should be made by taking each of the individual's preference

lists, blotting out completely the dead candidate's name, and considering only the orderings of the remaining candidates in going through the procedure of determining a winner.” Arrow implies that the voting has already occurred, but the final determination of the winner(s) has not been made. In fact Optimal Threshold voting would produce identical results for all the other candidates if the death occurs after the election takes place as is proven in the Appendix. If the candidate dies before the election takes place, however, polling results would not be accurate. If another poll were to be taken before the actual vote took place, then the Optimal Thresholds would have to be recomputed based on those polling results. However, the voters' ballots would be exactly the same since utilitarian values for the candidates are assumed not to change regardless of the composition of the set of candidates who are standing for election. Furthermore, the social choice will be made by taking each individual's utilitarian input (utilitarian data implies preferences), blotting out completely the dead candidate's name, and considering only the ratings (which imply orderings) of the remaining candidates in going through the process of determining a winner(s). Only the Optimal Thresholds will be different in each candidate environment due to polling data, and those are determined by the voting system itself independently of the voters. IIA is satisfied unconditionally by Optimal Threshold voting if polling data is not taken into account at all, and the essence of IIA, according to Arrow's verbal example, is satisfied if it is. Arrow's mathematical statement of IIA, however, does not accurately reflect his verbal statement.

Lehtinen<sup>7</sup> has considered the use of strategy to maximize the expected value of a voter's utility for the outcome of an election. He has shown that strategic voting behavior which violates one of Arrow's conditions is actually beneficial, and, therefore, one of Arrow's conditions is not normatively acceptable. We take a slightly different approach and show that Optimal Threshold Voting does not violate any of Arrow's normative conditions as long as polling data is fixed. If a candidate dies, the results of the election are determined by blotting out the dead candidate's name from each voter's list and then going about the process of determining the winners just as Arrow suggests. The winners, however, might be different if the candidate dies before the election has taken place and if there are subsequent polling results based on the candidates still running since the voters' polling results would be redistributed among the remaining candidates in an unpredictable way. If a candidate dies or drops out after the vote has taken place or at least after the last poll has been taken before voting, Optimal Thresholds will not change, and the results of the election will be the same with the dead candidate's results blotted out. The results will be the same with the score for the candidate who has died eliminated. If he was in the winning set, the next highest vote getter would be elevated to the winning set.

Arrow's other conditions are all fulfilled by inspection. It is assumed that utility values do not change before or after the election and polling data for each candidate does not change as a result of some candidate dropping out after the voting has taken place.

## **Summary and Conclusion**

We rely heavily on the analysis presented in my previous paper, “Optimal Threshold for Selection of Candidates in Multi-Winner Elections.” This paper assumed that polling data was unknown so that the probabilities associated with each candidate ending up in the winning set were the same. In general, if polling data is known, there will be a probability of being elected associated with each candidate. For each voter we take their utilitarian style input which represents their utility for each associated candidate, and using polling data, we compute a threshold such that the normalized expected utility of

the winning set, which is the average of the sum of expected utilities of the candidates in the winning set, is a maximum. We divide the candidates into two sets according to the threshold which is computed for each voter by the voting procedure itself. Those candidates above threshold are given an approval style vote of 1. Those candidates below threshold are given an approval style vote of -1. The votes are summed over all voters, and the candidates with the  $m$  highest vote totals are assigned to the winning set of size  $m$ .

In our approach the expected utility of each individual voter's ballot is maximized by the voting procedure or system itself. The voter has no incentive to vote insincerely since the voting procedure provides the optimum strategy for each voter. If a voter voted insincerely, they would get a suboptimum result for their approval style final ballot. Gibbard states: "An individual manipulates a system of voting if, by misrepresenting his preferences, he secures a result he prefers to the result that would obtain if he expressed his true preferences." Therefore, Optimal Threshold voting satisfies strategyproofness.

Satterthwaite<sup>8</sup> states: "This result, which is reminiscent of Arrow's general possibility theorem for social welfare functions, suggests a second question. What is the relationship between the requirement for voting procedures of strategyproofness and Arrow's requirements for social welfare functions of rationality, nonnegative response, citizens' sovereignty, and independence of irrelevant alternatives? I show that they are equivalent: a one-to-one correspondence exists between every strategy-proof voting procedure and every social welfare function satisfying Arrow's four requirements. This means that if a social welfare function violates any one of Arrow's requirements, then the voting procedure which is naturally derived from the social welfare function is not strategy-proof." Since we have demonstrated a voting procedure which is strategy proof, we have demonstrated a voting system which reverses Arrow's Impossibility Theorem and also the Gibbard-Satterthwaite results.

Preferences are easily derived from ratings both for the individual and society since ratings imply preferences. Thus by delving deeper into utility ratings for each voter, we have come up with a system that implies societal preferences in the final result.

The paradox of voting<sup>9</sup>, which Arrow's Impossibility Theorem is a generalization of, was known to Condorcet in 1793. It is sufficient to prove Arrow's theorem without the mathematical elaboration that Arrow provided. Thus both Arrow's and Gibbard-Satterthwaite's theorems are narrowly conceived and mathematically tautological.

I conclude with Lehtinen's observation: "This means that while Arrow's theorem and the Gibbard-Satterthwaite [theorem] are logically impeccable, they fail to have the devastating consequences for democracy that have sometimes been attributed to them."

## **Appendix**

Theorem: With Optimal Threshold voting, if a candidate drops out of an election after voting has occurred, the results of the election will not be changed

Proof:

For some particular voter let the expected value of above threshold average utility be

$$(p_1u_1 + p_2u_2 + \dots + p_nu_n)/n$$

Assume candidate  $j$ , an above threshold candidate, drops out after votes are cast.

Then, expected value of above threshold average utility is

$$(p_1u_1 + p_2u_2 + \dots + p_{j-1}u_{j-1} + p_{j+1}u_{j+1})/(n - 1)$$

Let  $p_1u_1$  be the least above threshold value. Then the vote is changed if the optimal threshold is raised to exclude this value.

That would be true if

$$\left( \frac{\sum_{i=2, i \neq j}^n p_i u_i}{n - 2} \right) > \left( \frac{\sum_{i=1, i \neq j}^n p_i u_i}{n - 1} \right)$$

But

$$\left( \frac{n - 1}{n - 2} \right) \sum_{i=2, i \neq j}^n p_i u_i \not> \sum_{i=1, i \neq j}^n p_i u_i$$

By inspection.

Therefore, the above threshold value of average expected utility is not increased by raising the threshold.

Consider lowering the threshold. We renumber the expected utilities such that  $p_1u_1$  is the value just below threshold, and there are  $n-1$  values above threshold. Then is

$$\left( \frac{\sum_{i=1, i \neq j}^n p_i u_i}{n} \right) < \left( \frac{\sum_{i=2, i \neq j}^n p_i u_i}{n - 1} \right)$$

We know that

$$\left( \frac{\sum_{i=1}^n p_i u_i}{n} \right) < \left( \frac{\sum_{i=2}^n p_i u_i}{n - 1} \right)$$

and

$$p_i u_i < p_j u_j \text{ for } j > 1$$

So

$$p_i u_i / n < p_j u_j / n < p_j u_j / (n - 1)$$

Therefore,

$$\left( \frac{\sum_{i=1, i \neq j}^n p_i u_i}{n} \right) < \left( \frac{\sum_{i=2, i \neq j}^n p_i u_i}{n - 1} \right)$$

Since the optimal threshold doesn't have to be recomputed if an above threshold candidate drops out after the votes are cast by the voters, Arrow's IIA condition is preserved and the vote tally remains the same as if the candidate had just been blotted out of the election results. The candidate with the next highest vote total would then be added to the winning set.

## References

1. Arrow, Kenneth J. (1951), *Social Choice and Individual Values*, Yale University Press.
2. Gibbard, A. (1973), "Manipulation of voting schemes: a general result," *Econometrica*, Vol. 41, No. 4, pp. 587–601.
3. Satterthwaite, M. A. (1975), "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, 10, pp. 187–217.
4. <http://www.rangevoting.org/>
5. Brams, Steven J. and Fishburn, Peter C. (1983), *Approval Voting*, Birkhauser-Boston.
6. Lawrence, J. C., "Optimal Threshold for Selection of Candidates in Multi-Winner Elections," [https://advance.sagepub.com/articles/2019\\_Optimal\\_Threshold\\_for\\_Selection\\_of\\_Candidates\\_in\\_Multi-Winner\\_Elections/8187755/1](https://advance.sagepub.com/articles/2019_Optimal_Threshold_for_Selection_of_Candidates_in_Multi-Winner_Elections/8187755/1)
7. Lehtinen, Aki (2015), "A welfarist critique of social choice theory: interpersonal comparisons in the theory of voting," *Erasmus Journal For Philosophy And Economics*, Vol. 8, No. 2, p. 77.
8. Satterthwaite, M. A. (1975), "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions," *Journal of Economic Theory*, 10, pp. 187–217.
9. Lawrence, J. C., "A Short Proof of Arrow's Impossibility Theorem," <http://www.socialchoiceandbeyond.com/scabpage36.html>