The Problem With Range Voting and Averaging Candidate Scores

The proposal known as "Range Voting" has gotten some attention as a result of William Poundstone's book, **Gaming the Vote**. While a useful tool for rating movies and products on the Internet, where the "voters" are predominantly disinterested judges who are not highly invested in the outcome, it has severe problems in high-stakes elections where the voters genuinely care about the outcome. Prof. Nicolaus Tideman, one of the top election method experts in the world, wrote in his recent book, **Collective Decisions and Voting**, that Range Voting is one of six "unsupportable" methods that "have defects that are so serious as to disqualify them from consideration." While Tideman was largely focused on the susceptibility to strategic manipulation, there are other more fundamental problems with the underlying assumptions of Range Voting that are addressed in this brief essay.

Range Voting seeks to add together individual cardinal "utility" or "satisfaction" scores for candidates assigned by each voter to come up with an average community score. Warren Smith, the primary promoter of Range Voting, argues that even if the community score is not an accurate number due to strategic voting, lack of voter information about candidates, etc., this voting method still tends to elect the "right " candidate more than any other voting method. Key to this assertion is Smith's definition of the "right" winner. He relies on something called Bayesian Regret -- a measurement of how dissatisfied the community would be **as a whole** with each candidate's victory.

Take as an example the election of Lyndon Johnson over Barry Goldwater in 1964 by 61% to 39%. If a mind-reading machine could measure the level of satisfaction of each voter (to avoid the problem of insincere strategic voting for the moment) and found out that the Goldwater supporters were far more upset with the election of Johnson and that the Johnson supporters would have been only mildly upset by the election of Goldwater, it is possible that the Bayesian Regret calculation could show that, society-wide, there would have been greater average satisfaction if Goldwater had won, even though most voters preferred Johnson. The notion is that intensity of dissatisfaction can counter-balance sheer numbers. Smith would declare that Goldwater was the correct winner, since his Bayesian Regret score was better.

Smith then goes on to compare all voting methods based on Bayesian Regret, and shows that no voting method is as good as Range Voting at electing the candidate with the lowest Bayesian Regret score. Since Smith and other Range Voting advocates believe that Bayesian Regret is the "gold standard," they assert this proves that Range Voting is the best voting method.

Nearly all political scientists, election methods experts, and members of the general public would reject the underlying assumptions about Bayesian Regret being the "gold standard." Most people believe that some variant of the principle of majority rule is the more appropriate standard for electing a mayor or president. But even if one accepted the premise of Range Voting -- that community-wide average Bayesian Regret was the best measure of the "right" winner -- there are other problems with the method.

Many economists and political scientists reject the notion that individual cardinal scores can be combined in this way. It is widely held that when looking at individual preferences of multiple individuals, such information can only meaningfully be reported using ordinal comparisons ("do you prefer A or B?") rather than cardinal numbers ("Is A worthy of a score of 4.3?"). This understanding underlies the "universal domain criterion" that Kenneth Arrow adopted in evaluating voting methods in his ground-breaking (and Nobel Prize-winning) work on what has come to be known as Arrow's Impossibility Theorem.

Let's examine the nature of cardinal scores for a moment. Within a single brain it is conceivable that an exceptionally well-disciplined mind could assign a satisfaction, or "utility" score to all eventualities, using a consistent scale …a new car is worth 8.3, seeing one's child perform well in a school play is a 7.6, electing one's third favorite candidate for lieutenant governor is a 2.4, etc. However, for most people these scores would be, at best, extremely rough. Now if we limit the scoring to candidates for a particular office, we are at least seemingly comparing apples to apples, so the task is more manageable. However, even here, I might wonder how large of a gap should I leave between the score for my third favorite candidate, whom I agree with on one important policy issue, and my fourth favorite candidate who I respect as honest and sincere? Also, it seems apparent that most people do not naturally use a strictly linear number-line in

scoring. For example, the "distance" between giving a candidate a 4 and a 6 may be a lot easier to bridge than moving a candidate from an 8 to a 10. That is, a candidate with a 6 may be only 5% better than a candidate receiving a 4, whereas a candidate receiving a 10 might need to be more than 20% better than a candidate given an 8. And this is all within a single voter's mind. It seems unlikely that two voters giving one candidate a score of 5 would genuinely balance out another voter giving a different candidate a 10, in terms of satisfaction, even though these two candidates have the same average score among the three voters.

Voters are also likely to only have clear opinions about candidates they put at the two extremes of their scoring -- the ones they favor and the ones they hate -- with only a vague sense about some of the lesser-known candidates, which they may give more or less random scores in the middle. Even at the extremes of one voter's set of preferences, how will a voter settle on a scale to score candidates? One voter may compare all candidates to an idealized hypothetical candidate who would be in perfect synch on policy positions and of stellar personal character. Thus, no real-world candidate is deserving of a perfect 10 in most elections. Another voter might decide to invent a score based on the scale of candidates actually on the ballot in a particular race, with the first-choice candidate getting a 10 (even if the voter considers this candidate to be barely mediocre) and the last-choice candidate getting a zero, with the candidates in between getting in between scores. Other voters may decide 5 is a neutral mid point, and consider anything above 5 to be good, and anything below 5 to be an indication of disapproval. But other voters will understand the tactical advantage of overstating their preferences to maximize their chance of having an impact on the election results, and score some candidate(s) 10 and others 0, with not middling scores at all. Here is a key failing of Range Voting: different voters are using different scales, and adding their scores together is a essentially a meaningless mathematical exercise. A minority of voters who understand how to take tactical advantage (going to the extremes of 10 and 0) can overwhelm a majority who simply try to follow the voter instructions and give sincere scores.

Now let's see what happens when we start adding different voters' scoring lists together and average the candidate's scores. Voters Adam, Betty and Cathy happen to have identical opinions about three of the candidates in a race. Because there is no common scale, Adam scores them 1, 4, 7, Betty scores them 4, 7, 9, and Cathy scores them 2, 6, 10. The cardinal scores may be mushy, but the ordinal information that CAN be gleaned from such scores is how each voter ranks these candidates relative to each other. So, using the Range cardinal values, we now have a mathematically precise average of 8.667 for their most preferred candidate. But it is apparent that rather than having an accurate value that reveals the "will" of the voters (that can be calculated to ten decimal places if desired), we only have the illusion of accuracy, and in fact we have mud. Combine this average with the average gleaned from some other voters with different opinions about the candidates, and the mud gets deeper and murkier. If we add just two more voters with opposite political views, who have learned to vote strategically, they might score the candidates 10, 0, 0. The third candidate's average has now dropped to 5.2, whereas the first candidate's average has risen from 2.333 to 5.4 (the Range winner, who the majority oppose).

Most people will never be willing to abandon the principle of majority rule in favor of Bayesian regret (the core concept of Range Voting advocates). If 55% of voters prefer candidate A and 45% prefer candidate B in a two candidate race, Range Voting promotes the concept that perhaps B should win ***if*** the 45% feel very strongly about their choice, while the 55% are only lukewarm about their choice. This characteristic of Range Voting, the fact that it expressly rejects the notion of majority rule, means that it will never be adopted for government elections.