Quantification of the Effects of Voter Protocols on the Outcome of Approval Voting

Sarah Kunkler [‡*] Zhuorong Mao [‡†]

Project Advisors: Susana Furtado [‡§] Charles R. Johnson [¶‖]

Abstract

Approval Voting over several alternatives asks each voter to choose a subset of the alternatives of which they "approve". Then, the alternative (or, perhaps, alternatives) approved by the most voters is selected. The outcome is then not only a function of the profile of individual voter preferences alone, but also the protocols (number of alternatives approved) chosen by each voter. We quantify the differences in outcome that result from differences in protocols in several ways, regarding hypotheses about individual preferences. Considered are the two natural protocols when there are three alternatives, for both small and large numbers of voters. For more alternatives, we consider all possible pure protocols. Mixed protocols are also considered to quantify protocol effects. Several methods were employed for varying numbers of voters. Consistent numbers result from all methods. We find that differences in outcome, due to differences in protocol alone, can be quite frequent and often rival preferences as a determinant of the outcome.

Keywords: approval voting, profile, protocol

AMS Classi ication: 91B12, 91B14, 05A99

[*]Email: sarah.j.kunkler@gmail.com

[†]Email: zmao@wm.edu

[‡]CEAFEL and Faculdade de Economia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-464 Porto, Portugal.

[§]Email: sbf@fep.up.pt The work of this author was supported by FCT- Fundação para a Ciência e Tecnologia, under project UIDB/04721/2020.

[¶]Department of Mathematics, College of William and Mary, Williamsburg, VA 23187-8795.

[‖]Email: crjohn@wm.edu. The work of this author was supported in part by National Science Foundation grant DMS-0751964.

1 Introduction

In "Approval Voting" (AV) [2] by n voters among s alternatives, each voter is asked which subset of the s alternatives he approves. The winning alternative is then the one approved most often. Of course each voter may still be assumed to have an individual preference ordering (IPO), and it may reasonably be assumed that each voter only approves of alternatives sufficiently high in his/her IPO. But, since the "approval protocols" may well differ, the outcome of AV is not a function just of the profile of IPO's. Whatever virtues AV may be argued to have [4], this ambiguity resulting from the approval protocols is an important feature of AV and should be better understood. That there can be outcome differences resulting from differences in approval protocols was recognized in [9], where it was shown that there are profiles from which any outcome may result by adjusting each individual's approval protocol. This feature, troubling to the authors of [9], was viewed as a virtue by the authors of [3] in that it gives voters greater flexibility to register their cardinal preferences. Our purpose here is to quantify this ambiguity that we will refer to as the "protocol-caused ambiguity" or "ambiguity" for simplicity. Despite considerable new, and often quantitative, information about AV that may be found in the Handbook on Approval Voting [6], this phenomenon has not been well recognized nor quantified. Simulations of other aspects of AV can be found, for example, in [8]. Among other papers on AV, we note the recent study [1].

By the approval protocol, we mean the number of alternatives down a voter's IPO that the voter approves. Let protocol k, $1 \leq k < s$, mean that a voter approves the top k alternatives on his/her list. Differences in protocol may occur for many reasons, some of which include strategic voting, interpretation of intent, and cardinal differences in preference, as well as random factors. One measure of the ambiguity of AV is the frequency of outcome differences, depending only upon protocol differences. For example, we may ask how often there is a difference between all voters following protocol 1 and all voters following protocol 2. No value judgment need be made about the relative merits of such protocols, but since voters may adopt any protocol for any reason, the resulting outcome difference frequency is of considerable interest. Of course, there can be differences in such analysis depending upon the assumed underlying distribution of profiles. Nonetheless, the extent of this ambiguity appears high under a variety of assumptions. For example, under a uniform distribution on profiles hypothesis, assuming 3 alternatives and a large number of voters, the outcome differs between protocol 1 for all voters and protocol 2 for all voters about 46% of the time.

This analysis is done in Section 2.1 in which the frequency of the ambiguity is also investigated for a small number of voters. In Section 2.2, we also study $s > 3$ alternatives and the frequency of the ambiguity when all voters follow protocol k and all voters follow protocol l, for $k, l < s$ and $k \neq l$. While in Section 2 we considered pure protocols, that is, all voters follow the same protocol, in Section 3 we allow mixed protocols, in the sense that in the same election different voters may choose different protocols. We study the frequency of the ambiguity in some special situations.

Our numerical experiments were done using Python 3.8.3 and also confirmed using Octave 6.1.0 and R [10].

2 Differences between Pure Protocols

2.1 With 3 alternatives

Consider the case of $s = 3$ alternatives and a large number n of voters. There are 6 possible IPO's. Let a_{ij}, $i \neq j$, $1 \leq i, j \leq 3$, be the fraction of voters with the IPO: $i > j$ ($> k$, the third alternative). By definition, a_{12}, a_{13}, a_{21}, a_{23}, a_{31}, and a_{32} sum to 1. Since n is large, we may think of (a_{12}, a_{13}, a_{21}, a_{23}, a_{31}, a_{32}) as an arbitrary vector in the unit 6-simplex with six vertices each corresponding to the event in which all voters have the same one of the six orderings. If we assume a uniform distribution over this simplex, (i.e. assume impartial anonymous culture), consider the relative volume in the 6-simplex of the region in which alternative 1 wins under protocol 1 and does not win under protocol 2. Under the assumption of independence between alternatives, this volume, times 3, is the probability of a difference in outcomes under the two uniform protocols. Algebra gives that this region of the 6-simplex is defined by the two inequalities:

$$a_{12} + a_{13} > \max\{a_{21} + a_{23}, a_{31} + a_{32}\}$$

and

$$a_{23} + a_{32} > \min\{a_{13} + a_{31}, a_{12} + a_{21}\}.$$

Monte Carlo integration [5][7] of the relative volume of this region, to at least two decimal places of accuracy gives 0.154. Multiplication by 3 gives 0.462 as the probability of a difference between the two uniform protocols. (Closed form, analytical integration appears intractable.)

To confirm the simplex analysis when n is large, we used what we call the convex combination method, that is, we simulated, by generating in each trial a random vector of frequencies summing to one, in which the ith entry is the frequency of voters with the ith IPO. Such a vector was obtained by generating a random number from a uniform distribution on $(0, 1)$ for each IPO and then dividing it by their sum. Each simulation was based upon 10^5 trials and, for simplicity, those rare trials exhibiting a tie in either protocol were excluded. (Inclusion of ties would have increased the frequency of the ambiguity.) The obtained frequency of ambiguity was 0.46155. There was usually agreement among simulations to at least two decimal places, suggesting stability in the probability of an ambiguity among distributions of IPO's near to uniform.

To better understand the frequency of the ambiguity when n is small, we present in Table 1 the exact frequencies of an ambiguity using protocols 1 and 2, for $3 \leq n \leq 9$. For varying numbers of $n \geq 10$ voters, we also simulated, by generating random IPO's with equal probability and then voted using protocols 1 and 2. Each simulation was based upon 10^5 trials. The results obtained in our simulations are given in Table 2. In both Tables 1 and 2, we consider the case in which any trial exhibiting a tie in either protocol is excluded, as well as the case in which ties are considered. In this case, the following procedure was applied: i) if only one of the protocols gives a tie, in case the winner of the other protocol is one of the winners of the protocol giving a tie, we count a difference of 0.5 in both protocols, otherwise we count 1; ii) if both protocols give a tie but the results do not coincide for both protocols, we count 0.5 if there are two alternatives that win in both protocols and 0.75 otherwise. This is only an example of a procedure to count more when we intuitively think the difference is more substantial; other reasonable measures could be followed without changing the general claim.

Observe that, according to the described procedure, including ties increases the frequency of an ambiguity. As ties become more rare under higher n, this difference (in ambiguity frequency) between ties and no ties decreases. Our primary interest is in the case of no ties, though for completeness we include data with ties.

n	3	4	5	6	7	8	9
frequency of no ties	0.50	0.421	0.463	0.599	0.533	0.558	0.647
frequency of ambiguity (no ties)	0.333	0.352	0.350	0.389	0.376	0.380	0.403
frequency of ambiguity (with ties)	0.431	0.451	0.454	0.453	0.458	0.457	0.455

Table 1: Exact frequencies of a protocol-caused ambiguity using protocols 1 and 2, for 3 alternatives and n voters.

n	10	11	12	15	35	55	75	95	1000	2000
frequency of no ties	0.600	0.613	0.680	0.704	0.767	0.814	0.850	0.856	0.954	0.968
frequency of ambiguity (no ties)	0.393	0.396	0.410	0.419	0.427	0.436	0.443	0.441	0.458	0.463
frequency of ambiguity (with ties)	0.459	0.459	0.459	0.460	0.463	0.462	0.463	0.463	0.466	0.466

Table 2: Simulated frequencies of a protocol-caused ambiguity using protocols 1 and 2, for 3 alternatives and n voters (10^5 trials).

The data exhibit some number theoretic non-monotonicity, especially for smaller n, but eventually stabilize and gradually grow. The number attained from the simplex approach appears to be a limiting value. Thus, the protocol-caused ambiguity appears to be intrinsic to AV.

2.2 More than 3 alternatives

A convex combination method[1] similar to the one for $s = 3$ (see Section 2.1) was carried out to compute differences in outcomes among protocols, for $s = 4$ and larger.

For $s \in \{4, 5, 6, 7\}$, we used simulation to study the frequency of such differences in protocols k and l, for $k, l < s$ and $k \neq l$. We consider 10^5 trials and in each trial generate a random vector of frequencies summing to one, as above, in which the ith entry is the frequency of voters with the ith IPO. Those rare trials exhibiting a tie in some protocol were excluded. In Table 3 we give the frequencies of all $s - 1$ protocols equal.

[1] The code of the method can be accessed through this link: `https://github.com/zhuorongmao/approval_voting`.

s	4	5	6	7
frequency of all protocols giving the same outcome	0.270	0.139	0.074	0.042

Table 3: Simulated frequencies of all protocols equal, for s alternatives (10^5 trials with ties excluded).

Next we present the $(s-1)\times(s-1)$ matrices A_s in which the i,j entry, $i<j$, is the frequency in the performed simulations of a protocol-caused ambiguity when all voters follow protocol i and all voters follow protocol j, when $s\in\{4,5,6,7\}$ alternatives are considered.

$$A_4=\begin{bmatrix} 0 & 0.497 & 0.614 \\ 0 & 0 & 0.497 \\ 0 & 0 & 0 \end{bmatrix};\ A_5=\begin{bmatrix} 0 & 0.527 & 0.634 & 0.705 \\ 0 & 0 & 0.487 & 0.633 \\ 0 & 0 & 0 & 0.526 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_6=\begin{bmatrix} 0 & 0.545 & 0.653 & 0.712 & 0.764 \\ 0 & 0 & 0.493 & 0.624 & 0.715 \\ 0 & 0 & 0 & 0.492 & 0.654 \\ 0 & 0 & 0 & 0 & 0.545 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix};\ A_7=\begin{bmatrix} 0 & 0.564 & 0.666 & 0.726 & 0.765 & 0.801 \\ 0 & 0 & 0.498 & 0.622 & 0.700 & 0.765 \\ 0 & 0 & 0 & 0.482 & 0.620 & 0.724 \\ 0 & 0 & 0 & 0 & 0.496 & 0.665 \\ 0 & 0 & 0 & 0 & 0 & 0.559 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Observe that the matrices A_s are close to symmetric with respect to the anti-diagonal, that is, the frequency of an ambiguity when considering protocols i and j, $i\neq j$, is similar to the one when considering protocols $s-i$ and $s-j$. For example, for $s=7$, the difference between protocols 1 and 2 is similar to the difference between protocols 5 and 6. It may be that approving a certain number of alternatives and disapproving the same number of alternatives have some symmetry.

Also note that the frequency of an ambiguity between different protocols increases when the number of alternatives increases. From Table 3 we see that the number of all protocols giving the same winner drops when the number of alternatives increases.

3 Mixed Protocols

The previous "pure protocol" analysis is designed to identify and quantify the protocol-caused ambiguity by comparing scenarios in which IPO's are translated into votes in two different, well-defined, ways. In fact, unpredictably mixed protocols may occur in an actual vote. It may be asked if this mitigates the ambiguity. To this end, simulations were run in which each voter was endowed with a (random) IPO and then, for an initial vote, a random (possibly different) protocol was chosen for each voter, determining an AV outcome. For a second vote, protocols were again randomized for each voter, determining another AV outcome. This concluded a trial. Multiple hundred thousand trials simulations show that, for 3 alternatives and a large number of voters, considering protocols 1 and 2, the outcome was different in about 32% of the trials. It is not surprising that the percentage is smaller when compared with the one for pure protocols 1 and 2 (32% vs 46%), as there is less distinction among the chosen protocols. The results of this experiment, in which ties are excluded, are presented in Table 4. For the sake of comparison, we also included small numbers of n.

n	**3**	**4**	**10**	**15**	**35**	**55**	**75**	**95**	**1000**	**2000**
frequency of no ties	0.442	0.494	0.630	0.685	0.685	0.818	0.844	0.860	0.955	0.968
frequency of ambiguity	0.128	0.144	0.208	0.234	0.268	0.282	0.288	0.290	0.318	0.322

Table 4: Frequencies of protocol-caused ambiguity when protocols 1 or 2 for each voter are randomly chosen, for 3 alternatives and n voters (10^5 trials with ties excluded).

In addition, we might imagine that the chosen protocol depends somehow upon the IPO's. As an experiment, two other mixed protocols were compared to the pure protocol 1: (i) one in which just the voters who had the protocol-1-winning alternative first on their list remained with protocol 1, while all others used protocol 2 and (ii) vice-versa. Clearly, (i) should yield fewer differences in outcome, and likely the latter should produce more. Each simulation was based upon 10^5 trials. Ties were excluded from the number of trials and the test was done for $s = 3$ alternatives. The results are presented in Tables 5 and 6. We observe that the frequency of no ties increases for odd n and for even n but, in Table 6, from an odd n to the next even n decreases.

n	10	15	20	35	40	55	60	75	100	101	200
frequency of no ties	0.793	0.822	0.802	0.864	0.893	0.908	0.912	0.922	0.933	0.925	0.947
frequency of ambiguity	0.026	0.015	0.009	0.007	0.006	0.003	0.003	0.001	0.001	0.000	0.000

Table 5: Frequencies of a protocol-caused ambiguity with protocol 1 versus protocol (i) above for 3 alternatives (10^5 trials with ties excluded).

n	10	15	20	35	40	55	60	75	100	101	200
frequency of no ties	0.566	0.754	0.648	0.855	0.761	0.907	0.801	0.922	0.845	0.925	0.883
frequency of ambiguity	0.850	0.940	0.946	0.988	0.990	0.997	0.998	0.999	1.000	1.000	1.000

Table 6: Frequencies of a protocol-caused ambiguity with protocol 1 versus protocol (ii) above for 3 alternatives (10^5 trials with ties excluded).

4 Conclusions

Here, we have studied the fact that the winning alternative in an approval vote depends not only on each voter's preference ordering but also on how far down the preferences the voter approves of (the voter's protocol). Under several assumptions, we have quantified the difference in the winning alternative under different protocols. Our numerical experiments indicate that, in case different pure protocols are used, the frequency of an ambiguity increases with the number of voters and stabilizes in a certain limiting value. Moreover, this frequency also increases when the number of alternatives increases. The inclusion of ties also increases the frequency of an ambiguity, although ties become less frequent when the number of voters increases.

We have also investigated some situations in which mixed protocols are used, that is, not all voters in an election use the same protocol. In this case, the frequency of an ambiguity may increase or not with the number of voters, depending on the specificities of the protocols used.

One fact emerges: the "protocol-caused ambiguity" we identify is substantial and rivals preferences in determining the outcome of an approval vote. This is very important and is our key conclusion.

References

[1] Ben-Haim, Y. *Approval and plurality voting with uncertainty: Info-gap analysis of robustness.* Public Choice 189 (2021), 239–256.

[2] Brams, Steven J. and Fishburn, Peter C. *Approval Voting.* New York: Springer Science+Business Media, 1983.

[3] Brams, Steven J., Fishburn, Peter C., and Merrill, S. *The Responsiveness of Approval Voting: Comments on Saari and van Newenhizen.* Public Choice 59 (1988), 121-131.

[4] Brams, Steven J. *Mathematics and Democracy.* Princeton: Princeton University Press, 2008.

[5] Hammersley, J. M.; Handscomb, D. C. *Monte Carlo Methods.* Methuen (1964). ISBN 978-0-416-52340-9.

[6] Laslier, Jean-François and Sanver, M. Remzi, editors. *Handbook on Approval voting.* Springer New York, 2010.

[7] Lepage, G. P. *A New Algorithm for Adaptive Multidimensional Integration.* Journal of Computational Physics (1978). 27 (2): 192–203. doi:10.1016/0021-9991(78)90004-9.

[8] Mercik, Jacek W. *Computer Simulations of Approval Voting: Positional in One Dimension Candidates' Characteristics*, European Journal of Political Economy 4, No. 4 (1988), 517–527.

[9] Saari, Donald G. and van Newenhizen, Jill. *The Problem of Indeterminacy in Approval, Multiple and Truncated Voting Systems.* Public Choice 59 (1988), 101-120.

[10] The R Project for Statistical Computing. 9 December 2008 <http://www.r-project.org>.