Voting schemes based on candidate-orderings or discrete choices considered harmful

Warren D. Smith
WDSmith@fastmail.fm

December 2001

Abstract —
We give a family of scenarios in which *any* voting system for single-winner elections in which each voter's vote is based purely on his perceived *ordering* of the 3 candidates, is entirely helpless – i.e. must regard the election as a 3-way tie – hence necessarily will elect a random candidate. By slight tie-breaking perturbations we could instead make it elect any particular candidate, e.g. the worst one. These scenarios are based on a new construction of "intransitive dice" having independent interest. We further argue that *any* voting system involving *discrete* vote choices must often elect bad candidates in these scenarios. Meanwhile, voting systems in which votes are permitted to continuously vary, need not regard the election as tied nor experience any difficulty. A specific scenario is given that is a "universal counterexample" causing every voting system so far proposed (except for "honest utility voting") to misbehave simultaneously.

Keywords — Range voting, nontransitive dice, universal counterexample.

1 Nontransitive dice

Let $x_1 < x_2 < x_3 < x_4 < x_5$ be reals. Let $g = (\sqrt{5}-1)/2 \approx 0.6180339887$ be the "golden" constant. Consider[1] the following 3 probability distributions A,B,C.

A**:** $\text{prob}(x_3) = 1$;

B**:** $\text{prob}(x_4) = g, \quad \text{prob}(x_1) = 1-g$;

C**:** $\text{prob}(x_5) = 1-g, \quad \text{prob}(x_2) = g$.

Notice these 3 probability distributions are "nontransitive." That is, if one extracts a sample S_A from A (and similarly S_B from B and S_C from C) then

$$\text{prob}(S_B > S_A) = g, \quad \text{prob}(S_C > S_B) = g(1-g) + 1 - g = g, \quad \text{prob}(S_A > S_C) = g. \tag{1}$$

Stephen Omohundro and I invented these three probability distributions in about 1995 and I proved that g was the maximum possible minimum-intransitivity for any three probability distributions. However, we soon realized we were not the first discoverers. Indeed, nontransitive dice seem to keep getting rediscovered [2][3]. Trybula [6][4] proved in 1965 that a cycle of N intransitive dice exists for any $N \geq 3$, and determined the maximum minimum-intransitivity value ι_N for each N. $\iota_3 = g \approx 0.618$, $\iota_4 = 2/3 \approx 0.667$, and $\iota_N \to 3/4 = 0.75$ as $N \to \infty$.

Bradley Efron noted that one can build a maximally-intransitive set of 4 ordinary cubical fair dice A,B,C,D by labeling their 6 faces as follows:

$$A : 004444, \quad B : 333333, \quad C : 222266, \quad D : 111555 \tag{2}$$

(or any real numbers having the same order-relations as 0,1,2,3,4,5,6 may be used instead; I recommend mysterious-sounding values such as π, e, $\sqrt{40}$ if you plan to build dice for use in bar bets). They obey $A \succ B \succ C \succ D \succ A$ where "$A \succ B$" means "A will produce a larger roll than B with probability 2/3."

[1] By "$\text{prob}(x_3) = 1$" we mean $\text{prob}(x = x_3) = 1$; i.e. A is a probability distribution on a real variable x which consists of a unit-mass "spike" located at $x = x_3$. Similarly B consists of two spikes, one of mass g and the other of mass $1-g$, located at x_4 and x_5 respectively.

2 A new kind of nontransitive dice

This paper will rest on the following new construction of 3 "nontransitive dice" A, B, C.

Let $x_0 < x_1 < x_2 < x_3 < x_4 < x_5 < x_6$ be 7 real numbers. Let $z \approx 0.6421298561$ be[2] a root of $3z^4 - 3z^3 + 2z - 1$.

Let A,B,C be the following 3 probability distributions:

A**:** $\text{prob}(x_0) = p_0$, $\text{prob}(x_3) = p_3$, $\text{prob}(x_6) = p_6$;

B**:** $\text{prob}(x_1) = p_1$, $\text{prob}(x_4) = p_4$;

C**:** $\text{prob}(x_2) = p_2$, $\text{prob}(x_5) = p_5$;

where

$$p_3 = 4z^3 - 2z^2 - 2z + \frac{5}{3} \approx 0.61682500, \quad p_2 = p_4 = z \approx 0.64212986,$$
$$p_1 = p_5 = 1 - z \approx 0.35787014, \quad p_6 = p_0 = \frac{1 - p_3}{2} \approx 0.19158750. \tag{3}$$

Let $p_- \equiv p_0 p_2 p_4$ and $p_+ \equiv p_2 p_3 p_4$. These 3 distributions obey $A \prec B \prec C \prec A$ where "$A \prec B$" means "A will produce a smaller sample than B with probability $p = p_- + 2p_+ \approx 0.58766925$." That by itself is somewhat less impressive than the maximally intransitive dice-triplet (given before), which has $p = g \approx 0.61803399$. What *is* impressive, is that not only are these dice beautifully symmetrically intransitive when just considering *pairs* of samples, but they also are beautifully symmetrically intransitive when considering *triples* of samples from all three distributions:

Theorem 1 (Permutation probability equalities for these 3 dice) *Let S_A be a sample from A, S_B from B, and S_C from C where A,B,C are the probability distributions above. Then*

$$\textit{prob}(S_A > S_B > S_C) = \textit{prob}(S_B > S_C > S_A) = \textit{prob}(S_C > S_A > S_B) = p_- \approx 0.07899742 \tag{4}$$

$$\textit{prob}(S_A < S_B < S_C) = \textit{prob}(S_B < S_C < S_A) = \textit{prob}(S_C < S_A < S_B) = p_+ \approx 0.25433591 \tag{5}$$

Proof. Verify the following equalities by performing exact arithmetic on algebraic numbers. (One could also just use 10-decimal floating point arithmetic for a convincing non-proof.)

event	its probability
$S_A > S_B > S_C$	$p_6 p_4 p_2 = p_-$
$S_B > S_C > S_A$	$p_4 p_2 p_0 = p_-$
$S_C > S_A > S_B$	$p_1 p_3 p_5 = p_-$
$S_A < S_B < S_C$	$p_5 p_4 (p_3 + p_0) + p_5 p_1 p_0 + p_2 p_1 p_0 = p_+$
$S_B < S_C < S_A$	$p_5 p_6 + p_2 (p_3 + p_6) p_1 = p_+$
$S_C < S_A < S_B$	$p_2 p_3 p_4 = p_+$

Q.E.D.

3 A particular model of voter beliefs

Associate 3 candidates A,B,C with (to abuse notation) the respective probability distributions A,B,C of §2.

Suppose that there are 7 actions the candidates could take after they are elected, having respective utilities (known to and agreed on by all) $x_0, x_1, x_2, x_3, x_4, x_5, x_6$. As usual nowadays, the candidates refuse to answer direct questions about which of the 7 actions they actually will take, but they do try vaguely to suggest different biases in their stances in different places at different times. Consequently, a fraction p_3 of the voters become convinced that candidate A will do action 3, a fraction p_6 of the voters become convinced A will do action 6, and the remaining voters think he will do action 0, which is exactly the probability distribution "A" from §2. Similarly, we'll suppose the perceived probabilities correspond to the probability distributions A, B, C.

In other words, each voter independently does the following mental exercise, to determine his perceived utility of electing each candidate:

1. Extract (independent) samples S_A, S_B, S_C, from probability distributions A, B, C respectively.
2. Regard candidate A's utility as S_A, candidate B's as S_B, and candidate C's as S_C.

[2] This may be expressed in closed form with the aid of the quartic formula: Let $R = (12 + 4\sqrt{41})^{1/6}$ and $Q = \sqrt{2R^4 + 3R^2 - 16}$. Then $12z = 3 - QR^{-1}\sqrt{3} + \sqrt{6}\sqrt{3 - R^2 + 8R^{-2} + 13\sqrt{3}R/Q}$.

4 What will happen in various voting systems

To keep things simple, we will assume there are a *large* number V of voters (take the limit $V \to \infty$), all of the type in §3.

Theorem 2 (Voting systems based on candidate-orderings) *With the voters described in §3,* any *voting system in which each voter provides (and only provides) a vote that is a function of his perceived* ordering *of the 3 candidates (and we assume the "winner" output by this voting system is a function only of the* votes *and not of the* names *of the 3 candidates, i.e. behaves invariantly under a simultaneous permutation of the candidates and votes, i.e. does not know which candidate is named "A") will be incapable of distinguishing between A, B, and C, i.e. must elect a random winner.*

This is a trivial corollary of theorem 1.

But in general there *is* reason to prefer one of A, B, or C. Specifically, we (as Bayesians) want the candidate with the maximum expected utility. For the probability distributions of §2, the expectation values are

$$A: p_0x_0 + p_3x_3 + p_6x_6, \quad B: p_1x_1 + p_4x_4, \quad C: p_2x_2 + p_5x_5. \tag{6}$$

Thus for any particular choice of (and depending on this choice) $\vec{x} = (x_0, x_1, x_2, x_3, x_4, x_5, x_6)$, there generically is – from the Bayesian point of view – a clear winner: the candidate with the greatest expectation value in EQ 6.

Theorem 3 (Any Bayesian ordering is possible) *By choosing $(x_0, x_1, x_2, x_3, x_4, x_5, x_6) \in \mathbf{R}^7$ appropriately with $x_0 < x_1 < x_2 < x_3 < x_4 < x_5 < x_6$, any of the $6 = 3!$ orderings of the 3 candidate expectation values is achieveable.*

Proof. Consider tiny perturbations of the point

$$\vec{x}^* = (-1, \ -p_4, \ -p_5, \ 0, \ 1-p_4, \ 1-p_5, \ 1) \tag{7}$$

at which all 3 expectation values are 0. By perturbing x_3, x_4, and x_5 slightly we will perturb the expectation values of A, B, C by proportionate amounts. Q.E.D.

Thus, we have a contrast. *Every* voting system based purely on candidate-orderings is utterly incapable of distinguishing between these 3 candidates. That includes such voting systems as "Borda Count," Condorcet, Copeland, Dabagh, Plurality, Bullet, Tideman's "Ranked Pairs," Black's system, Hare's "single transferable vote" (also called "instant runoff voting"), Weighted positional systems, Coombs's "single transferable vote," Carey's system, etc. (see [5] for descriptions of these and other voting systems).

But on the other hand, there generically is a clear preference ordering (based on expected utility summed over all voters) for the 3 candidates. And that ordering *can* be determined by a voting system *not* based on candidate orderings:

Theorem 4 (Honest utility voting works) *Suppose there are V voters, $V \to \infty$, and each voter supplies* as his vote*, in a c-candidate election (c fixed) precisely his c-tuple of his perceived utilities of the c candidates. Also suppose each voter obtains his c perceived candidate-utilities by sampling from c respective probability distributions (each of which, we assume, has an expectation value). Suppose the voting system is: These tuples are summed and a candidate with the maximum sum is declared the winner.*

Then: This system will, with probability$\to 1$, elect a candidate with maximum expected utility.

Proof. This is simply the "law of large numbers" from probability theory. Q.E.D.

First Conclusion: Voting systems – in particular "honest utility voting" – allowing continuously variable votes are capable of handling[3] situations which voting systems based purely on candidate orderings can't.

We can go further. Instead of $\vec{x}^*$ from EQ 7, we could in fact use any member of the 4-dimensional polytopal open subset P of $\mathbf{R}^7$ defined by the inequalities $x_0 < x_1 < x_2 < x_3 < x_4 < x_5 < x_7$ and the equalities $x_0 + x_6 = 2x_3$, $x_2p_2 + x_5p_5 = x_3$, and $x_1p_1 + x_4p_4 = x_3$. Any such $\vec{x}$ yields a 3-way tie for "best candidate" in terms of expected utility. By perturbing any such tie by an arbitrarily small amount in 7-space, we can break the tie according to any desired ordering of the 3 candidates. A voter (or anybody), purely by examining a single sample from A, B, and C (since it consists, in total, of only 3 real numbers) cannot determine which $\vec{x}$ in this 4-dimensional continuum, we are at, nor (even after the tie-breaking perturbation) can he tell which candidate is best nor even decide which of any two of them is better. Therefore, if the voters employ *any* deterministic algorithm to determine their vote from (S_A, S_B, S_C), and if that vote is required to be in a *discrete* set, then there must be choices of $\vec{x}$ causing that voting scheme to deliver (with probability 1 as $V \to \infty$) the wrong winner. Proof: suppose not. In that case we can by an

[3] "Handling" means "not being utterly unable to distinguish the candidates in any way" and "being able to determine a clear winner."

arbitrarily small perturbation of $\vec{x}$ change the best candidate, but discrete voting schemes cannot detect infinitesimal perturbations unless $\vec{x}$ is exactly on a decision boundary. But with any finite number V of voters, no matter how large, decision boundaries must constitute measure-0 of P, since any finite or countable union of measure-0 sets has measure zero, and our previous remarks show that each particular voter has decision boundaries that have measure 0 in P. Q.E.D. Therefore:

Second Conclusion: Voting systems – in particular "honest utility voting" – allowing continuously variable votes are capable of handling situations which cannot be handled by voting systems allowing only votes in a discrete set.

5 Range voting, Bayesian regret, and universal counterexamples

A wide variety of proposed voting systems for c-candidate, single-winner elections were surveyed in the author's comprehensive paper "Range Voting" [5]. That paper, after surveying both theoretical properties and extensive computer simulation studies for both honest, strategic, and ignorant voters (all in many scenarios), came to the clear conclusion that "Range Voting" was superior to all the other voting systems surveyed.

5.1 Range voting

Definition. Range voting is this: Each voter supplies a c-tuple of real numbers (each in the interval $[0, 1]$, or equivalently any other finite interval could be used) as his vote. The tuples are summed. A candidate corresponding to a maximum entry in the summed tuple is declared the winner.

Range voting is also the only voting system I know of (among those proposed so far) which allows continuously variable – as opposed to discrete – votes. It was also shown in [5] that K.J.Arrow's famous "impossibility theorem" [1] could, to some extent, be evaded by range voting (and especially by honest utility voting). That argument was partly based on the fact that Arrow had *assumed* in his definition of "voting system" that votes could only be discrete and based on candidate-orderings – restrictions that here are argued to be silly and harmful.

In view of this, it might seem that the present paper is highly supportive of range voting. That is true, but unfortunately the situation is not so simple. We will soon construct choices of $\vec{x} \in \mathbf{R}^7$ in our scenario of §3, that cause range voting actually to be *worse* than ordering-based voting methods (i.e., worse than picking a winner randomly). The reason that range voting is nevertheless superior to ordering-based voting methods is that such $\vec{x}$ are rare.

Here are 3 kinds of range (or range-like) voting:

U: Honest (*u*nscaled) *u*tility voting. This will elect the candidate with maximum expected summed-utility, by theorem 4.

S: "Range voting with honest voters," also called (in [5]) "*s*caled utility voting." Here each voter gives the candidate with maximum perceived utility a vote of 1, the one with minimum perceived utility gets 0, and the remaining candidate gets a vote linearly interpolated (according to his perceived utility) between 0 and 1.

A: In range voting (or "*a*pproval voting") with strategic voters, each voter gives the candidate with maximum perceived utility a vote of 1, the one with minimum perceived utility gets 0, and the remaining candidate gets 0 if his perceived utility is below the average of the best and worst candidates's utilities, and otherwise 1.

One may verify (by considering the 12 possible values of the tuple (S_A, S_B, S_C), their probabilities, and what vote consequently happens in each case) that, in the limit $V \to \infty$ of a large number of voters, with probability 1, these 3 voting systems, applied to the scenario of §3, will elect a candidate corresponding to a maximum entry in the following respective 3-tuples (one for each voting system; I have rounded various coefficients to 8 decimals):

$$\mathbf{U}: \; (0.19158750x_6+0.61682500x_3+0.19158750x_0,\; 0.64212986x_4+0.35787014x_1,\; 0.35787014x_5+0.642129856x_2) \quad (8)$$

$$\begin{aligned}\mathbf{S},\mathbf{A}: \Big(&0.25433592\,F(\frac{x_3-x_2}{x_4-x_2}) + 0.07899742\,F(\frac{x_3-x_1}{x_5-x_1}),\; 0.07899742\,F(\frac{x_4-x_2}{x_6-x_2}) + 0.14174583\,F(\frac{x_4-x_3}{x_5-x_3})\\ &+0.04402664\,F(\frac{x_4-x_0}{x_5-x_0}) + 0.02453681\,F(\frac{x_1-x_0}{x_5-x_0}) + 0.04402664\,F(\frac{x_1-x_0}{x_2-x_0}),\; 0.04402664\,F(\frac{x_5-x_4}{x_6-x_4}) \quad (9)\\ &+0.02453681\,F(\frac{x_5-x_1}{x_6-x_1}) + 0.04402664\,F(\frac{x_2-x_1}{x_6-x_1}) + 0.14174583\,F(\frac{x_2-x_1}{x_3-x_1}) + 0.07899742\,F(\frac{x_2-x_0}{x_4-x_0})\Big)\end{aligned}$$

where, in EQ 9, for **S**, use $F(x) \equiv x$ while for **A** use

$$F(x) \equiv \begin{cases} 0 & \text{if } x < 1/2 \\ 1 & \text{if } x > 1/2 \end{cases} . \tag{10}$$

The fact that all 3 of these rangevote procedures are capable of giving *different* results is revealed by considering

$$\vec{x} = (-154,\ -42,\ -34,\ 10,\ 48,\ 112,\ 122) \tag{11}$$

which causes

U: C wins, since the tuple in EQ 8 is $(0.037,\ 15.79,\ 18.25)$

S: B wins, since the tuple in EQ 9 is $(0.1631,\ 0.1792, 0.1320)$

A: A (i.e., the worst candidate in terms of summed-expected-utility) wins; tuple$\approx$ $(0.2543,\ 0.1671,\ 0.1476)$.

It is also possible for **S** to elect the *worst* candidate in terms of summed-expected-utility, as is shown by the example

$$\vec{x} = (-130,\ -25,\ 15,\ 16,\ 40,\ 42,\ 270). \tag{12}$$

in which *both* **A** and **S** elect the worst candidate B instead of the best (A).

5.2 Bayesian regret

The reader of [5] will recall that, in computer studies of artificial elections with numerous utility-generating models, all models tried led to the conclusion that Range Voting was the best voting system (as measured by "Bayesian regret") among all voting systems so far proposed.

Definition. The "Bayesian regret" [5] of a single-winner voting system (in the presence of some randomized method by which each voter is assumed to acquire perceived utilities of election for each candidate) is the expected difference in utility between the winner in that voting system, and the hypothetical utopian (maximum expected-utility) winner. (Here "utility" is summed over all voters, of course. Voting systems with smaller regret are better.)

Those computer studies might now lead a optimist to speculate that *every* reasonable sounding randomized method for generating voters's perceived utilities of candidates, would cause Range Voting to yield smaller Bayesian Regret than all other so-far-proposed voting systems. But the example of EQ 12 is a counterexample to that speculation because the worst candidate has lower expected utility than a random candidate. I.e.:

Theorem 5 (Range voting can also be bad) *There exist reasonable sounding randomized methods (in particular, our method of §3 with the parameter choice $\vec{x} \in \mathbf{R}^7$ in EQ 12) for generating voters's perceived utilities of candidates, which cause both* **S** *and* **A** *range voting to be* worse *(i.e. yield greater Bayesian regret) than picking a random winner or (equivalently, in this example) using any ordering-based voting scheme.*

5.3 Universal counterexamples, and the fact that range voting is usually superior

Indeed, it is interesting that, for the parameter choice EQ 12 in the above scenario, *every voting system so far proposed* (at least, all the ones surveyed in [5]) other than, (of course) honest utility voting **U**, *fails* to deliver the candidate A that is uniquely best in a Bayesian sense. So the scenario of §3 instantiated by EQ 12 is, in some sense, a *universal counterexample.*

Fortunately, the two particular choices of $\vec{x}$ in EQ 11, 12 seem atypically nasty. If, in the scenario of §3, we pick $\vec{x}$ to be the result of sorting a 7-tuple of random numbers, then **S** and **A** both *usually* do substantially *better* than picking a random winner (**R**) or equivalently using the winner given by any ordering-based voting scheme (**O**), as the following table of Bayesian Regret values (found by doing 10^7 Monte-Carlo experiments) shows:

$\vec{x}$ \ voting method	**U**	**A**	**S**	**R=O**
Uniform[0,100] distrib.	0	0.973	0.714	4.45
Normal(0,1) distribution	0	0.0383	0.0299	0.142

6 Moral

Voting systems, such as range voting, in which the voters are allowed to *continuously* vary their votes, should be preferred a priori over systems in which the allowed votes lie in a discrete set.

References

[1] Peter C. Fishburn: Arrow's Impossibility Theorem: Concise Proof and Infinite Voters, J. Economic Theory 2,1 (March 1970) 103-106.

[2] Mathematical games column of Scientific American 223 (Dec 1970) 110-114; 231 (Oct 1974) 120-125; and 277 (Nov 1997) 110-113. The first two are reprinted in Martin Gardner: Nontransitive paradoxes, in *Time Travel and Other Mathematical Bewilderments*, W.H. Freeman 1987; Nontransitive dice and other probability paradoxes, pp. 44-50 in *Wheels, Life, and Other Mathematical Amusements*, W.H. Freeman 1983.

[3] Ivars Peterson's "Math Trek" column, MAA Online, 22 December 1997 and 6 Oct 1997. (Describes yet another recent rediscovery of nontransitive dice by Nathaniel Hellerstein and Bradley Efron.)

[4] R.P.Savage: The paradox of the nontransitive dice, Amer. Math. Monthly 101,5 (1994) 429-436.

[5] W.D.Smith: Range voting. Available electronically on `http://math.temple.edu/~wds/homepage/works.html` as #56.

[6] S.Trybula: On the paradox of N real variables, Zastos. Matem. 8 (1965) 143-154.