THE PROFILE STRUCTURE FOR LUCE'S CHOICE AXIOM

DONALD G. SAARI

ABSTRACT. A geometric approach is developed to explain several phenomena that arise with Luce's choice axiom such as where differences occur when a ranking is determined from best, or worst, first. New ways to compute probabilities and to increase the applicability of the choice axiom are introduced; e.g., with ten alternatives, Luce's formulation allows nine degrees of freedom for profiles while the approach described here allows millions of degrees of freedom. A profile decomposition identifies and explains the effects of using different methods to determine outcomes and compute ranking probabilities. Three and four alternatives are emphasized for reasons of exposition, but most of the conclusions hold for any number of alternatives.

1. INTRODUCTION

With the introduction of different axiomatic systems, one by Arrow (1963) for group processes and another by Luce (1959, 1961, 1962, 1977) for individual decisions, the 1950s proved to be a seminal period for decision and choice theory. These approaches differ in fundamental ways: Arrow emphasized the properties of the decision rules while Luce imposed a relationship on the outcomes. While Luce emphasizes a particular decision rule and its associated profiles, I adopt a different perspective by treating Luce's relationships as providing a way to implicitly define all admissible decision rules and their associated profiles. An advantage of this approach is that it endows Luce's choice axiom with more flexibility and potential applicability.

To motivate both Arrow's and Luce's systems, suppose 78 voters rank the four alternatives A, B, C, D as follows where "$\succ$" means "is strictly preferred to."

Number	Ranking	Number	Ranking
5	$A \succ B \succ C \succ D$	9	$B \succ D \succ A \succ C$
7	$A \succ C \succ B \succ D$	8	$C \succ B \succ A \succ D$
9	$A \succ D \succ B \succ C$	11	$C \succ D \succ A \succ B$
4	$B \succ A \succ C \succ D$	8	$D \succ B \succ A \succ C$
7	$B \succ C \succ A \succ D$	10	$D \succ C \succ A \succ B$

(1.1)

The plurality election outcome (where each voter votes for one alternative) is $A \succ B \succ C \succ D$ with the 21:20:19:18 tally. Dropping *any* alternative flips the election outcome

This research was supported by NSF grant DMI 0233798. My thanks to D. Luce for introducing me to this topic and for his several useful suggestions and comments about the earlier drafts of this paper. Also, my thanks L. Narens for several conversations about this subject, and to a referee who suggested the reorganization to start the paper, and place more emphasis on recent results in "voting theory." .

to agree with the reversed $D \succ C \succ B \succ A$. (For instance, by dropping D, the new $C \succ B \succ A$ ranking has a 29:28:21 tally.) But dropping *any two* alternatives flips the pairwise majority vote rankings to agree with the original $A \succ B \succ C \succ D$ outcome.

To convert Eq. 1.1 into a individual decision problem, normalize (divide each number by 78) to create a table that describes the probabilities a subject assigns to each ranking. This simple connection between aspects of voting theory and individual decision suggests that we can use recent results from voting theory (Sect. 2) to uncover new conclusions about individual decision making. This is a theme of this paper. For the rest of this introductory section, issues are raised, notation is given, and an outline for the paper is given.

1.1. **Some issues.** Although the Eq. 1.1 example had nothing to do with Arrow's and Luce's initial motivation, it is useful to view their contributions as exploring how to avoid these kinds of problems by finding ways to ensure consistent outcomes. But "consistency" may require imposing restrictions on preferences: a problem with restrictions is that they limit the applicability of any theory. A natural goal, then, is to find sufficiently lenient conditions that include interesting, useful settings. I identify alternative approaches that significantly relax the restrictions.

To be more specific about Arrow's approach, his "binary independence" condition (or "IIA") can be thought of as constructing a societal ranking by using only the information about how each voter ranks each pair of alternatives. Arrow proved that with binary independence, the only way to assure transitive societal rankings is to use a dictatorship for the decision rule. The reason, as we now know (Saari 1998, 2001b), is that a decision rule satisfying Arrow's assumptions dismisses the explicitly specified information that the voters have transitive preferences. As described later, this "loss of transitivity" phenomenon is caused only by certain configurations of preferences. It is interesting how the data component that causes Arrow's problem helps Luce's system (Sect. 6).

If binary independence dismisses the crucial transitive preference assumption, then to sidestep the difficulties Arrow identified, we could relax his binary independence condition to allow a decision rule to use the information about the transitivity of voters' preferences. By doing so, Arrow's dictator is replaced by several rules including the *Borda Count:* a decision rule that plays an interesting role in voting and in our discussion of Luce's axiom (Sects. 3-7).

Definition 1. *For n-alternatives, the* Borda Count *tallies each voter's ballot by assigning $(n-j)$ points to the voter's jth ranked alternative, $j = 1, \dots, n$. The alternatives are ranked according to the sum of received votes.*

1.2. **Luce's approach.** With his interest in individual decision making, Luce considers settings that can be captured with probability measures. For notation, letters in the early part of the alphabet such as $A, \dots, D$, correspond to alternatives, X, Y, Z are variables, R, S, T represent sets, and lower case letters represent probabilities; e.g., $P(A) = a$. The probability measure P_T satisfies the usual axioms for space T; i.e.,

(1) For $S \subset T$, $0 \le P_T(S) \le 1$.
(2) $P_T(\emptyset) = 0$, $P_T(T) = 1$.
(3) If $R, S \subset T$ and $R \cap S = \emptyset$, then $P_T(R \cup S) = P_T(R) + P_T(S)$.

For $S \subset T$, P_S denotes the conditional probability measure defined over set S and $P(A, B)$ represents $P_{\{A,B\}}(A)$: the probability of selecting A in $S = \{A, B\}$.

Luce avoids the Table 1.1 type of inconsistencies by *requiring* consistent outcomes where the probabilities in any set T determine all consistent subset outcomes. He accomplishes this feat by imposing the following axiom that is called the "choice axiom" in what follows.

Axiom 1. *Luce's Choice Axiom (Luce, 1959). For any $n \ge 2$, let $T = \{A_1, A_2, \dots, A_n\}$ be a set of n alternatives. A probability measure P_T satisfies the choice axiom if the following are true.*

(1) *For every non-empty $S \subset T$, P_S is defined.*
(2) *If $P(A_i, A_j) \ne 0, 1$ for all $A_i, A_j \in T$, then for $R \subset S \subset T$*

$$P_T(R) = P_S(R) P_T(S); \tag{1.2}$$

(3) *If $P(A_i, A_j) = 0$ for some $A_i, A_j \in T$, then for every $S \subset T$*

$$P_T(S) = P_{T-\{A_i\}}(S - \{A_i\}). \tag{1.3}$$

Because $P(A, B) = 0$ implies that $P(B, A) = 1$, part 3 describes an individual with perfect discrimination between alternatives A_i and A_j. Luce (1959, p. 8) imposed this condition to avoid questionable behavior that can be associated with perfect discrimination. While I also emphasize imperfect discrimination (namely, for all $A, B \in T$ assume that $P(A, B) \ne 0, 1$), Condition 3 is briefly mentioned.

With imperfect discrimination, conditional probability is defined as

$$P_T(R|S) = \frac{P_T(R \cap S)}{P_T(S)}. \tag{1.4}$$

Luce noted for $R \subset S \subset T$ that with $P_S(R) = P_T(R|S)$, his axiom is equivalent to the standard independence relationship. A subtle but important distinction is that because Luce does *not* require an universal set, his setting holds for those realistic situations where the set of alternatives changes. Namely, alternatives can be added to the system without needing to renormalize to ensure that the sum of the probabilities equals unity. A central reason this property holds is that, as described below, Luce's axiom has the effect of endowing each alternative with an intrinsic level of likelihood that is independent of the particular set. Thus, while P_T denotes a probability defined over set T, this notation does *not* imply that T is the universal set.

1.3. **Rankings.** I consider only settings where a subject ranks alternatives. Following Luce (1959, pp. 68-74), for the set of alternatives T, let "$P_T(A)$ denote the probability that A is judged to be the superior element in T according to some specified criterion." The reason for the "some specified criterion" phrase is that, as Luce states and as I show (Sects. 3, 7), selecting a criterion is equivalent to adding conditions to the choice axiom. Independent of what criterion is selected, the choice axiom and standard probability conditions force the following structure. (Luce uses a more general representation.)

Proposition 1. *(Luce) For $S \subset T$ and $A_i \in S$,*

$$P_S(A_i) = \frac{P_T(A_i)}{\sum_{A_j \in S} P_T(A_j)}. \tag{1.5}$$

As Prop. 1 demonstrates, Luce's conditions impose an orderly structure on the outcomes where the $\{P_T(A_j)\}_{j=1}^n$ values determine all remaining $P_S(A_j)$ probabilities. Notice how Eqs. 1.4, 1.5 provide the intended sense that *attached to each alternative A_j is an intrinsic $P_T(A_j)$ level of likelihood.* This property suggests that P_T can be constructed from the pairwise probabilities: this is illustrated for $T = \{A, B, C\}$.

Proposition 2. *If X, Y, Z represent the three alternatives of $T = \{A, B, C\}$, then, under the assumption of imperfect discrimination,*

$$P_T(X) = \frac{P(X,Y)P(X,Z)}{P(X,Y) + P(X,Z) - P(X,Y)P(X,Z)} \tag{1.6}$$

and

$$P(B,C) = \frac{P(A,C)P(B,A)}{P(A,B)P(C,A) + P(B,A)P(A,C)}. \tag{1.7}$$

This result, which may be known, asserts that the probability of selecting X from T is determined by how this alternative fares in both pairwise comparisons. A related relationship holds for any $n \geq 4$.

Proof. Let $P_T(A) = a$, $P_T(B) = b$, $P_T(C) = c$. Just substitute the $P(X,Y) = \frac{x}{x+y}$ values into these expressions and collect terms. $\square$

Notice how the choice axiom specifies the relationship among outcomes without mentioning the decision rules. This suggests exploring whether the choice axiom can be treated as implicitly defining a class of admissible decision rules (i.e., probability computations) and their associated profile restrictions. This is a main theme of the paper. First a "profile" must be defined.

Definition 2. *For $n \geq 3$ alternatives, there are $n!$ strict (complete, linear) transitive rankings. A voting profile lists the number of voters who have each ranking as a personal preference. A probability profile $\mathbf{p}$ lists the probability that a particular ranking occurs. That is, each component of $\mathbf{p}$ is non-negative and the sum of the components is unity.*

With a voting profile, ballots are tallied. With a probability profile, the various ranking probabilities are determined by the summation rules of probability. To illustrate, when Table 1.1 is converted into a probability profile (by dividing each number by the total number of votes), the inconsistency of the Table 1.1 rankings means that if the probability measure os equivalent to the plurality vote, profiles of the Table 1.1 type are not admitted by the choice axiom. (By assigning an intensity of likelihood value to each alternative, profiles satisfying Luce's axiom have the property that alternatives retain the same relative ranking with different subsets.) If this profile is not allowed, what is? The admitted profiles are characterized starting in Sect. 3.

The choice axiom introduces intriguing mysteries when constructing $T = \{A, B, C\}$ rankings. One of them compares the effects of an individual constructing a ranking with a "best-first" approach versus a "worst-first" method. For a "best-first" approach, it is reasonable for the likelihood of $A \succ B \succ C$, denoted by $R(A \succ B \succ C)$, to be

$$R(A \succ B \succ C) = P_T(A)P(B, C) \tag{1.8}$$

The subject first selects the top ranked A with likelihood $P_T(A)$, and then selects B from the remaining two alternatives with probability $P(B, C)$. In contrast, a worst-first approach starts with likelihood that C is judged to be inferior, $P^*_T(C)$, and then the likelihood that B is judged the inferior choice of the remaining two, $P^*(B, A)$, to obtain

$$R^*(A \succ^* B \succ^* C) = P^*_T(C)P^*(B, A) \tag{1.9}$$

where P^*_T satisfies the choice axiom with $P^*(X, Y)$.

Under Luce's reasonable assumption that pairwise decisions for either approach agree, that is, $P(A, B) = P^*(B, A)$, one might expect $R(A \succ B \succ C) = R^*(A \succ^* B \succ^* C)$. But Luce proves that this relationship requires a peculiar, restrictive condition.

Theorem 1. *(Luce, 1959 p. 69). Let P and P^* be defined as above for $T = \{A, B, C\}$, and assume that they both satisfy the choice axiom where all pairwise discriminations are imperfect and $P(A, B) = P^*(B, A)$. A necessary and sufficient condition for $R(A \succ B \succ C) = R^*(A \succ^* B \succ^* C)$ is that $P(A, B) = P(B, C)$.*

This unexpected behavior is nicely examined by Yellott (1977, 1997) in terms of Thurstone's theory of comparative judgment. Other insightful papers in this direction include Fishburn (1994) and Marley (1982, 1997). Estes (1997) describes the flavor of the thinking of the time when this theorem and the choice axiom were discovered. The discussion of Thm. 1 given here differs and explains this condition in several ways.

1.4. **Debreu, Tversky, and reversal problems.** While Thm. 1 is discussed throughout this paper, it admits a particularly simple "number of equations and number of unknowns" explanation. A way to motivate the argument is to address G. Debreu's (1960) criticism in a review of Luce's book and to show why Tversky's elimination-by-aspect approach is subject to the same difficulty.

To capture the flavor of Debreu's example that involves the $T = \{A, B, C\}$ alternatives, let A and B represent the same CD featuring Sibelius, but from different stores, so I am indifferent between them. Let C be a CD featuring Mozart where I am indifferent between purchasing a CD featuring Mozart or Sibelius. These assumptions create an example where the likelihood of selecting either alternative from each pair is a half. But according to Eq. 1.6, the choice axiom requires the unreasonable $P_T(A) = P_T(B) = P_T(C) = \frac{1}{3}$. A more realistic choice has an equal likelihood for selecting either a Mozart or Sibelius CD where the Sibelius choices split to define $P_T(A) = P_T(B) = \frac{1}{4}$ while $P_T(C) = \frac{1}{2}$. Debreu argues that, "To meet this difficulty one might say that the alternatives have not been properly defined. But how far can one go in the direction of redefining the alternatives to suit the axiom without transforming the latter into a useless tautology?"

While several explanations of Debreu's comment have been made, the purpose of the simple one advanced here is to introduce the "equation counting" argument used in this paper. The main point is that the choice axiom can model only settings where a natural intensity is attached to each alternative: in contrast, Debreu's example captures a natural situation where the intensity changes with the setting. Restated in terms of the numbers of equations and unknowns, by attaching an intensity to each alternative, Luce's axiom allows $(n-1)$ degrees of freedom for $n \geq 3$ alternatives, so the choice axiom can be used only for settings that can be described with no more than $n-1$ independent expressions (equations). (The n choices of $P_T(A_j), j = 1, \dots, n$, are constrained by $\sum_j P_T(A_j) = 1$.) Debreu's example involves five independent expressions, so rather than a criticism of the choice axiom, his example underscores the reality that we cannot model situations with more independent expressions than allowed degrees of freedom.

It is interesting, and may have been noticed by others, that a similar restriction applies to Tversky's elimination-by-aspect method (1972a, b). This is because with n alternatives the number of aspects defines $2^n - 2$ independent variables coming from the

2^n possible subsets minus the empty set and the set where all aspects agree. But each of the $\binom{n}{j}$ subsets with $n-j$ alternatives define $(n-j-1)$ different $P_S(A_j)$ probabilities. Thus over all subsets of alternatives it is possible to describe

$$\tau(n) = \sum_{j=0}^{n-2} \binom{n}{j}(n-j-1) \tag{1.10}$$

different independent expressions. Consequently, once $\tau(n) > 2^n - 2$, which is for $n \geq 4$, examples with the Debreu flavor can be constructed where Tversky's elimination-by-aspect framework is not applicable. While an example requires more than $2^n - 2$ expressions, the Eq. 1.1 example indicates how to construct such situations.

A slightly more subtle twist on this "number of equations and number of unkowns" argument explains Thm. 1. For $T = \{A, B, C\}$, let $x = P_T(A), y = P_T(B)$, and $1 - x - y = P_T(C)$. Similarly, let $u = P^*(A), v = P^*(B)$, and $1-u-v = P^*(C)$. With the choice axiom, the three conditions $P(A, B) = P^*(B, A)$, or $\frac{x}{x+y} = \frac{v}{u+v}$, $P(A, C) = P^*(C, A)$, and $P(B, C) = P^*(C, B)$ impose two constraints leaving two degrees of freedom. (There are two independent constraints because, according to Eq. 1.7, the probabilities for any two pairs uniquely determines those for the third.) As

$$R(A, B, C) - R^*(A, B, C) = x(\frac{y}{y + (1 - x - y)}) - (1 - u - v)(\frac{v}{u + v}) \tag{1.11}$$

is a single equation with two degrees of freedom, we must expect a line of solutions. Thus Eq. 1.11 has the desired unique zero solution only after imposing another independent condition: the seemingly strange $P(A, B) = P(B, C)$ requirement of Thm. 1 is one choice for this extra condition, and there are others (Sect. 5).

The real objective is to have all ranking probabilities agree. By repeatedly using Thm. 1 with all six rankings, all six ranking probabilities agree if and only if

$$P(A, B) = P(B, C) = P(C, A),\ P(B, A) = P(A, C) = P(C, B) \tag{1.12}$$

where the terms for the first string of equalities come from the rankings

$$A \succ B \succ C, \quad B \succ C \succ A, \quad C \succ A \succ B \tag{1.13}$$

and those for the last string come from $B \succ A \succ C$, $A \succ C \succ B$, $C \succ B \succ A$. These configurations are called *Condorcet triplets,* and, as we will see starting in Sect. 2, they play a central role in voting, probability computations, and Arrow's theorem. According to Eq. 1.12, to satisfy $R(\sigma) = R^*(\sigma)$ for all rankings σ, we have at most one degree of freedom, say the $\frac{a}{a+b}$ value. An equation counting argument shows that the actual restriction is more severe; all rankings have probability $\frac{1}{6}$. This is what Luce (1959), (Luce, Bush, Galanter, 1965) proved; different explanations are in Sects. 5, 6.

1.5. **Pairwise Computations.** While there is no fixed definition for the "likelihood of a pairwise ranking," the intuitive approach defines $P(A,B)$ as the sum of probabilities over rankings where $A \succ B$. This is not necessary; probability computations such as the one given later in Def. 10 could assign more weight to rankings such as $A \succ C \succ B$, which separate A and B, than to $A \succ B \succ C$ where A and B are adjacent.

For notation, if $S \subset T$ and σ is a ranking of the alternatives in S, let $R_S(\sigma)$ denote the probability of ranking σ. It follows from Eq. 1.6 that a probability $P'_T(X)$ (Eq. 7.9) can be constructed that satisfies the choice axiom where

$$P'(A,B) = \sum_{\rho \text{ where } A \succ B} \frac{5}{6} R_T(\rho) + \sum_{\rho \text{ where } B \succ A} \frac{1}{6} R_T(\rho) = \frac{1}{6} + [\frac{2}{3} \sum_{\rho \text{ where } A \succ B} R_T(\rho)]. \tag{1.14}$$

Observe that P' and the standard approach always agree in how the alternatives are ranked according to likelihoods, but the *values* differ; e.g., P' *never* allows perfect discrimination. A natural question is to determine when a standard approach must be used.

Theorem 2. *Luce (1959). Let R_S and P_S be defined for all $S \subset T$ and suppose that*

1. $R_{\{A,B\}}(A \succ B) = P(A,B)$, $R_T(A \succ \rho) = P_T(A) R_{T-\{A\}}(\rho)$.
2. *P_S satisfies the choice axiom,*
3. *all pairwise discriminations are imperfect;*

then

$$P(A,B) = \sum_{\rho \text{ where } A \succ B} R_T(\rho). \tag{1.15}$$

Call Eq. 1.15 the "standard pairwise computation." The next more general definition is used in what follows.

Definition 3. *Let σ be a ranking of the alternatives in $S \subset T$. The* linear computation *of $R_S(\sigma)$ is where $R_S(\sigma)$ is a specified linear combination of $R_T(\rho)$ for all rankings ρ. The* standard computation *of $R_S(\sigma)$ is the special case where*

$$R_S(\sigma) = \sum_{\rho \text{ where } \sigma \text{ is the relative ranking}} R_T(\rho). \tag{1.16}$$

Equation 1.14 illustrates a particular non-standard linear pairwise computation; there exist many other reasonable ways, even nonlinear approaches, to compute these probabilities. While it is interesting to speculate whether individuals use other computational approaches, the standard pairwise approach is emphasized until Def. 10.

1.6. **The "Ranking Axiom".** Replacing $P(A, B)$ with $P'(A, B)$ from Eq. 1.14 establishes the existence of rules over different subsets of alternatives that have different numerical scores for alternatives but always share the same ordinal ranking (based on the probabilities). In other words, it is possible to relax the precision of the choice axiom while preserving its spirit in order to admit settings where a subject does not exhibit Luce's required numerical precision. This alternative axiom, which captures the spirit of Arrow's and Luce's approaches, imposes consistency in how a subject *ranks* alternatives without requiring *precise probability values.*

Axiom 2. [Ranking Axiom]. *For the alternatives $T = \{A_1, \dots, A_n\}$, let S denote a subset. Suppose P is a decision rule defined over all subsets of alternatives where P_S represents the rule restricted to subset S. Decision rule P and an associated subset of profiles $\mathcal{PR}_P$ satisfies the* ranking axiom *if the following hold for each $S \subset T$, $|S| \geq 2$.*

1. *For each S and $\mathbf{p} \in \mathcal{PR}_P$, $P_S(\mathbf{p})$ is a complete, binary, transitive ranking of the alternatives in S.*
2. *The ranking of $P_S(\mathbf{p})$ is the same as the relative ranking of the alternatives from S in $P_T(\mathbf{p})$.*

If a probability measure satisfies the choice axiom for the alternatives in T, it also satisfies the Ranking Axiom. But as Eq. 1.14 demonstrates, a decision rule can fail the choice axiom and still satisfy the Ranking Axiom. Similarly, if a profile restriction admits a method where Arrow's Binary Independence holds for the ranking of all sets S, then it satisfies the Ranking Axiom. The added flexibility offered by this axiom is indicated in Sect. 6 where I introduce a continuum of decision rules that satisfy this condition but not the choice axiom.

1.7. **Outline.** To outline this paper, in Sect. 2, I describe a new approach to analyze voting rules: it decomposes a voting or probability profile into those components that cause different outcomes when used by different kinds of rules. This decomposition of probability profiles explains some mysteries associated with Luce's axiom.

In Sects. 3 and 4, results from voting theory are used to create new probability measures that satisfy the choice axiom, and it is proved that a significantly larger set of probability profiles, much larger than previously expected, satisfy the choice axiom. With these new probability measures, we find new ways to describe the concerns introduced above in Thm. 1, and one of the measures, based on Def. 10, has remarkable properties in that it does not require any profile restrictions and it avoids many of the choice axiom difficulties that occur with other measures. In Sect. 5, different ways to model the Thm. 1 reversal behavior are explored; in all cases ranking probabilities agree only at complete indifference. Section 6 uses the voting material introduced in Sect. 2

to explain the source of several choice axiom mysteries. In this section I explain why the Condorcet triplet (Eq. 1.13) that causes Arrow's negative conclusions plays a vital positive role for Luce's choice axiom. In Sect. 7 it is shown that the new probability measures that satisfy the choice axiom are only special cases, and new consequences of the axiom are described.

2. A decomposition of profiles

The purpose of this section is to describe new results from voting theory that explain all possible election outcomes that can arise with a large class of voting rules. When applied to probability measures and probability profiles, this theory explains a variety of the choice axiom behaviors including the earlier described reversal results. I start by showing how to list profiles so that geometry sorts the profile entries to make it easier to compute election outcomes. To introduce the notions with $n = 3$ alternatives, label the 3! strict, complete, transitive rankings as specified next.

Type number	Ranking	Type number	Ranking
1	$A \succ B \succ C$	4	$C \succ B \succ A$
2	$A \succ C \succ B$	5	$B \succ C \succ A$
3	$C \succ A \succ B$	6	$B \succ A \succ C$

(2.1)

The rankings of $n \geq 3$ alternatives can be geometrically represented with an equilateral n-gon that resides in a $(n-1)$-dimensional space. For $n = 3$, this figure is an equilateral triangle; for $n = 4$ it is an equilateral tetrahedron. Assign each alternative to an unique vertex, and assign rankings to each point in the equilateral n-gon according to the point's distance from each vertex where "closer is better." In Fig. 1a, for instance, a point in the small triangle with a "1" is closer to the A vertex than the B vertex, so it has a $A \succ B$ binary ranking. All points on the Fig. 1a vertical line (connecting the midpoint of the A-B edge with the C vertex) are equidistance from A and B, so they represent where A is tied (or indifferent to) B denoted by $A \sim B$.

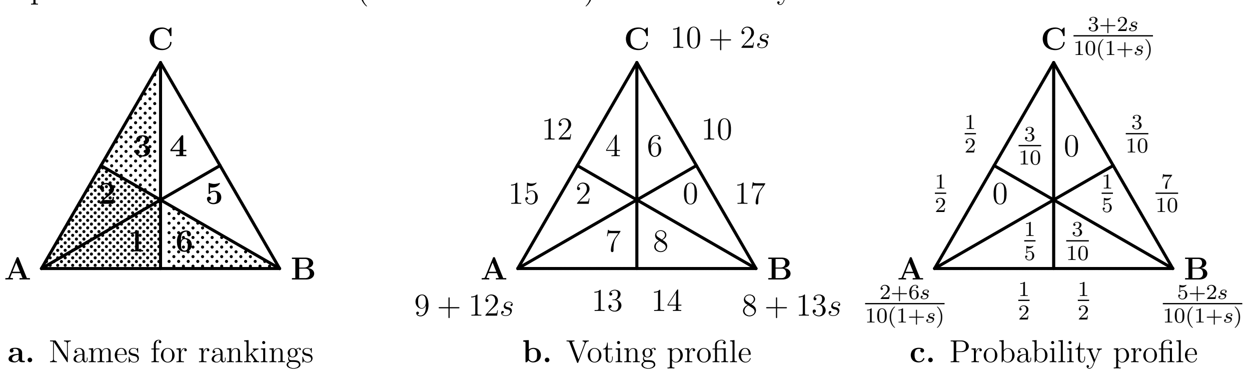


Fig. 1. Representation triangle

In this manner, the n-gon is divided into *ranking regions*. The numbers placed in the small triangles of Fig. 1a identify which ranking region has which Table 2.1 strict ranking; e.g., for a point in the small triangle with a "2," the closest vertex is A, the next closest vertex is C, and the farthest is B, so the assigned ranking is $A \succ C \succ B$. The remaining rankings include the complete indifference $A \sim B \sim C$ outcome (the triangle's barycenter) and six rankings with one pairwise indifference. These last six rankings are represented by the line segments that separate open triangular regions. (For instance, the line segment representing $A \succ B \sim C$ separates regions 1 and 2.) Indifference rankings are excluded to allow a cleaner exposition.

2.1. **Profiles and tallying ballots.** With this notation, the "voting profile space" is

$$\mathcal{VP}^3 = \{\mathbf{n} = (n_1, \dots, n_6) \in R^6 \,|\, n_j \text{ is a non-negative integer}\} \tag{2.2}$$

where n_j is the number of voters with the j^{th} preference ranking. This six-dimensional $\mathcal{VP}^3$ is difficult to envision, so I use the Fig. 1 triangle to represent $\mathcal{VP}^3$ profiles by placing n_j, in the open region representing the jth ranking, $j = 1, \dots, 6$. In this way, Fig. 1b represents the profile $(7, 2, 4, 6, 0, 8)$. Similarly, a "probability profile space" is

$$\mathcal{P}^3 = \{\mathbf{p} = (p_1, p_2, p_3, p_4, p_5, p_6) \in R^6 \mid \sum_{j=1}^{6} p_j = 1,\, p_j \geq 0 \;\forall j\} \tag{2.3}$$

where p_j is the probability that the jth ranking occurs; $j = 1, \dots, 6$. Fig. 1c has the probability profile $(\frac{1}{5}, 0, \frac{3}{10}, 0, \frac{1}{5}, \frac{3}{10})$. (For n-alternatives, use the obvious definitions of $\mathcal{VP}^n$ and $\mathcal{P}^n$.)

To see how this geometric profile representation facilitates computing election tallies, notice that everyone who prefers $A \succ B$ is represented by the numbers to the left of the Fig. 1b vertical line; thus, A's tally in an $\{A, B\}$ election is the $4 + 2 + 7 = 13$ sum of numbers on this side. Similarly, B's tally is the $6 + 0 + 8 = 14$ sum of numbers on the right side. All pairwise tallies are similarly computed and listed by the appropriate edge of the triangle. Likewise, using the standard way to compute probabilities, e.g., $P(B, A) = p_4 + p_5 + p_6$, the computations for all pairs of the Fig. 1c probability profile are posted next to the appropriate edge.

"Positional elections" (or "positional election rules") is the name attached to election methods where ballots are tallied by assigning points to candidates according to how they are positioned (that is, ranked) on the ballot. The widely used "plurality vote," where each voter votes for one candidate and the candidate with the largest number of votes wins, assigns one point for a top-ranked candidate and zero for all others. The antiplurality vote is where one point is given to all but the last-ranked candidate who is assigned zero points.

Definition 4. *A n-alternative positional voting decision rule is defined by voting vector $\mathbf{w}^n = (w_1, w_2, \dots, w_n)$, $w_1 = 1, w_n = 0$, $w_j \geq w_{j+1}$ for $j = 1, \dots, n-1$, where w_j points are assigned to a voter's jth ranked alternative, $j = 1, \dots, n$. The alternatives are ranked according to the sum of assigned points. In "positional probabilistic voting," the number of points assigned to each alternative for a specified ranking is multiplied by the probability of the ranking; the alternatives are ranked according to the sums. (To define "choice probabilities," further normalize the weights so that $\sum w_j = 1$.)*

The (normalized) Borda Count (BC) is

$$\mathbf{b}^n = (1, \frac{n-2}{n-1}, \frac{n-3}{n-1}, \dots, \frac{1}{n}, 0).$$

When different choices of $\mathbf{w}^n$ are used to tally the same ballots, quite different election outcomes can occur. With ten candidates, for instance, a profile can be created whereby *millions* of different election rankings result by varying the choice of $\mathbf{w}^{10}$; indeed, it can be that each alternative wins with some decision rules but is bottom ranked with others (Saari 1992, 2000b). (The same assertion holds for probability profiles.) Because different election rules can create different election rankings, we might want to characterize all possible election outcomes that can result from all $\mathbf{w}^n$, explaining why each occurs, and describing all supporting profiles: this project is completed for any number of alternatives (Saari, 1999, 2000a, b). The $n = 3$ results needed for our purposes are described next. (For details, proofs, and references, see Saari (1999, 2001a).)

The normalized three-candidate voting vector can be expressed as $\mathbf{w}_s = (1, s, 0)$ where s is a specified value satisfying $0 \leq s \leq 1$; e.g., for any s, A's positional $\mathbf{w}_s$ tally is

> [number of voters with A top-ranked]$+s$[number of voters with A second-ranked].

Using Fig. 1a to represent a profile, the values for the first and second brackets would be the sum of numbers, respectively, from the heavily and the lightly shaded regions; e.g., with the Fig. 1b profile, A's $\mathbf{w}_s$ tally is $[2 + 7] + s[4 + 8]$. All $\mathbf{w}_s$ tallies are similarly computed and listed by the relevant vertex. The similar tallies for the Fig. 1c probabilistic profile are further normalized to become probabilities by dividing by $1 + s$.

The Fig. 1b s-coefficients differ among the candidates, so it is reasonable to suspect that this profile admits conflicting election outcomes with different election rules; indeed, candidate A is the Borda winner ($s = \frac{1}{2}$) with the Borda election ranking $A \succ B \succ C$, B is the Condorcet winner (she beats each of the other two candidates in pairwise majority vote elections), and C is the plurality winner ($s = 0$) with the $C \succ A \succ B$ plurality outcome, so each candidate wins with some election rule. A long standing question in social choice theory has been to understand and explain all conflicts of this sort. The basic ideas behind the approach answering these questions (see Saari (1999, 2001a) for $n = 3$ candidate and Saari (2000a, b) for $n \geq 3$ candidates) are described next.

2.2. **Basic idea.** There are certain configurations of voter preferences where it is arguable that the outcome should be a complete tie. Indeed, if a voting rule fails to have a tie in such settings, it is reasonable to suspect that its tallies introduce a bias favoring certain candidates. This leads to the conjecture, which turned out to be true, that all election differences among standard voting rules are caused by this kind of bias.

For three candidates, only two configurations of preferences are needed to explain all differences in voting rules. The first is a *reversal configuration* where two voters have directly opposite rankings such as $A \succ B \succ C$ and $C \succ B \succ A$. To see why the majority vote outcome for any pair is a tie, notice that if a pair is ranked one way in the first ranking, it is ranked in the opposite manner in the second ranking. For positional methods, however, a tie need not occur. As the $A : B : C$ $\mathbf{w}_s$-tally is $1 : 2s : 1$, the outcome is a tie iff $s = \frac{1}{2}$, or iff the Borda rule is used. If $s < \frac{1}{2}$, as true with the plurality vote, the reversal configurations favor (or introduce a bias in favor of) $A \sim C$ over B. Conversely, choices of $s > \frac{1}{2}$ (such as the antiplurality rule) favor B over $A \sim C$. Because the reversal configuration affects only positional methods, we might anticipate, and it is true, that all possible differences among positional methods are caused by profile configurations of this type. Moreover, reversal profile components cause positional election outcomes to differ from the pairwise election outcomes.

To describe the next configuration, what I call the *Condorcet n-tuple*, start with any ranking, say $A_1 \succ A_2 \succ \cdots \succ A_n$. To create the next ranking, move the top-ranked alternative to the bottom, so $A_2 \succ \cdots \succ A_n \succ A_1$: continue until all alternatives have been top-ranked once. To illustrate by starting with $A \succ B \succ C$, the Condorcet triplet, that already has cropped up with the choice axiom in Eq. 1.13, is

$$A \succ B \succ C, \quad B \succ C \succ A, \quad C \succ A \succ B. \tag{2.4}$$

Each candidate in the Condorcet triplet (and n-tuple) is ranked in each position precisely once, so it is arguable that the outcome should be a complete tie. The completely tied behavior is realized by all positional methods with their $A \sim B \sim C$ outcome where each candidate receives $1 + s$ points. The pairwise vote, however, generates the cycle $A \succ B, B \succ C, C \succ A$, each by a 2:1 tally. The reason the pairwise vote fails to have a tie is that, by concentrating on pairs, it cannot recognize nor utilize the broader symmetry that demands the tied outcome. But, by not being a tie, it becomes arguable that the cyclic tallies for this configuration bias the pairwise election outcomes.

As one must anticipate, Condorcet triplets cause pairwise outcomes and tallies to differ from positional outcomes. The important fact (Saari 2000a) is that the Condorcet

n-tuples are completely responsible for all difficulties encountered in pairwise comparisons. For instance, I have shown elsewhere that these Condorcet n-tuples are totally responsible for Arrow's dictator conclusion; e.g., if we restrict attention to profiles that have no Condorcet n-tuple components, Arrow's dictator is replaced with the Borda Count.

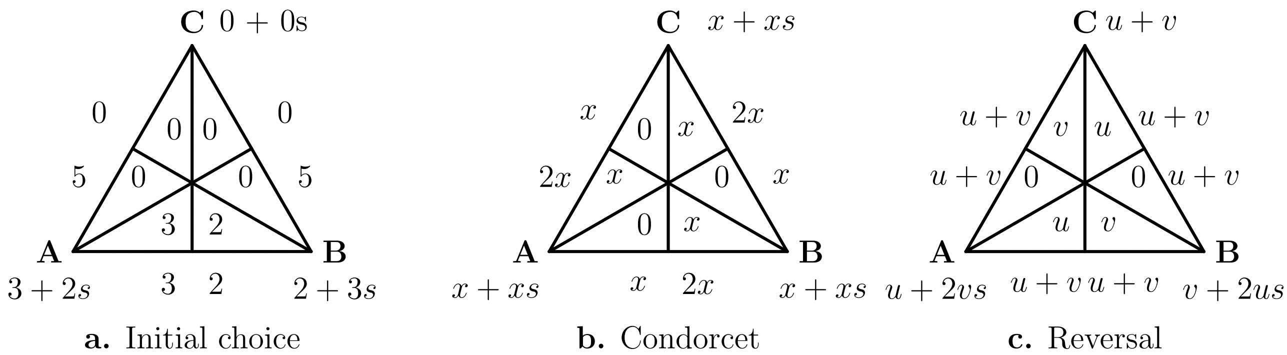


Fig. 2. Creating an example

To illustrate how these biases change election outcomes, let me show how I constructed the Fig. 1b profile. Start with the innocuous Fig. 2a profile where its various tallies are listed by the edges and vertices: with the exception of the antiplurality vote ($s = 1$), all pairwise and positional tallies agree with the $A \succ B \succ C$ outcome. To change the Condorcet winner from A to B without affecting any $\mathbf{w}_s$ ranking, it follows from the above that we must use Condorcet triplets. So, introduce x units of the Condorcet triple $B \succ A \succ C, A \succ C \succ B, C \succ B \succ A$ that favors B over A as represented in Fig. 2b. By adding the pairwise tallies of Figs. 2a, b, it follows that the combined profile has the pairwise majority vote $B \succ A$ and $B \succ C$ rankings iff the inequalities $2 + 2x > 3 + x$, $5 + x > 2x$ are satisfied; e.g., iff $1 < x < 5$. I used $x = 2$.

To change the positional outcomes without affecting pairwise and Borda outcomes, the above comments show that we must add reversal configurations. Select a ranking for any positional method: I selected $C \succ A \succ B$ for the plurality vote ($s = 0$). To favor C, use the two reversal configurations (Fig. 2c) where C is top-ranked. Adding the values by the vertices of Figs. 2a, c (with $s = 0$), this plurality election outcome occurs iff $u + v + 0 > 3 + u > 2 + v$, or $v > 3, u > v - 1$. To create the Fig. 1b profile, use $u = v = 4$.

In this simple manner, profiles exhibiting any collection of admissible election outcomes can be created, and any three-alternative differences in election outcomes can be explained in terms of the reversal and Condorcet components. The important point is that the reversal configurations affect only the positional outcomes, the Condorcet configurations affect only the pairwise outcomes. This interaction between the two types of configurations of preferences will arise in much of what follows.

2.3. **Profile decomposition.** To convert the above description into an analytical tool, it is described in terms of a coordinate system on the space of profiles. To keep the description simple, I consider only three candidates and jump between a profile with integer components (from $\mathcal{VP}^3$) and its probabilistic representation (from $\mathcal{P}^3$). The approach is to decompose a profile into components that reflect the reversal and the Condorcet triplets. The origin of the coordinate system for $\mathcal{P}^3$ is at the *neutral point* denoted by $\mathbf{p}_{Neu} = (\frac{1}{6}, \dots, \frac{1}{6})$.

If $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}^3$, then the sum of the components of $\mathbf{p}_1 - \mathbf{p}_2$ is zero, and $\mathbf{p}_1 - \mathbf{p}_2$ lies in the tangent plane for $\mathcal{P}^3$ denoted by $\mathcal{TP}^3$. (Or, view $\mathcal{TP}^3$ as the plane passing though the origin of R^6 that is orthogonal to $\mathbf{p}_{Neu}$.) The following definition provides a name for the difference between any two probability profiles.

Definition 5. *A* profile differential *is a vector in R^6 where the sum of its components equals zero.*

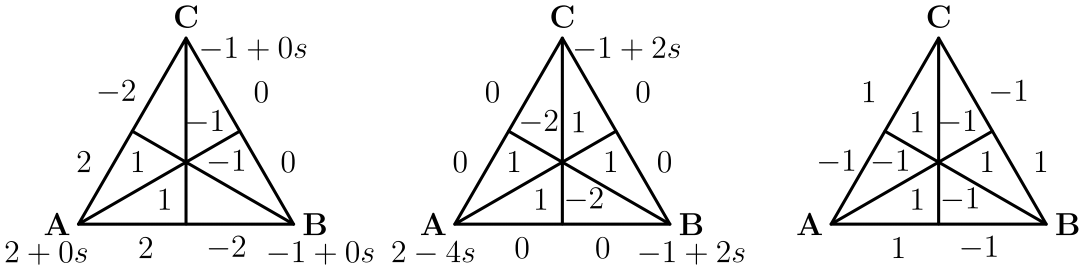
For motivation, notice that the 24-voter profile $(3, 2, 7, 1, 5, 6)$ can be represented as $(3, 2, 7, 1, 5, 6) = (4, 4, 4, 4, 4, 4) + (-1, -2, 3, -3, 1, 2)$; i.e., think of this sum as starting with a *Neutral profile* and using a profile differential to reassign voters from some preference rankings to others to create the desired profile. The coordinates of the following system are profile differentials.

Definition 6. *(Saari 1999) An A_j-*basic profile differential*, $\mathbf{B}_{A_j}$, assigns one point to each ranking where A_j is top-ranked and -1 points to each ranking where A_j is bottom ranked. The basic profile space, consisting of all* basic profile differentials*, is the $\mathcal{TP}^3$ subspace spanned by $\{\mathbf{B}_{A_j}\}_{j=1}^3$.*

*The A_j-*reversal profile differential*, $\mathbf{R}_{A_j}$, assigns one point to each ranking where A_j is top or bottom ranked and -2 points to each ranking where A_j is middle ranked. The reversal subspace, consisting of all reversal profile differentials, is the $\mathcal{TP}^3$ subspace spanned by $\{\mathbf{R}_{A_j}\}_{j=1}^3$.*

The Condorcet profile differential*, $\mathbf{C}$, assigns one point to each of the rankings in the Condorcet triplet $A \succ B \succ C, B \succ C \succ A, C \succ A \succ B$, and -1 points to each of the three remaining rankings that also form a Condorcet triplet.*

The profile differentials for alternative A are illustrated in Fig. 3. The pairwise tallies indicated in the figure use the standard pairwise computation; the $\mathbf{w}_s$ outcomes, if not all zero, are listed next to the appropriate vertices. As it should be clear, the $\mathbf{R}_X$ reversal terms capture the earlier mentioned reversal configurations of the $A \succ B \succ C$ and $C \succ B \succ A$ type while the Condorcet term $\mathbf{C}$ is based on the Condorcet triplet. As described above, the role played by these differentials is that ties occur with certain kinds of voting rules, but not with others: this is indicated by the tallies in Fig. 3.



a. *A*-Basic **b.** *A*-Reversal **c.** Condorcet

Fig. 3. Profile decomposition

Theorem 3. *(Saari 1999) Vectors from the basic, reversal, and Condorcet spaces are mutually orthogonal. The basic and reversal subspaces are two-dimensional; they are spanned by any two of the three defining vectors. Space $\mathcal{TP}^3$ is spanned by the basic, reversal, and Condorcet spaces. The properties of these differentials follow.*

(1) *All $\mathbf{w}_s$ tallies of a basic profile agree; for any $\mathbf{w}_s$ method, the common $\mathbf{B}_{A_j}$ tally assigns 2 points to A_j and -1 points to each of the two remaining alternatives. The $\mathbf{w}_s$ ranking also is consistent with the rankings of the pairs when using the standard pairwise computation; this pairwise computation for $\mathbf{B}_{A_j}$ over a pair including A_j assigns 2 points to A_j and -2 points to the other alternative. If the pair does not involve A_j, both alternatives receive zero points.*

(2) *All pairwise tallies of a reversal differential end in $0:0$ ties. Similarly, the Borda tally (using $\mathbf{w}_{1/2}$) is a complete tie where each alternative receives zero points. For all remaining $\mathbf{w}_s$, the outcome is determined by the relationships were the $\mathbf{w}_s$ tally for $\mathbf{R}_{A_j}$ assigns $2-4s$ points to A_j and $2s-1$ points to each of the other two alternatives. Consequently, for s greater than, or less than, $\frac{1}{2}$, the ranking reverses.*

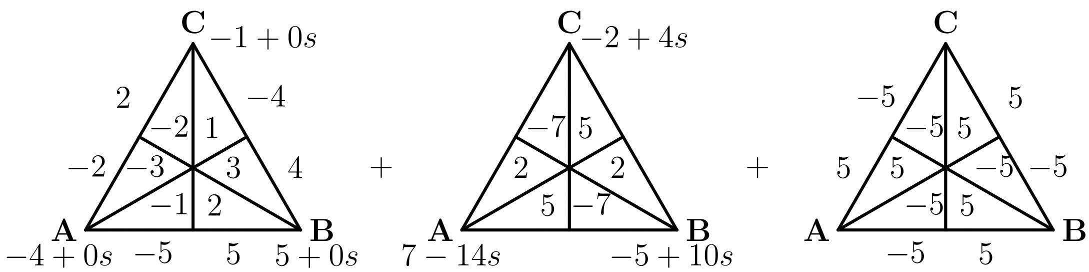
(3) *All $\mathbf{w}_s$ tallies of $\gamma\mathbf{C}$ are a complete tie where each alternative receives zero points. The pairwise computation of $\mathbf{C}$ creates a cycle whereby $A \succ B, B \succ C, C \succ A$ all by $1:-1$ tallies.*

According to Thm. 3, any three-alternative probability profile can be uniquely expressed as

$$\mathbf{p} = \mathbf{p}_{Neu} + \mathbf{p}_{Ba} + \mathbf{p}_{Rev} + \mathbf{p}_{Con}$$

where the Ba, Rev, Con subscripts identify a profile component in, respectively, the Basic, Reversal, and Condorcet $\mathcal{TP}^3$ subspaces. As Thm. 3 asserts, the $\mathbf{p}_{Ba}$ differential has the delightful property that all positional and pairwise decision rules agree: no conflict can occur in tallies or rankings. Consequently, knowing the $\mathbf{w}_0$ plurality tallies for the $\mathbf{p}_{Ba}$ term determines its $\mathbf{w}_s$ tallies for all s (they all agree) and all pairwise outcomes. Conversely, the pairwise tallies can be used to determine all $\mathbf{w}_s$ tallies.

Since $\mathbf{p}_{Ba}$ does not allow any conflict in outcomes among decision rules and subsets of alternatives, it follows that *all conflict among decision rules and subsets of alternatives is caused by the components* $\mathbf{p}_{Con}$ *(for the pairwise tallies) and* $\mathbf{p}_{Rev}$ *(for the positional outcomes).* The $\mathbf{p}_{Con}$ terms distorts the pairwise outcomes by adding a cyclic effect to the values. As Thm. 3 indicates and as shown earlier, the $\mathbf{p}_{Rev}$ term has opposing effects on outcomes depending on whether $s < \frac{1}{2}$, as with the plurality system, or $s > \frac{1}{2}$, as with the antiplurality approach.



a. $\mathbf{p}_{Ba} = 3\mathbf{B}_B + \mathbf{B}_C$ **b.** $\mathbf{p}_{Rev} = 4\mathbf{R}_A + \mathbf{R}_C$ **c.** $\mathbf{p}_{Con} = -5\mathbf{C}$

Fig. 4. Constructing an example

A way to illustrate Thm. 3 is to create a probability profile with the plurality ranking of $A \succ B \succ C$, pairwise rankings that reflect the opposite $C \succ B \succ A$, and the Borda ranking $B \succ C \succ A$. Since only the Borda method is immune to the $\mathbf{p}_{Rev}$ and $\mathbf{p}_{Con}$ terms (Thm. 3), the desired Borda $B \succ C \succ A$ ranking requires using a Basic component such as $\mathbf{p}_{Ba} = 3\mathbf{B}_B + \mathbf{B}_C = (-1, -3, -2, 1, 3, 2)$. (See Fig. 4a.)

According to Thm. 3, the only way the plurality ranking can differ from the Borda outcome is to add appropriate $\mathbf{p}_{Rev}$ terms to the profile. In doing so, notice that the plurality tallies of $\mathbf{p}_{Ba} + x\mathbf{R}_C + y\mathbf{R}_A$ for A, B, C are, respectively, $-4 - x + 2y$, $5 - x - y$, and $-1 + 2x - y$. Thus a plurality $A \succ B \succ C$ ranking requires choosing $y > 3$ and $2 > x \geq 0$. The $x = 1, y = 4$ choices define the $\mathbf{p}_{Rev}$ illustrated in Fig. 4b. According to Thm. 3, $\mathbf{p}_{Rev}$ does not affect the Borda nor the pairwise rankings.

The desired conflict in pairwise rankings can occur only by adding a Condorcet component. By computing the $\mathbf{p}_{Ba}$ pairwise outcomes, and using the Thm. 3 values for Condorcet outcomes, simple algebra shows that adding $\mathbf{p}_{Con} = -5\mathbf{C}$ to $\mathbf{p}_{Ba}$ (Fig. 4c) generates the desired pairwise rankings; $\mathbf{p}_{Con}$ does not affect the positional rankings.

As $\mathbf{p}_{Ba} + \mathbf{p}_{Rev} + \mathbf{p}_{Con} = (-1, 4, -14, 11, 0, 0)$ (add the values for each region from the Fig. 4 triangles), the components become nonnegative by adding 14 to each term: this corresponds to adding the Neutral profile. The total is $(13, 18, 0, 25, 14, 14)$, so a probability profile with the desired properties is

$$\mathbf{p} = (\frac{13}{84}, \frac{18}{84}, 0, \frac{25}{84}, \frac{14}{84}, \frac{14}{84}).$$

Recall from Thm. 3 that the $\mathbf{w}_s$ outcome for $\mathbf{p}_{Rev}$ is one ranking for $s < \frac{1}{2}$ and the opposite for $s > \frac{1}{2}$. A similar phenomenon should occur whenever a profile has a sufficiently large $\mathbf{p}_{Rev}$ component as with the above $\mathbf{p}$: the plurality tally is $(\frac{31}{84}, \frac{28}{84}, \frac{25}{84})$ with an $A \succ B \succ C$ ranking while the antiplurality tally is $(\frac{45}{168}, \frac{56}{168}, \frac{57}{168})$ with the opposite $C \succ B \succ A$ ranking.

3. Ranking probabilities

I now use the structure of voting procedures to analyze lLuce's choice axiom with respect to ranking probabilities. (In what follows, if a proof for a statement is not provided, it is in Sect. 9.) For other papers on ranking probabilities and related questions, see Fishburn (1994, 2002), Luce (1959, 1961, 1962, 1977), Marley (1968, 1982), and Yellott (1977, 1997).

Luce claims that his "best-first" (Eq. 1.8) model, where

$$\mathcal{R}(A \succ B \succ C) = a \times \frac{b}{b+c} = a\frac{b}{1-a}, \tag{3.1}$$

"is, in spirit, closely related to [the choice axiom]; however, it is logically independent of it." A quick way to recognize this logical independence (for $n = 3$) is to use a "number of equations and number of unknowns" argument: the choice axiom does not introduce enough independent equations to determine the six ranking probabilities in a way to satisfy the choice axiom. The refined Eq. 3.1 probabilities impose new assumptions, or, equivalently, add new independent equations to the system. Denote Luce's profile by

$$\mathbf{p}_L(a, b) = (\frac{ab}{1-a}, \frac{ac}{1-a}, \frac{ca}{1-c}, \frac{cb}{1-c}, \frac{bc}{1-b}, \frac{ba}{1-b}). \tag{3.2}$$

To satisfy the choice axiom, the Eq. 3.1 ranking probabilities mandate how the associated probability values of $P_T(X) = x$, $X = A, B, C$ must be computed; the computations must resemble a plurality vote. For instance, to ensure that $P_T(A) = a$, $P_T(A)$ must be the sum of probabilities for the two rankings where A is top ranked; i.e.,

$$P_T(A) = a\frac{b}{1-a} + a\frac{c}{1-a} = a\frac{b+c}{1-a} = a. \tag{3.3}$$

For $n \geq 3$ alternatives where $P(A_j) = a_j$, Eq. 1.8, 3.1 generalizes to

$$R(A_1 \succ \cdots \succ A_n) = a_1 \times \frac{a_2}{1-a_1} \times \ldots \times \frac{a_{n-1}}{1 - \sum_1^{n-2} a_j} \tag{3.4}$$

where to satisfy the choice axiom, Eq. 3.4 implicitly defines an unique way to compute probabilities: it is equivalent to using the plurality vote on each subset S.

As a brief aside, recall (Sect. 1.1) that the choice axiom does not require an universal set; alternatives can be added without renormalizing to make the sum of the probabilities unity. This feature is nicely demonstrated with Eq. 3.4 by observing that the transformation from $R(A \succ B \succ C) = a\frac{b}{1-a}$ to $R(A \succ B \succ C \succ D) = [a\frac{b}{1-a}]\frac{c}{c+d}$ and $R(A \succ B \succ D \succ C) = [a\frac{b}{1-a}]\frac{d}{c+d}$ splits the original $a\frac{b}{1-a}$ value into two parts.

This description of Luce's choice of Eq. 3.1 suggests ways to generalize his notions. The alternative approach used here shows that:

(1) Specifying the profile dictates how to compute $P_T(X)$ probabilities, but the profile restriction is overly severe. A much larger space of admissible profiles (for $n \geq 4$ alternatives), with added modelling flexibility, emerges by specifying only how to compute the probabilities. The admissible profiles are then implicitly defined by the computational decision rule and the choice axiom.

(2) There exists a continuum of natural ways to compute the $P_T(X)$ probabilities. A class that can be identified with voting decision rules provides new insights into the choice axiom.

3.1. **Implicitly defined profiles.** For the rest of this paper, after introducing alternative ways to compute the probabilities are introduced, the choice axiom is used to implicitly define all admissible profiles.

Definition 7. *A* Luce-Plurality probability decision rule *for set S is where the probability for alternative A_j is the sum of the probabilities for each ranking in S where A_j is top-ranked.*

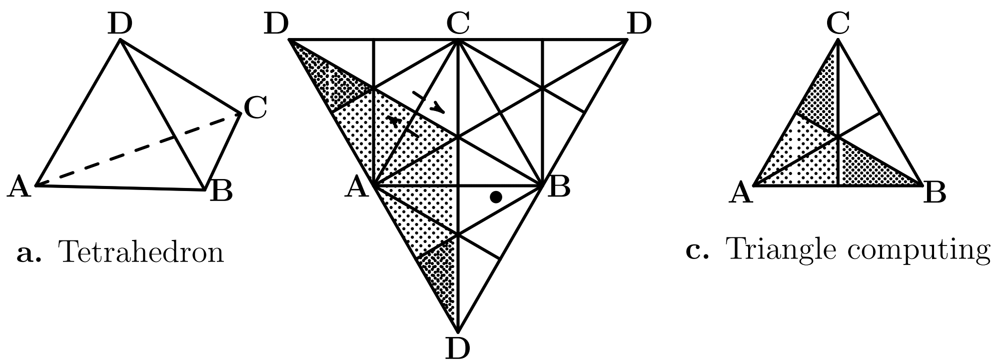
Next it is shown that this approach of implicitly defining the profiles uncovers many new ones. To use an analogy to suggest the source of these new profiles, consider the function $f(x, y) = \sqrt{x^2 + y^2}$. If we concentrate on a specific domain point, such as $(3, 4)$, we have that $f(3, 4) = 5$. But, by letting the outcome 5 define the supporting domain points, we have that $f^{-1}(5)$ is a circle of points. To see the connection with the arguments given next, identify Luce's Eq. 3.4 profile with the specific domain point $(3, 4)$, the positional probability rule with $f(x, y)$, and the relationships imposed by the choice axiom with the f image value of 5. Just as $f^{-1}(5)$ identifies a much larger set of supporting domain points, the approach of implicitly defining the profiles from the choice axioms defines a much larger set of supporting probability profiles. Indeed, this implicit approach augments the Eq. 3.4 profiles with so many new profiles that they must be described in terms of dimensional differences (or degrees of freedom).

Theorem 4. *For $n = 3$ alternatives, the only profiles satisfying the Luce-Plurality decision rule and the choice axiom with the standard pairwise computation form a two-dimensional manifold of $\mathcal{P}^3$; this profile space is parametrically described by Eq. 3.1.*

For $n \geq 4$, the profiles satisfying the Luce-Plurality decision rule and the choice axiom form a $(n-1) + [n! - 2^{n-1}(n-2) - 2]$ dimensional subspace of $\mathcal{P}^n$. This space *is characterized by the $(n-1)$-dimensional space parameterized by Eq. 3.4 and a $[n! - 2^{n-1}(n-2) - 2]$-dimensional "kernel" space of profiles.*

The proof of this theorem is in Sect. 9

To indicate the advantage offered by Thm. 4, while Luce's Eq. 3.4 restricts profiles to a $(n-1)$ dimensional submanifold of $\mathcal{P}^n$, by implicitly defining profiles in terms of the computational method and the choice axiom, a much richer profile space emerges. With $n = 5$ alternatives, for instance, Thm. 4 augments each Eq. 3.4 profile with a 70-dimensional subspace of profiles. When $n = 6$, each profile in the original five-dimensional space defined by Luce's Eq. 3.4 has an associated 590-dimensional space of profiles with the same outcomes and properties. Then, with $n = 10$, Eq. 3.4 defines a nine dimensional space of profiles; the implicit approach attaches to each profile a much richer 3,624,704 dimensional profile subspace.



a. Tetrahedron

b. Unfolded tetrahedron

c. Triangle computing

Fig. 5. Representation triangle and tetrahedron

Now that we know there are other profiles that satisfy the choice axiom, the next task is to identify them. This wider class of $n = 4$ profiles promised by Thm. 4 is described in terms of the *Representation Tetrahedron* (Saari, 2000a); this higher dimensional version of the representation triangle (Fig. 1) represents four alternative profiles. To convert this tetrahedron into a planar figure, choose a vertex (D in Fig. 5a) and cut along the three adjacent edges. By folding the three faces down (Fig. 5b), the 24 ranking regions are displayed in the flat figure. Again, the ranking associated with each region is determined by its distance to the vertices. For instance, the • is in the region closest to a B vertex, next closest to A, then to D, and finally to C, so it represents $B \succ A \succ D \succ C$.

As in Fig. 1b, represent a profile by placing the ranking likelihood in the appropriate ranking region. The Luce-Plurality value for $P_T(A)$ is the sum of the regions where A is top-ranked. The description for the three alternatives of $T = \{A, B, C\}$ was given earlier: recall that this computation sums the values in the two lightly shaded regions of Fig. 5c. Similarly, with the four alternatives $T = \{A, B, C, D\}$, $P_T(A)$ is the sum of values in the six lightly shaded regions of Fig. 5b. The Luce-Plurality value of $P_S(A)$ for

$S = \{A, B, C\} \subset T$ is the sum all Fig. 5b ranking probabilities where A is top-ranked among $\{A, B, C\}$: it is the sum of the six regions in the lightly *and* the two more heavily shaded regions. (In the heavier shaded regions, the missing D is top-ranked and A is second-ranked.) For four alternatives, $P(A, B)$ is the sum of values in the twelve regions to the left of the vertical line in the flattened tetrahedron of Fig. 5b. Other $P_S(U)$ values for $T = \{A, B, C, D\}$ are determined in a symmetric manner.

3.2. **Finding all $n = 4$ profiles.** The proof of Thm. 4 (Sect. 9) has two parts. The first shows that any Eq. 3.4 profile satisfies the choice axiom. The second part comes from Saari (2000a) where it is shown that there exists a $[n! - 2^{n-1}(n-2) - 2]$-dimensional subspace of $\mathcal{P}^n$ where the profiles have no effect upon the plurality and pairwise tallies.

The $n = 4$ profile changes, which augment the Eq. 3.4 probabilities and satisfy the choice axiom, are indicated by the two arrows in Fig. 5b. These arrows mean that after transferring the same probability value in the two directions among the four rankings, all plurality and pairwise tallies remain unchanged. These Fig. 5b arrows, then, represent where the Eq. 3.4 probabilities are altered by

$$\begin{array}{ll}\mathcal{R}(A \succ C \succ B \succ D) = \frac{acb}{(c+b+d)(b+d)} - \gamma_{AC}, & \mathcal{R}(A \succ C \succ D \succ B) = \frac{acd}{(c+d+b)(d+b)} + \gamma_{AC}, \\ \mathcal{R}(C \succ A \succ D \succ D) = \frac{cad}{(a+d+b)(d+b)} - \gamma_{AC}, & \mathcal{R}(C \succ A \succ B \succ D) = \frac{cab}{(a+b+d)(b+d)} + \gamma_{AC},\end{array}$$

where γ_{AC} is any value, positive or negative, constrained only by the requirement that the resulting four ranking probabilities remain non-negative. The subscript indicates that the changes cross the tetrahedron edge connecting vertices A and C. All possible kernel terms come from the six tetrahedron edges where values are symmetrically transferred in the same fashion; the choices from the six edges define the six extra dimensions of profiles promised by Thm. 4. By using the above geometric tallying process, a simple computation shows that such profile changes do not affect the probability values.

According to Thm. 4, each choice of a, b, c, d values for the choice axiom generates a six-dimensional subspace of $\mathcal{P}^4$ rather than just the unique Eq. 3.4 profile. To determine this space, use the a, b, c, d values to define a specific Eq. 3.4 four-alternative profile. The six-dimensional space of associated profiles (with identical probability outcomes for each set S) are obtained from the six γ_{UV} variables associated with changes across the edge connecting vertex U and V, $U, V = A, B, C, D$. The resulting values define a six-dimensional linear space constrained only by the requirement that each ranking probability is between zero and unity.

4. Other probability decision rules

Luce's claim that his definition of $P_T(X)$ is logically independent of the choice axiom suggests that there must exist other approaches. Later I will show how to find a surprisingly wide spectrum of examples, but first a continuum of natural choices, motivated by the $\mathbf{w}_s$ positional voting methods, are described. View these probability measures as representing where a subject makes a distinction between the likelihood of selecting A_j and the likelihood A_j is top-ranked. Namely, in computing $P_T(A_j) = a_j$ that A_j is most desirable, the subject goes beyond using the likelihood that A_j is top-ranked to include information about where A_j is second ranked.

Definition 8. *The three-alternative* positional probability method $\mathbf{w}_s = (1, s, 0)$, $0 \leq s \leq 1$, *is where* $P^s_T(A_j)$ *is the sum of the likelihoods of the two rankings where* A_j *is top-ranked and* s *times the sum from the two rankings where* A_j *is middle ranked all divided by* $(1 + s)$*. Call the three decision rules defined by* $s = 0, \frac{1}{2}, 1$*, respectively, the Luce-Plurality method, the Borda method, and the antiplurality method.*

Using Fig. 5c, compute $P^s_T(A)$ by adding an s multiple of the sum of the entries in the two heavily shaded regions to the sum of the entries in the lighter shaded regions; the final sum is divided by $(1 + s)$. The $P^s_T(A)$, $P^s_T(B)$, $P^s_T(C)$ values for Fig. 1c are, respectively, $\frac{2+6s}{10(1+s)}$, $\frac{5+2s}{10(1+s)}$, $\frac{3+2s}{10(1+s)}$. The term "antiplurality" for $\mathbf{w}_1$ reflects the decision rule's feature of equally weighting a first and second place alternative; it only distinguishes between the top two and the bottom ranked alternative. As $\mathbf{w}_1$ identifies and penalizes a bottom-ranked alternative, $\mathbf{w}_1$ is a reasonable choice to consider when developing profiles from "worst-first" instead of the "best-first" approach.

4.1. Ranking probabilities.

Combined with the choice axiom, each $\mathbf{w}_s$ rule implicitly defines the associated ranking probabilities. A convenient way to find these rankings, and to associate them with the computational scheme, is to expand the six rankings into a space of 12 rankings. Do so by dividing each ranking, say $A \succ B \succ C$, into a larger and smaller part of respective sizes $1/(1 + s)$ and $s/(1 + s)$. Treat $P^s_T(A)$ as adding the larger portion of the two rankings where A is top ranked and the smaller portion of the two rankings where A is middle ranked. So, $\mathbf{w}_{1/4} = (1, \frac{1}{4}, 0)$ divides the jth ranking into a portion of size $1/(1 + \frac{1}{4}) = \frac{4}{5}$, with probability $\frac{4}{5}p_j$, and a smaller portion of size $\frac{1}{5}$, with probability $\frac{1}{5}p_j$. Computing $P^{1/4}_T(A)$ in this manner leads to

$$P^{1/4}_T(A) = \frac{4}{5}(p_1 + p_2) + \frac{1}{5}(p_3 + p_6).$$

This division admits an interesting conditional probability interpretation that might roughly describe how a subject assembles ranking probabilities. To illustrate with $s = \frac{1}{4}$ and the type two ranking of $A \succ C \succ B$, the larger $\frac{1}{1+(1/4)}p_2 = \frac{4}{5}p_2$ could represent the probability of selecting A to be top-ranked given that C is second-ranked, while the smaller $\frac{1}{5}p_2$ is the likelihood of selecting C as second-ranked given that A is top-ranked. Thus, $p_2 = \frac{4}{5}p_2 + \frac{1}{5}p_2$ is the sum of probabilities of the two ways to assemble the $A \succ C \succ B$ ranking. In this fashion, $P_T^{1/4}(A)$ is the sum of the conditional probabilities where A is top and second-ranked. Thus we can view a $\mathbf{w}_s$ method as where for any particular ranking, say $A \succ C \succ B$, the ratio of the likelihood C is second-ranked given that A is top-ranked to the likelihood A is top-ranked given C is second-ranked is s.

As described next in Thm. 5, each $\mathbf{w}_s$ rule determines a two-dimensional section of $\mathcal{P}^3$ profiles. A contribution of this theorem is that it significantly expands the applicability of the choice axiom by identifying a continuum of alternative ways to compute the probabilities. Also, wWith the dissimilarity between the plurality and antiplurality profiles, we must anticipate reversal and other behaviors with the dissimilarity between the plurality and antiplurality profiles, we must anticipate reversal and other behaviors.

Theorem 5. *For three alternatives and each $s \in [0,1]$, $s \neq \frac{1}{2}$, there is a uniquely defined, smooth two-dimensional submanifold $\mathcal{M}_s$ of $\mathcal{P}^3$ profiles for which the $\mathbf{w}_s$ method satisfies the choice axiom. The ranking probability for ranking σ is denoted by $\mathcal{R}^s(\sigma)$.*

The space of probability profiles $\mathcal{M}_1$ for the antiplurality method is parameterized by

$$\mathcal{R}^1(A \succ B \succ C) = b(2 - \frac{1}{1-c}) = \frac{b(1-2c)}{1-c} \tag{4.1}$$

(where the five remaining probabilities are obtained by appropriate permutations of the names of the alternatives) for $\{(a,b,c) \mid a+b+c=1, 0 \le a,b,c \le \frac{1}{2}\}$.

The $\mathcal{M}_s$ space of profiles associated with $\mathbf{w}_s$, $s \neq \frac{1}{2}$, are

$$\mathcal{R}^s(A \succ B \succ C) = \frac{1-s}{1-2s}\mathcal{R}^0(A \succ B \succ C) - \frac{s}{1-2s}\mathcal{R}^1(A \succ B \succ C) \tag{4.2}$$

There do not exist $\mathcal{P}^3$ profiles where the Borda method probabilities (i.e., $s = \frac{1}{2}$) for $P_T^{1/2}(X)$ satisfy the choice axiom with the standard pairwise probability computations.

Proof. Add the probabilities for the rankings in the plurality and antiplurality ways to determine that the pairwise and $P_T^0(X)$ and $P_T^1(X)$ outcomes are as specified. The $\mathbf{w}_s$ conclusions follow from the weighted sum definition of these profiles. The uniqueness assertion follows immediately by specifying the number of equations and unknowns. The assertions about the Borda method are described later □

Notice that Eq. 4.2 is not defined for $s = \frac{1}{2}$, the Borda method, and that the two coefficients change sign at this $s = \frac{1}{2}$ value. While this behavior can be explained in terms of the equations, a deeper explanation is given later and in Sect. 6.

The Eq. 4.2 profile is illustrated in Fig. 6. For a description, start with the $\mathcal{R}^1(\sigma)$ values illustrated in the triangle on the right side of Fig. 6, and view the space as consisting of twelve, rather than six, rankings. The $P^1_T(X)$ computations divide each ranking into two equal parts; the $P^1_T(A)$ computation takes *one of the two portions* from each ranking where A is not bottom ranked. Thus the largest possible $P^1_T(X)$ value is a half; this explains the $a \leq \frac{1}{2}$ upper bound.

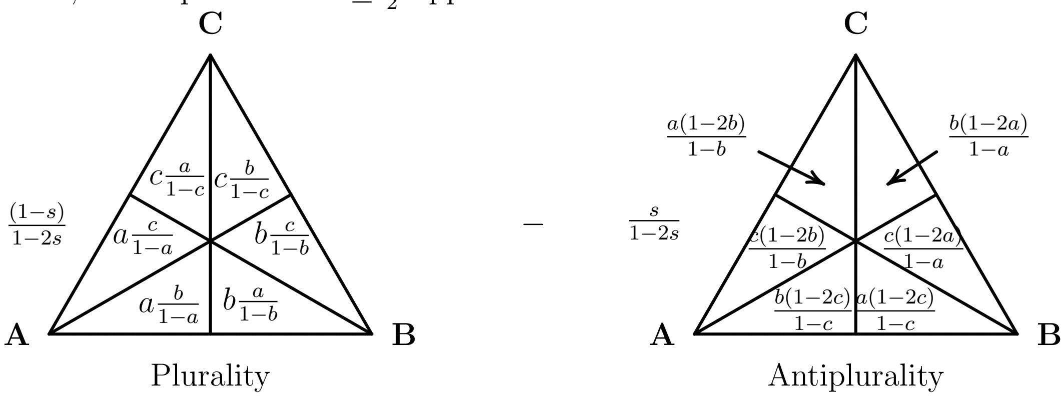


Fig. 6. Defining $\mathbf{w}_s = (1, s, 0)$ profiles

Since Fig. 6 shows that $\mathcal{R}^s(\sigma)$ is a weighted combination of the plurality, $\mathcal{R}^0(\sigma)$, and antiplurality, $\mathcal{R}^1(\sigma)$, ranking probabilities, a way to describe $\mathcal{R}^s(\sigma)$ is to explain the plurality and antiplurality ranking probabilities. Luce derives the plurality rankings from a "best-first" approach, so it remains to examine the $\mathcal{R}^1(\sigma)$ antiplurality computations. As the antiplurality emphasizes the bottom ranked alternative (the only alternative not receiving points), it is reasonable to examine the $\mathcal{R}^1(\sigma)$ values with an Eq. 1.9 "worst-first" scenario.

To start, the probability C is bottom-ranked is p_2+p_3 (the sum of ranking probabilities where C is bottom ranked, see Fig. 1a). To satisfy the choice axiom, the antiplurality outcome is $P^1_T(C) = \frac{1}{2}(p_1+p_4+p_5+p_6) = c$. Combining these equations with $\sum_{j=1}^6 p_j = 1$, it follows that the probability C is bottom ranked is $p_2 + p_3 = (1 - 2c)$. The next Eq. 1.9 step is to compute the likelihood B is bottom ranked in $\{A, B\}$; according to Luce's $P(U, V) = P^*(V, U)$ condition, this is $1 - \frac{a}{a+b} = \frac{b}{a+b} = \frac{b}{(1-c)}$. Multiplying these two values in a Eq. 1.9 sense yields the Fig. 6 and Thm. 5 value of

$$\mathcal{R}^1(A \succ B \succ C) = \frac{b}{(1-c)}(1-2c). \tag{4.3}$$

Thus, as anticipated, $\mathcal{R}^1(\sigma)$ admits a natural "worst-first" explanation.

With this analysis, the $\mathcal{R}^s(\sigma)$ probabilities of Fig. 6 can be viewed as determining ranking probabilities with a weighted combination of "best-first" and "worst-first" considerations – an interpretation compatible with the earlier "conditional probability" representation of assembling ranking probabilities. Moreover, since the $\mathcal{R}^s(A \succ B \succ C)$

value changes with s, different ways to compute the ranking probabilities lead to different values. In this manner, Thm. 5 generalizes the spirit of Luce's reversal result (Thm. 1) to a wider class of approaches and issues.

To conclude this subsection, the "worst-first" ranking probabilities for all $n \geq 3$ are described. As the antiplurality voting vector casts votes for all but the bottom ranked alternative, a probability representation requires dividing each ranking into $(n-1)$ equal portions. The antiplurality outcome is the probability of the union of one portion from each ranking where the specified alternative is not bottom ranked. Over subsets $S \subset T$, each ranking is divided into $|S|-1$ parts; divide each ranking into an appropriate number of portions where $1/(|S|-1)$ of them are used for computations of set S. This leads to a representation with the Eq. 3.4 flavor.

$$\mathcal{R}^{AP}(A_1 \succ \cdots \succ A_n) = (1-(n-1)a_n)[1-\frac{(n-2)a_{n-1}}{1-a_n}]\dots[1-\frac{a_2}{1-\sum_{j=1}^{n-2}a_j}]. \tag{4.4}$$

As with Thm. 4, attached to each Eq. 4.4 choice is a $[n! - 2^{n-1}(n-2)-2]$ dimensional "kernel" space of profiles. All profiles from this larger space satisfy the choice axiom with the standard pairwise computation and the $P^1_S(X)$ computation for each set S.

Corollary 1. *For $n \geq 3$ alternatives, Eqs. 3.4 and 4.4 define probability profiles that with, respectively, the Luce-Plurality and antiplurality methods satisfy the choice axiom. The first can be viewed as where the subject ranks alternatives from "superior choice first and then downwards," while the second represents profiles where a subject ranks them from "inferior choice first and then upwards." Attached to each profile is a $[n! - 2^{n-1}(n-2)-2]$ dimensional subspace with the same properties.*

If subjects use a combination of "best-first" and "worst-first" analyses, or uniformly include information about where an alternative is second ranked, the above identifies the profiles and decision rules.

4.2. **Other consequences.** As a twist on reversal effects (with $n=3$), it is reasonable to wonder whether reversing a $\mathbf{w}_s$ probability profile reverses the $\mathbf{w}_s$ outcome. While this issue differs from Luce's Thm. 1, it most surely is related in spirit. First we need a precise definition for "reversing a profile."

Definition 9. *Let ρ be the operation that reverses a ranking; e.g., $\rho(A \succ B \succ C) = C \succ B \succ A$. For $\mathbf{p} = (p_1, p_2, p_3, p_4, p_5, p_6)$, let $\rho(\mathbf{p}) = (p_4, p_5, p_6, p_1, p_2, p_3)$ be where the probability assigned to a ranking with $\mathbf{p}$ now is assigned to the reversed ranking in $\rho(\mathbf{p})$.*

To develop intuition as to whether reversing a profile provides any useful relationships, experiment with $\mathbf{p}_L(a,b)$ to obtain

$$\rho(\mathbf{p}_L(a,b)) = (\frac{cb}{1-c}, \frac{bc}{1-b}, \frac{ba}{1-b}, \frac{ab}{1-a}, \frac{ac}{1-a}, \frac{ca}{1-c}).$$

The Luce-Plurality outcome with $\mathbf{p}_L(a,b)$ is $P_T^1(A) = a$, $P_T^1(B) = b$, $P_T^1(C) = c$ while the respective $\rho(\mathbf{p}_L(a,b))$ values are $\frac{bc(1+a)}{(1-b)(1-c)}, \frac{ac(1+b)}{(1-a)(1-c)}, \frac{ab(1+c)}{(1-a)(1-b)}$. Beyond proving that if the P_T^0 ranking is $A \succ B \succ C$ for $\mathbf{p}_L$ then the P_T^0 ranking for $\rho(\mathbf{p}_L)$ is the reversed $C \succ B \succ A$, it seems unlikely to find more useful reversal relationships.[1]

A more complicated reversal relationship, however, does exist. To motivate the statement, recall that since the antiplurality method emphasizes the bottom, not the top, ranked alternative it is, in a real sense, the reversal of the plurality method. To explain, since $(1,1,0) = (1,1,1) - (0,0,1)$, a way to compute $P_T^1(X)$ is to add the ranking probabilities where X is *bottom ranked*, subtract this value from 3, and then normalize the values. This is equivalent to subtracting the $P_T^0(X)$ value of $\rho(\mathbf{p})$ from 3. (The value of "3" comes from summing the components of $(1,1,1)$.) Similarly, $\mathbf{w}_{1-s}$ can be viewed as the reversal of $\mathbf{w}_s$ as $(1, 1-s, 0) = (1,1,1) - (0,s,1)$ where $(0,s,1)$ refers to computing the $\mathbf{w}_s$ value of $\rho(\mathbf{p})$.

Denote the "tally" of a profile by $\tau^s(\mathbf{p}) = (P_T^s(A), P_T^s(B), P_T^s(C))$. The following theorem states that an appropriate combination of these reversals always has tallies satisfying a precise numerical relationship; this relationship trivially extends to all $n \geq 3$ values.

Theorem 6. *The following relationship among $P_T^s(X)$ values is satisfied for any probability profile $\mathbf{p}$ and $s \in [0,1]$.*

$$(1+s)\tau^s(\mathbf{p}) + (1 + (1-s))\tau^{1-s}(\rho(\mathbf{p})) = (1,1,1). \tag{4.5}$$

Proof. It suffices to establish Eq. 4.5 for one alternative, say A. According to the definition, the $\mathbf{w}_s$ outcome for $\mathbf{p}$ is

$$(1+s)P_T^s(A) = p_1 + p_2 + sp_3 + sp_6$$

while the $\mathbf{w}_{1-s}$ computation for $\rho(\mathbf{p})$ is

$$(1 + (1-s))P_T^{1-s}(A) = p_4 + p_5 + (1-s)p_3 + (1-s)p_6.$$

The sum of the two expressions is $\sum_j p_j = 1$; this proves the theorem. □

[1]Notice that $\frac{bc(1+a)}{(1-b)(1-c)} > \frac{ac(1+b)}{(1-a)(1-c)}$ if and only if $\frac{1-a^2}{a} > \frac{1-b^2}{b}$. Since $\frac{1-x^2}{x}$ is a strictly decreasing function, the P_T^0 rankings for $\mathbf{p}_L$ and $\rho(\mathbf{p}_L)$ always reverse each other. This relationship need not hold for other probability profiles nor for P_T^1.

It is worth emphasizing that Thm. 6 holds for any probability profile, not just those specified in Thm. 5 and Fig. 6. To suggest applications of Thm. 5, notice that since $\tau^0(\mathbf{p}_L(a,b)) = (a,b,c)$, we obtain $\tau^1(\rho(\mathbf{p}_L(a,b)) = (\frac{1}{2}(1-a), \frac{1}{2}(1-b), \frac{1}{2}(1-c))$ Thus as $a > b > c$ creates a Luce-Plurality ranking of $A \succ B \succ C$, the antiplurality ranking of the reversal of $\mathbf{p}$, of $\rho(\mathbf{p})$, is the opposite $C \succ B \succ A$. Similarly, the $P^0_T(X)$ values of the reversal of the Fig. 6 profile $\mathbf{p}$ for $s = 1$ must be $\tau^0(\rho(\mathbf{p})) = (1-2a, 1-2b, 1-2c)$. In general, applying $\mathbf{w}_s$ to the reversal of the Fig. 6 profile for $1-s$ leads to

$$\tau^{1-s}(\mathbf{p}) = (a,b,c), \quad \tau^s(\rho(\mathbf{p})) = \frac{1}{1+s}(1-(2-s)a, 1-(2-s)b, 1-(2-s)c). \tag{4.6}$$

To illustrate with the Fig. 1c probability profile that does not satisfy the choice axiom, $\tau^0(\mathbf{p}) = (\frac{1}{5}, \frac{1}{2}, \frac{3}{10})$, so $\tau^1(\rho(\mathbf{p})) = (\frac{2}{5}, \frac{1}{4}, \frac{7}{20})$.

A lesson learned from Thm. 6 is that the discussion about reversal effects and Thm. 1 may have concentrated on the wrong relationships. Instead, Eq. 4.5 offers a more accurate reversal argument that holds for *all* probability profiles and goes beyond "best-first" and "worst-first" considerations to relate other reversal effects. More precisely, while we would like $\mathbf{w}_s$ and $\mathbf{w}_{1-s}$ to give the same ranking for a profile $\mathbf{p}$, we come close; $\mathbf{w}_{1-s}$ gives the *reversed ranking* to the *reversed* profile $\rho(\mathbf{p})$. (See Sect. 6.)

It is interesting to notice from Thm. 6 that only the Borda method ($s = \frac{1}{2}$) admits a relationship between a profile and its reversal. This assertion seems unfortunate because in general, the Borda method fails to satisfy the choice axiom with the standard pairwise computation (Thm. 5). But, as shown next, both this new reversal effect and the choice axiom are satisfied by using a nonstandard, nonlinear pairwise computation.

Definition 10. *For $n = 3$, let $P_{Borda}(A,B) = \frac{2(p_1+p_2)+p_3+p_6}{1+2(p_1+p_6)+(p_2+p_5)}$. The numerator of $P_{Borda}(A,B)$ is the sum of ranking probabilities for the rankings where A is middle ranked and twice the ranking probabilities where A is top-ranked. The denominator is the sum of the numerators for $P_{Borda}(A,B)$ and $P_{Borda}(B,A)$.*

$P_{Borda}(A,B)$ determines whether A is more likely than B by differentiating information about where A is top ranked from information where A is second ranked; $P_{Borda}(A,B)$ even uses information from the $B \succ A \succ C$ ranking where B is ranked above A. In terms of Fig. 5c, $P_{Borda}(A,B)$ is the normalized value of twice the sum of the terms in the lightly shaded region (placing more value on where A is top ranked) plus the sum of the terms in the heavily shaded region (where A is second ranked). What adds interest to this nonstandard pairwise computation is that *with P_{Borda}, the Borda method satisfies the choice axiom for* all *probability profiles;* there are no restrictions. While the Luce-Plurality approach holds only when restricted to a two-dimensional selection of probability profiles, the Borda method satisfies the choice axiom for all profiles.

Theorem 7. *For $T = \{A, B, C\}$, if $P_T(X)$ is computed by using the Borda, $s = \frac{1}{2}$, method and pairwise probabilities are computed by using $P_{Borda}(U, V)$, then the system satisfies the choice axiom for all probability profiles.*

The $P_T^{1/2}(X)$ values for a probability profile $\mathbf{p}$ and its reversal $\rho(\mathbf{p})$ are related in the following manner. For any $\mathbf{p}$, we have $\tau^{1/2}(\rho(\mathbf{p})) = (\frac{2}{3}, \frac{2}{3}, \frac{2}{3}) - \tau^{1/2}(\mathbf{p})$.

Proof. The first statement involves an elementary computation, the second is a direct consequence of Eq. 4.5. □

The Borda system satisfies the choice axiom without needing to impose severe constraints on the choice of profiles, and it satisfies a reasonable reversal effect. From a mathematical perspective, this system is ideal for selecting ranking probabilities and using the choice axiom: it appears to be superior to the plurality approach. Whether the system models the behavior of subjects is a different, unexamined issue.

We encounter a mystery. Why does the Borda approach satisfy the choice axiom for all probability profiles while the other $\mathbf{w}_s$ methods are constrained to highly restricted choices? A "number of equations and number of unknowns" answer comes from comparing the information used to compute $P_T^s(A)$ and $P(A, B)$. With Fig. 1a, notice that $P_T^0(A) = p_1 + p_2 = a$ while $P(A, B) = p_1 + p_2 + p_3 = a + p_3$: namely the two decision rules use different information. Consequently, without imposing the severe restriction that $p_3 = \frac{a}{a+b} - a$, the choice axiom cannot be satisfied. Carrying this argument one step further, recall that the choice axiom requires $P(A, B) = \frac{P_T^0(A)}{P_T^0(A)+P_T^0(B)}$. But with $T = \{A, B, C\}$, the information needed to compute $P(A, B)$ is $p_1 + p_2 + p_3$. Compare this information with the different information needed for the right-hand side that is $\frac{P_T^0(A)}{P_T^0(A)+P_T^0(B)} = \frac{p_1+p_2}{(p_1+p_2)+(p_5+p_6)}$; one equation depends on p_3 information while the other requires information about p_5 and p_6. With this radical difference in informational content, very strong restrictions are required in order to satisfy the choice axiom.

In other words, $P_T^0(A)$ uses information coming only from where A is top-ranked while $P(A, B)$ uses this information *and* certain information about where A is second ranked. Obviously, if two decision rules do not use the same information, agreement (e.g., satisfying the choice axiom) occurs only by imposing constraints on the information: the Fig. 6 constraints are mandatory.

This argument has some other subtle messages. In particular, because $P(A, B)$ uses information different from $P_T^0(A), P_T^0(B)$, *the standard pairwise computation violates the Eq. 1.5 spirit that the choice axiom attaches an intrinsic intensity to each alternative.* In contrast, since the Borda method and $P_{Borda}(A, B)$ use the same information to compute outcomes, agreement is possible for all profiles. A similar argument holds for all $\mathbf{w}_s$ methods (where the numerator for $P(A, B)$ is $(p_1 + p_2) + s(p_3 + p_6)$ and the denominator is the sum of the numerators for $P(A, B)$ and $P(B, A)$) when the same form of information is used for all computatons, for different S, and for $n \geq 3$. A different explanation is in Sect. 6.

To illustrate the Thm. 7 statement about the Borda probabilities, since $\tau^{1/2}(\mathbf{p}) = \frac{2}{3}(0.5, 0.6, 0.4) = (\frac{1}{3}, \frac{2}{5}, \frac{4}{15})$ for the Fig. 1c profile, $\tau^{1/2}(\rho(\mathbf{p})) = \frac{2}{3}(0.5, 0.4, 0.6) = (\frac{1}{3}, \frac{4}{15}, \frac{2}{5})$; observe the striking symmetry. To interpret this relationship, notice that $[P_T^{1/2}(X) - \frac{1}{3}]$ describes how the likelihood of selecting X differs from $\frac{1}{3}$: the $\frac{1}{3}$ value is the average likelihood of selecting an alternative. Since $[\frac{2}{3} - P_T^{1/2}(X)] = \frac{1}{3} - [P_T^{1/2}(X) - \frac{1}{3}]$, the second statement asserts for any probability profile that the $\tau^{\frac{1}{2}}(\mathbf{p})$ and $\tau^{\frac{1}{2}}(\rho(\mathbf{p}))$ tallies are the same distance from the average $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ values but, as we might hope, reversing a profile reverses the sign.

These last comments provide an interpretation for Eq. 4.5. Dividing both sides of this equation by three shows that the $\tau^s(\mathbf{p})$ and $\tau^{1-s}(\rho(\mathbf{p}))$ tallies describe a balanced weighted difference relative to the average $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ values.

5. Comments about reversal

The above comments about the Borda method suggest that maybe Thm. 1 can be understood in terms of informational differences in the computations of "best-first" and "worst-first" probability computations. To underscore the point, I include other ways to describe reversal conditions. With the Thm. 5 profiles, for instance, we might wonder when the ranking probabilities associated with $\mathbf{w}_0$ and $\mathbf{w}_1$ agree: agreement holds only for complete indifference.

Corollary 2. *For $n = 3$, the $\mathbf{w}_0$ and $\mathbf{w}_1$ ranking probabilities agree if and only if $a = b = c = \frac{1}{3}$. More generally with the exception of the Borda method, (i.e., for $s_1, s_2 \in [0, \frac{1}{2}) \cup (\frac{1}{2}, 1], s_1 \neq s_2$), the $\mathbf{w}_{s_1}$ and $\mathbf{w}_{s_2}$ values for their ranking probabilities agree if and only if $a = b = c = \frac{1}{3}$. Thus the ranking probabilities agree if and only if all rankings have the same probability of $\frac{1}{6}$.*

Proof. For the $\mathcal{R}^s(A \succ B \succ C)$ and $\mathcal{R}^s(B \succ A \succ C)$ values to agree for $s = 0, 1$, we need that $\frac{ab}{1-a} = \frac{b(1-2c)}{1-c}$ and $\frac{ba}{1-b} = \frac{a(1-2c)}{1-c}$. Cancelling b from the first expression and a from the second leads to

$$\frac{a}{1-a} = \frac{1-2c}{1-c} = \frac{b}{1-b}.$$

The first and last terms require $a = b$. From the first two terms, $c = \frac{1-2a}{2-3a} = \frac{1-(a+b)}{2-3a} = \frac{c}{2-3a}$, or $a = \frac{1}{3}$. That is, $a = b = c = \frac{1}{3}$. Substituting these values into the Fig. 6 expressions force all ranking probabilities to equal $\frac{1}{6}$. The $s_1 \neq s_2$ assertion follows from their different Fig. 6 weights. □

While the above technical proof verifies the assertion, this argument should be viewed in terms of the kind of information used by the different decision rules. For instance, $P^0_T(A) = p^0_1 + p^0_2$ while $P^1_T(A) = p^1_1 + p^1_2 + p^1_3 + p^1_6$ where p^s_j is the Fig. 6 jth ranking probability using the s weight. Since P^1_T goes beyond a $p^1_1+p^1_2$ tally to include $p^1_3+p^1_6$, the significant difference in sources of information requires us to anticipate agreement only in highly restrictive settings. Restating this comment in a mathematical perspective where different sources of information are interpreted as different variables, Cor. 5 must be anticipated because (as shown in Sect. 7) the pairwise probability computations required for the choice axiom define a four-dimensional subspace of $\mathcal{P}^3$ where, generically, two two-dimensional surfaces from P^0 and P^1 meet only in points: this intersection point is the profile of complete indifference.

5.1. **Various reversal approaches.** There are at least three ways to construct the "worst-first" probability for the $A \succ B \succ C$ ranking.

(1) Luce introduces a second probability P^*_T that satisfies his axiom to determine what it means for C to be judged the inferior choice; this leads to Thm. 1.
(2) The antiplurality profile uses the likelihood that C is bottom ranked leading leads to the Fig. 6 values for $s = 1$ and Cor. 2. (The likelihood of being bottom ranked need not be the likelihood of being judged the inferior choice.)
(3) A third approach does not change decision rules (as Luce does) by always using a specified decision rule. The probability C is judged inferior is the probability that C is judged superior for $\rho(\mathbf{p})$.

All three approaches lead to an extreme Cor. 2 type conclusion. To indicate why this happens with the last approach, first define a ranking probability using a "best-first" approach and a $\mathbf{w}_s$ system. For profile $\mathbf{p} = (p_1, \dots, p_6)$, the ranking probability for $A \succ B \succ C$, the p_1 value, must satisfy the $P^s_T(A)P(B,C)$ expression or

$$p_1 = \frac{1}{1+s}[p_1 + p_2 + s(p_3 + p_6)](p_1 + p_5 + p_6). \tag{5.1}$$

A similar expression holds for each p_j.

The "worst-first" outcome of $P^*(C)P^*(B,A)$, where $P^*_T(C)$ is the $P^s_T(C)$ value for $\rho(\mathbf{p}) = (p_4, p_5, p_6, p_1, p_2, p_3)$ and $P^*(B,A) = P(A,B)$, defines the p_1 value equation

$$p_1 = \frac{1}{1+s}[p_1 + p_6 + s(p_2 + p_5)](p_1 + p_2 + p_3) \tag{5.2}$$

with similar expressions for the other p_j values. Although Eqs. 5.1, 5.2 are based on the same information (i.e., the same variables), the weight placed on this information is combined in different independent ways. Consequently, agreement can be expected only in restrictive settings. Stated mathematically, these expressions involve different, independent equations for the variables, so a "number of equations and unknowns" argument indicates that a solution is highly restricted. Indeed, it can be shown (with algebraic manipulations) that agreement holds only for the profile of complete indifference.

5.2. **Worst-first.** Rather than being surprised by disagreements between the "best-first" and "worst-first" ranking probabilities, we should be surprised *if* Thm. 1, or any of the various modifications offered here, allowed agreement with even one degree of freedom. After all, if two two-dimensional surfaces (one defined by the "best-first" approach and the other by the "worst-first" construction) in the four-dimensional subspace of $\mathcal{P}^3$ (required for the pairs to satisfy the choice axiom) intersect in a line, the expressions must have a strong degree of dependency; a dependency that might prove to be valuable.

To explain this dependency comment with more familiar geometry, consider a straight line and a two-dimensional plane in R^3. If the line and plane never intersect, the line has the strong relationship of being parallel to the plane. If the straight line intersects the plane in several points, we have the stronger relationship that the line is *in the plane*; the line is a special case of the plane. The general and expected situation is where the line meets the plane in a single point. Similarly, the general condition for two two-dimensional surfaces in the appropriate $\mathcal{P}^3$ subspace is to meet in a point (or, if not "flat" affine spaces, in isolated points). Consequently, if the surfaces meet in a line, they must enjoy a strong relationship; a relationship that might provide insight into individual decision making.

This extra relationship does not hold for the choice axiom; the only way $\mathcal{R}(\sigma) = \mathcal{R}^*(\sigma)$ agreement can be achieved over all $T = \{A, B, C\}$ rankings is with the probability profile of complete indifference. For $n = 3$ this conclusion is due to Luce (1959), (Luce, Bush, Galanter, 1965). Block and Marschak (1960) extended the result to all n. (Also see Yellott (1997).) What makes this conclusion surprising, as reflected by the attention it has received, is that, intuitively, we expect the subject to use the same information to determine both $\mathcal{R}(\sigma)$ and $\mathcal{R}^*(\sigma)$, so both outcomes should agree. The following proof that $\mathcal{R}(\sigma) = \mathcal{R}^*(\sigma)$ requires "complete indifference" shows that different information is used with the two approaches. As demonstrated, the real source of the problem stems from the standard way to compute pairwise probabilities; e.g., these problems do not arise when the Borda method is used with P_{Borda}. Indeed, the source of the problem becomes apparent just by recognizing that $P(A)P(B, C) = (p_1 + p_2)(p_1 + p_5 + p_6)$ while $P^*(C)P(A, B) = (p_1 + p_6)(p_1 + p_2 + p_3)$: different information is used in each computation.

The first step in computing $\mathcal{R}^*(A \succ B \succ C)$ is to determine $P^*_T(C)$. A natural way to determine the likelihood that C is judged the inferior alternative is to compute the probability that C is the superior choice with $\rho(\mathbf{p}_L)$. This likelihood is

$$p_1 + p_6 = \mathcal{R}(A \succ B \succ C) + \mathcal{R}(B \succ A \succ C) = a\frac{b}{a+b} + b\frac{a}{a+b}. \tag{5.3}$$

Luce shows that the value is different; it is (see Luce, 1959, Thm. 1, page 16)

$$P^*(C) = \frac{ab}{ab + ac + bc}. \tag{5.4}$$

This Eq. 5.4 value dictates the complete indifference conclusion, so it is important to understand, beyond algebraic computations, why it arises.

While the $P_T^*(C)$ computation is the $p_1 + p_6$ sum, the reason the Eq. 5.4 value occurs is for P^* to satisfy the choice axiom. Namely, these p_j terms have different values; call them p_j^*. The source of these different p_j^* values comes from the fact that $P^*(C) = p_1^* + p_6^* = c^*$ while $P^*(C, B) = p_1^* + p_6^* + p_5^* = \frac{c^*}{b^*+c^*}$ where, with the $P(B, C) = P^*(C, B)$ assumption, $P^*(C, B) = \frac{b}{b+c}$. In other words, $P^*(C, B)$ uses more information than $P^*(C)$, so a strict constraint must be imposed on the p_5^* value. (By symmetry, a constraint is imposed on all p_j^* values.) The $P(B, C) = P^*(C, B)$ assumption relates the x and x^* values, $x = a, b, c$. Compare this source of p_5^* with the very different source of information defining p_5; e.g., $p_5 = P(C, A) - P_T(C)$. The following result (particularly Eq. 5.5), shows that this interpretation of the $\mathcal{R}^*(\sigma)$ values is consistent with Eq. 5.4.

Proposition 3. *For $n = 3$, assume P_T and P_T^* are defined over T and its subsets with imperfect discrimination. Suppose P_T satisfies the choice axiom, and that $P_S^*(A_j)$ is the probability that A_j is judged inferior in S. Assume that $P^*(B, A) = P(A, B)$ and that P_T^* satisfies the choice axiom. If $R^*(A \succ B \succ C) = P_T^*(C)P^*(B, A)$, we have that*

$$R^*(A \succ B \succ C) = [\frac{a}{a+b}][\frac{ab}{ab+ac+bc}] = P^*(B, A) - P_T^*(B). \tag{5.5}$$

With the same assumptions for $n \geq 3$, the probability profile is given by

$$R^*(A_1 \succ \ldots \succ A_n) = \prod_{j=2}^{n} P_{S_j}^*(A_j) \tag{5.6}$$

where S_j consists of the top j ranked alternatives $\{A_1, \ldots, A_j\}$ and

$$P_{S_j}^*(A_j) = a_1 a_2 \cdots a_{j-1} / \sum_{i=1}^{j} (a_1 a_2 a_3 \cdots a_j / a_i).$$

Proof. The product representations are direct consequences of (Luce, 1959, Thm. 1, page 16). The subtraction representation follows from Fig. 1 where B is bottom ranked only with type two and three rankings, so $P_T^*(B) = p_2^* + p_3^*$. But since $P^*(B, A)(= P(A, B)) = p_1^* + p_2^* + p_3^*$, we have that $p_1^* = R^*(A \succ B \succ C) = P^*(B, A) - P_T^*(B)$. To check whether this geometric representation agrees with the assumptions, substitute $P^*(B) = \frac{ac}{ab+bc+ac}$ and $P(A, B) = \frac{a}{(a+b)}$ into the second expression; algebraic computations prove that, with imperfect discrimination, equality holds. (By use of Fig. 4, a related, more complicated expression follows for $n = 4$.) □

This difference in the kind of information used to define p_j and p_j^* imposes a strong constraint in order to achieve agreement.

Proposition 4. *Using the assumptions of Thm. 1, a necessary and sufficient condition for* $R(A \succ B \succ C) = R^*(A \succ^* B \succ^* C)$ *is* $P(A,B) = P(B,C)$ *or*

$$P(B,C) - P_T(B) = P(A,B) - P_T^*(B). \tag{5.7}$$

Moreover, $P(B,C) = P(A,B)$ *if and only if* $P_T^*(B) = P_T(B)$.

In words, for ranking probabilities to agree whether computed top-down or bottom-up, a necessary and sufficient condition is that P_T likelihood of the middle ranked alternative being the superior choice equals its P_T^* likelihood of being the inferior choice.

Proof. With imperfect discrimination, $P(A,B) = P(B,C)$ if and only if $ac = b^2$. But $R(A \succ B \succ C) = R^*(A \succ B \succ C)$ if and only if

$$\frac{ab}{(b+c)} = (\frac{a}{a+b})(\frac{ab}{ab+ac+bc}). \tag{5.8}$$

By cross multiplying and collecting terms, equality holds iff $ab(1-a-b-c) = c(b^2-ac)$. As the left hand side is zero, it follows from imperfect discrimination that equality holds iff $b^2 = ac$. This proves the assertion.

Equation 5.7 follows from the p_1 and p_1^* representations. To prove that $P_T^*(B) = b$ iff $P(A,B) = P(B,C)$, notice that $P_T^*(B) = b$ iff $ac = b(ab+ac+bc)$, or $ac(1-b) = b^2(1-b)$. With imperfect discrimination, equality holds iff $ac = b^2$; this proves the assertion. □

As an alternative is middle ranked in precisely two strict rankings, the "worse-first" and "best-first" surfaces agree along the three lines $P^*(X) = P(X)$, $X = A, B, C$. But these lines of agreement intersect only at the point of complete indifference, so the only way all ranking probabilities agree, whether computed from top down or bottom up, is with the profile of complete indifference. This is Luce's (1959) conclusion; also see Yellott (1997). Namely, the combination of the standard way to compute pairwise probabilities with the choice axiom force very different information to be used when computing $\mathcal{R}(\sigma)$ and $\mathcal{R}^*(\sigma)$; these differences permit agreement only with complete indifference.

Theorem 8. *Under the assumptions of Thm. 1,* $p_j = p_j^*$ *for all* $j = 1, \dots, 6$, *iff* $P^*(A) = a, P_T^*(B) = b, P_T^*(C) = c$. *This condition, which requires* $a = b = c = 1/3$, *requires the profile to be of complete indifference where each ranking probability is* $\frac{1}{6}$.

Proof. The equivalence in the first statement follows from the above discussion. According to the above, $P_T^*(B) = b$ iff $P(A,B) = P(B,C)$ iff $p_1 + p_2 + p_3 = p_1 + p_5 + p_6$ iff $p_2 + p_3 = p_5 + p_6$ iff $p_2^* + p_3^* = p_5^* + p_6^*$ iff $P_T^*(B) = P_T^*(C)$ iff $b = c$. The $a = b$ equality is found similarly. The fact that the profile is one of complete indifference follows either from the form of $\mathbf{p}_L(a,b)$ or by solving the three equations for $P_T(X) = \frac{1}{3}$ and the expressions $P(A,B) = P(B,C) = P(A,C) = \frac{1}{2}$ for the p_j values. □

6. Explanations and the Ranking Axiom

To answer some of the remaining mysteries we have encountered, I use the earlier (Sect. 2) results from the positional and pairwise voting literature.

6.1. **Choice Axiom.** Theorem 3 shows that all conflict among outcomes of the pairs and the triplet disappears by restricting attention to the basic profiles, or by using Borda's method and restricting profiles so they have no Condorcet term. Alternatively, we could search for a judicious combination of Condorcet and reversal profiles that alters the pairwise outcomes in just the correct amount to agree with the adjusted $P_T^0(X)$ values of the triplet. This last scenario characterizes precisely what happens with the choice axiom and Luce's profile of Eqs. 3.2, 3.4 as well as all the profiles of Thm. 4, Fig. 6.

All conflict due to the pairwise vote and tallies is caused by the $\mathbf{p}_C$ term (which does not affect $\mathbf{w}_s$ outcomes). This component causes all cycles and non-transitive pairwise rankings, all differences among methods using pairwise votes, all differences between the tallies of the pairwise majority votes and other voting rules and so forth. The explanation, as developed in Saari (1999, 2000a, b) is that the Condorcet term can be viewed as replacing this portion of *transitive preferences* with *cyclic preferences*.

As this last observation explains properties of the choice axiom, it is worth being more explicit. The Condorcet triplet $A \succ B \succ C$, $B \succ C \succ A$, $C \succ A \succ B$ creates a cycle manifested here by two $A \succ B$ and one $B \succ A$ rankings, two $B \succ C$ and one $C \succ B$ rankings, and two $C \succ A$ and one $A \succ C$ rankings. But the same sequence of pairwise rankings results from profiles that violate transitivity by having cyclic preferences. (Two choices, for instance, are the $A \succ B \succ C$, $C \succ B \succ A$ and $A \succ B$, $B \succ C$, $C \succ A$ preferences and the two $A \succ B$, $B \succ C$, $C \succ A$ cyclic choices with one $B \succ A$, $C \succ B$, $A \succ B$.) Even if transitive preferences are intended, it can be shown that the decision rule interprets the data as coming from cyclic preferences.[2] In particular, the Condorcet term causes all problems; indeed, if profiles have no Condorcet

[2]This conflict between actual preferences and the way a decision rule must "interpret" them explains Arrow's Theorem and suggests ways to circumvent his negative assertion (Saari 2001).

terms, then Arrow's conditions are satisfied by the Borda method rather than just a dictator. Also see Saari, Sieberg 2001.)

With probability profiles, the Condorcet term reflects a subject's cyclic ambiguity in ranking the alternatives. It is interesting how the choice axiom implicitly mandates this ambiguity when the standard pairwise computation is used.

The Condorcet term only influences the $P(A, B)$ values, so other profile components must be introduced to influence the $P_T^s(X)$ values to achieve Luce's numeric precision. These are the $\mathbf{p}_R$ terms that have no influence on pairwise or Borda outcomes, but they create differences among $P_T^s(X)$ outcomes (for $s \neq \frac{1}{2}$). The Thm. 5 ranking profiles, then, require a precise level of $\mathbf{p}_R$ components to correspond with the profile's Condorcet component. It is this needed precision among reversal and Condorcet components that imposes the restrictions on the profiles that satisfy the choice axiom.

Recall from Part 2 of Thm. 3 that this $\mathbf{p}_R$ term has an opposing effect on the Luce-Plurality computation (which emphasizes the top-ranked alternative) as with the anti-plurality computation (which emphasizes the bottom ranked alternative). Thus, the larger the magnitude of this $\mathbf{p}_R$ term relative to $\mathbf{p}_B$, the greater the difference when emphasizing the best or the worst alternative. As different P_T^s computations have different $\mathbf{p}_R$ outcomes, a different component level of the Condorcet term is needed to coordinate the $P_T^s(X)$ and $P(A, B)$ values to satisfy the choice axiom; the phrase "compensating errors" is a way to describe this affect of adding components that should cancel. This description, where different amounts of profile differentials are need for different s values, shows that the ranking probabilities can be expected to coincide only when the Condorcet and $\mathbf{p}_R$ values are zero: this is the profile of complete indifference.

6.2. **Borda properties.** Other remaining mysteries are to explain why the Borda $P_T^{1/2}(X)$ is excluded from Thm. 5 and why the Eq. 4.2 coefficients approach infinity as $s \to \frac{1}{2}$. To explain, since the Condorcet component of a profile affects the $P(A, B)$ values, something must be done to appropriately adjust the Borda $P_T^{1/2}(X)$ values. But Thm. 3 proves this is impossible because reversal and Condorcet terms do not effect the Borda outcome. Similarly, as $s \to \frac{1}{2}$, it follows from Thm. 3 that the P_T^s value for $\mathbf{p}_R$ must approach zero. To compensate, larger amounts of $\mathbf{p}_R$ is needed. But as s passes through $\frac{1}{2}$, the reversal terms has a reversed effect on the $P_T^s(X)$ ranking, so the Condorcet term needs to be reversed; the coefficients of Eq. 4.2 reverse sign.

But by replacing the standard pairwise computation with $P_{Borda}(A, B)$, all of the difficulties with the Borda method disappear (Thm. 7). As explained next, $P_{Borda}(A, B)$ is not affected by Condorcet terms, so there is no need to adjust $P_T^{1/2}(X)$ values.

Theorem 9. *For $n = 3$ alternatives, the $P_{Borda}(U, V)$ outcome of any reversal or Condorcet component for any pair is zero. The only terms in a profile that affects the $P_{Borda}(U, V)$ outcome are the neutral $\mathbf{p}_N$ and basic $\mathbf{p}_B$ components.*

Proof. This is a direct computation; e.g., use Fig. 3. □

6.3. **Decompositions.** A way to underscore the above comments about the need for coordination among the components is to identify the $\mathbf{p}_R$ and $\mathbf{p}_C$ components in the Thm. 5 profiles. The next theorem computes this decomposition for any probability profile; the result is then applied to determine the decomposition of Thm. 5 profiles.

Theorem 10. *By vector addition, a profile expressed in the*

$$\mathbf{p} = a_B\mathbf{B}_A + b_B\mathbf{B}_B + a_R\mathbf{R}_A + b_R\mathbf{R}_B + \gamma\mathbf{C}^3 + k\mathbf{K} \tag{6.1}$$

form can be written in the $\mathbf{p} = (p_1, \dots, p_6)$ format. Conversely, for a given probability profile $\mathbf{p} = (p_1, \dots, p_6)$, the vector $\mathbf{v} = (a_B, b_B, a_R, b_R, \gamma, k)$ of coefficients of the Eq. 6.1 profile decomposition are obtained from the matrix expression $\mathbf{v} = T(\mathbf{p})$ where

$$T = \frac{1}{6}\begin{pmatrix} 2 & 1 & -1 & -2 & -1 & 1 \\ 1 & -1 & -2 & -1 & 1 & 2 \\ 0 & 1 & -1 & 0 & 1 & -1 \\ -1 & 1 & 0 & -1 & 1 & 0 \\ 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{6.2}$$

Proof. Expressing Eq. 6.1 in a matrix representation $\mathbf{p} = \mathcal{A}(\mathbf{v})$, we have that matrix $T = \mathcal{A}^{-1}$ converts profile $\mathbf{p}$ into its profile decomposition format. □

The decomposition of the Thm. 5 profiles, determined by using matrix T, follows.

Theorem 11. *The profile decomposition for the Luce-Plurality profiles is*

(1) $a_B = \frac{1}{6}[\frac{ab}{b+c} + a - c - \frac{bc}{a+b} + -b\frac{a-c}{a+c}]$,
(2) $b_B = \frac{1}{6}[a\frac{b-c}{b+c} - c - \frac{ac}{a+b} + b + \frac{ba}{a+c}]$,
(3) $a_R = \frac{1}{6}[ac[\frac{1}{b+c} - \frac{1}{a+b}] + b\frac{c-a}{c+a}]$
(4) $b_R = \frac{1}{6}[a\frac{c-b}{b+c} + bc[\frac{1}{a+c} - \frac{1}{a+b}]]$
(5) $\gamma = \frac{1}{6}[\frac{c(a-b)}{a+b} + \frac{a(b-c)}{b+c} + \frac{b(c-a)}{c+a}]$,

The decomposition for the Antiplurality profiles is

(1) $a_B = \frac{1}{6}[\frac{(2b+a)(1-2c)}{a+b} + \frac{(c-a)(1-2b)}{a+c} - \frac{(2b+c)(1-2a)}{b+c}]$,
(2) $b_B = \frac{1}{6}[\frac{(a+2b)(1-2c)}{a+b} - \frac{(2a+c)(1-2b)}{a+c} + \frac{(c-b)(1-2a)}{b+c}]$,
(3) $a_R = \frac{1}{6}[-\frac{a(1-2c)}{a+b} + \frac{(c-a)(1-2b)}{a+c} - \frac{c(1-2a)}{b+c}]$
(4) $b_R = \frac{1}{6}[-\frac{b(1-2c)}{a+b} + \frac{c(1-2b)}{a+c} + \frac{(c-b)(1-2a)}{b+c}]$
(5) $\gamma = \frac{1}{6}[\frac{(b-a)(1-2c)}{a+b} + \frac{(a-c)(1-2b)}{a+c} + \frac{(c-b)(1-2a)}{b+c}]$,

The decomposition of the $\mathbf{w}_s$ *profiles is* $(1-s)/(1-2s)$ *times the Plurality term minus* $s/(1-2s)$ *times the Antiplurality term.*

Proof. This is $T(\mathbf{p})$ where $\mathbf{p}$ has a Eq. 1.8 representation. □

Notice the careful coordination between Reversal and Condorcet terms. For instance, since $a+b=1-c$, we see from the denominators of a_R and γ how, as a or $a+b$ approach unity, the terms become quite dominant. Also, by experimenting with different a, b values, one can see the different signs of the reversal and Condorcet terms for $s=0,1$. The careful coordination needed to achieve the numerical precision required by the choice axiom is fascinating.

An alternative way to use $\mathbf{w}_s$ methods to define ranking approaches is to find a profile decomposition of the Thm. 3 form that emphasizes the $\mathbf{w}_s$ voting rule. This program has been carried out (Saari 2002): for convenience, definitions are given.

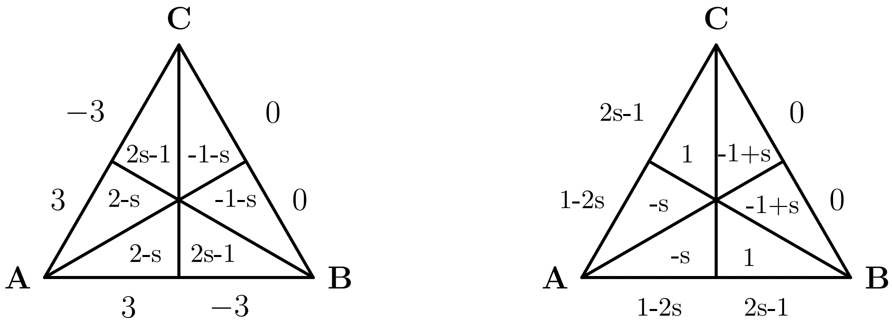
Definition 11. *(Saari 2002) For a specified* $\mathbf{w}_s$ *computation method,* $0 \le s \le 1$*, an* A_j-Basic profile differential*,* $\mathbf{B}^s_{A_j}$*, assigns* $2-s$ *points to each ranking where* A_j *is top-ranked,* $2s-1$ *points to each ranking where* A_j *is middle ranked, and* $-(1+s)$ *points to each ranking where* A_j *is bottom ranked. The* $\mathbf{w}_s$ *Basic profile space, consisting of all* Basic profile differentials*, is the* $\mathcal{TP}^3$ *subspace spanned by* $\{\mathbf{B}^s_{A_j}\}_{j=1}^3$.

The A_j-Orthogonal profile differential*,* $\mathbf{O}^s_{A_j}$*, assigns* $-s$ *points to each ranking where* A_j *is top ranked,* 1 *point where* A_j *is middle ranked, and* $-1+s$ *points to each ranking where* A_j *is bottom ranked. The* $\mathbf{w}_s$ *Orthogonal subspace, consisting of all Orthogonal profile differentials, is the* $\mathcal{TP}^3$ *subspace spanned by* $\{\mathbf{O}^3_{A_j}\}_{j=1}^3$.

Only the $\mathbf{w}_s$ Basic terms affect a $\mathbf{w}_s$ computation (Saari 2002). So, for the Luce-Plurality method, the space of profiles given by

$$\mathbf{p} = a^0_B(2,2,-1,-1,-1,-1) + b^0_B(-1,-1,-1,-1,2,2) + k(1,1,1,1,1,1)$$

describes all outcomes; terms in the Condorcet and Orthogonal direction affect, respectively, pairwise and other $\mathbf{w}_s$ and pairwise computations. What is interesting, and relates to the reversal concerns, is that while the Luce-Plurality method is not affected by any profile components in the $\mathbf{w}_0$ Orthogonal directions of $(0,0,1,-1,-1,1)$ and $(1,-1,-1,1,0,0)$, these profile components *do* affect all other $\mathbf{w}_s$ methods. This conclusion captures, in another sense, our earlier arguments.



a. $\mathbf{w}_s$ Basic for A **b.** $\mathbf{w_s}$ Orthogonal for A

Fig. 7. $\mathbf{w}_s$ standard profiles

It remains to briefly discuss the Ranking Axiom of Sect. 1. A quick approach is to notice that the main difference from the choice axiom is that Ranking Axiom foregoes numerical precision for consistency in rankings. Consequently, it follows from continuity considerations that for each $\mathbf{w}_s$ method there is an open set surrounding the manifold of its Fig. 6 profiles where the ranking axiom is satisfied but the choice axiom is not. In fact, with the profile decomposition, it is possible to describe this region in terms of simple algebra by using an approach similar to that given in Saari (1999) to describe when all voting paradoxes occur. The advantage is obvious; the conditions now become robust in that they hold for a much wider selection of probability profiles. Also, as described above, in a neighborhood about the basic profiles, the Ranking Axiom holds while Luce's does not. This establishes the existence of a continuum of examples that satisfy the Ranking but not the choice axiom.

7. Generalizations and a binary decomposition

It remains to examine more closely the binary probabilities and to show that many of the above results can be generalized. Again, the Fig. 1 geometric profile representation is used. As an aside to show how this approach facilitates computations, recall from Fig. 1a that the standard $P(B, A)$ computation (Def. 3, Eq. 1.16) is the sum of the probabilities of the three rankings in the shaded region; i.e., $P(B, A) = p_4 + p_5 + p_6$. To find all profiles satisfying Debreu's three pairwise conditions where the sum of each pairwise probability is $\frac{1}{2}$, place x, y, and z in each of two diametrically opposing ranking regions. (So, we are using the reversal configurations of Sect. 2) Since all choices of these values require $x + y + z = \frac{1}{2}$, any choice from this continuum suffices.

7.1. **Choice profile sets.**

To implicitly define the profiles and decision rules that satisfy the choice axiom, it is important to describe the restrictions that are imposed on the admissible profiles when its condition 2 is applied *only* to pairs of alternatives.[3] The next definition emphasizes the consequences of the choice axiom for binaries. As we already have seen, further restrictions arise when Luce's conditions are applied to $|S| = j$, $j = 3, \ldots, n$.

Definition 12. *Let the* choice binary outcomes *for the n alternatives $\{A_1, \ldots, A_n\}$ be where there exist n positive scalars $\{a_1, \ldots, a_n\}$, $\sum_{j=1}^{n} a_j = 1$, so that for all $i \neq j$, $P(A_i, A_j) = a_i/(a_i + a_j)$. In the $(n! - 1)$-dimensional profile space $\mathcal{P}^n$, the* choice profile set*, denoted by $\mathcal{CB}^n$, is the set of all profiles $\mathbf{p}$ where the pairwise outcomes under a specified computation are choice binary outcomes.*

To explain this definition, the choice axiom requires $a_i/(a_i+a_j)$ values for the pairwise computations for some choices of a_j. But the manner the profiles are combined to determine these values can vary. So, the "specified computation" refers to how pairwise probabilities are computed. Space $\mathcal{CB}^n$ are those profiles where, with the specified computation of pairwise probabilities, the probabilities have the required form.

To verify that $\mathcal{CB}^3$ restricts profiles, notice that the Fig. 1c profile is not in $\mathcal{CB}^3$. If it were, the standard pairwise computation giving its $A \sim B$ and $A \sim C$ tied outcomes would require, from Eq. 1.7, that $P(B, C) = \frac{1}{2}$: this value contradicts the $\frac{7}{10}$ to $\frac{3}{10}$ dominance of B over C. So, a way to understand the choice axiom is to determine the $\mathcal{CB}^n$ structure. (Also see Fishburn 2002.) While the following result is stated for the standard pairwise computation, it holds for any linear pairwise computation that includes an open set of choice binary outcomes.

Theorem 12. *For $n \geq 2$ alternatives and the standard pairwise computation, $\mathcal{CB}^n$ is a smooth submanifold of profile space $\mathcal{P}^n$ with codimension $\binom{n}{2} - (n-1)$.*

For three alternatives, $\mathcal{CB}^3$ is a codimension $\binom{3}{2} - (3-1) = 1$ submanifold. By excluding only one dimension, $\mathcal{CB}^3$ is a four-dimensional surface in the five-dimensional profile space $\mathcal{P}^3$. As already shown, profiles satisfying $\mathbf{w}_s$ outcomes and the choice axiom form a two-dimensional section of this four-dimensional space. For five alternatives, $\mathcal{CB}^5$ is a $[5! - 1] - [\binom{5}{2} - (5-1)] = 119 - 6 = 113$ dimensional surface in the 119-dimensional space $\mathcal{P}^5$. In contrast, recall that P_{Borda} (Def. 10) significantly extends this dimensional restriction on $\mathcal{CB}^3$.

By being a smooth lower dimensional submanifold (i.e., a smooth surface) of profile space, $\mathcal{CB}^3$ constitutes an unlikely event (i.e., a set with Lebesgue measure zero). Consequently, it is unlikely for an arbitrarily selected profile to satisfy even the binary part

[3]Logically, this description should precede Sect. 3, and it did in an earlier version. But following a referee's suggestion, the material was reorganized to make the paper easier to follow.

of the choice axiom. As "unlikely assertions" are standard for profile restrictions, different restrictions can be compared in terms of the dimensions of their spaces of admitted profiles. This dimensional comparison shows that $\mathcal{CB}^n$ is a reasonably relaxed restriction. For instance, as Black's (1958) condition (imposed in voting to achieve binary consistency) for three alternatives excludes the two voter types (hence two dimensions) where a particular alternative is bottom ranked, it defines a three-dimensional surface of profiles. With the three choices of the specified alternative, Black's condition defines the union of three three-dimensional submanifolds. By excluding two dimensions, rather than only one, Black's approach constitutes a more severe profile restriction than $\mathcal{CB}^3$. But even though Black's condition is stricter, it does not achieve the numerical precision of Luce's system and it allows many other ranking inconsistencies (Saari and Valgones, 1999).

7.2. **Representation cube.** A hurtle in determining the geometry of $\mathcal{CB}^n$ is its large dimension. So, rather than a direct analysis, an indirect approach exploiting the linearity of pairwise tallies is developed. The idea is that since the tally is linear, the pairwise outcomes inherit certain geometric traits from $\mathcal{CB}^n$. For instance, if the pairwise outcomes (the image set) fail to have certain convexity properties, then $\mathcal{CB}^n$ also fails to have these structures. In this indirect manner we prove that $\mathcal{CB}^n$ is a nonlinear submanifold.

While the approach and conclusions hold for any number of alternatives, I emphasize $n = 3$ alternatives so that the conclusions can be described in terms of the familiar three-dimensional geometry. To do so, I use the *representation cube* (Saari, 1995) defined by coordinates $(x_{A,B}, x_{B,C}, x_{C,A})$ where

$$x_{U,V} = P(U,V) - P(V,U), \quad U,V = A,B,C. \tag{7.1}$$

It follows that

$$-1 \leq x_{U,V} \leq 1, \quad x_{U,V} = -x_{V,U}$$

where $x_{U,V} = 1, -1$ represent, respectively, perfect discrimination in selecting U and V.

For any $\mathbf{p} \in \mathcal{P}^3$, the associated $\mathbf{x} = (x_{A,B}, x_{B,C}, x_{C,A})$ resides in the *orthogonal cube;* this is the R^3 cube $[-1,1]^3$ given by the eight vertices $(\pm1, \pm1, \pm1)$. Let $\mathbf{E}_j \in \mathcal{P}^3$ be the probability defined by $p_j = 1$, $\mathbf{E}_j$ is where a selection is made from each pair with perfect discrimination. Thus the six probabilities $\{\mathbf{E}_j\}_{j=1}^6$ define the six vertices

(7.2)

Probability	**Vertex $\mathbf{V}_j$**	**Probability**	**Vertex $\mathbf{V}_j$**
$\mathbf{E}_1$	$(1,1,-1)$	$\mathbf{E}_4$	$(-1,-1,1)$
$\mathbf{E}_2$	$(1,-1,-1)$	$\mathbf{E}_5$	$(-1,1,1)$
$\mathbf{E}_3$	$(1,-1,1)$	$\mathbf{E}_6$	$(-1,1,-1)$

As $\mathbf{p} = (p_1, \ldots, p_6) \in \mathcal{P}^3$ is the convex sum $\sum_{j=1}^{6} p_j \mathbf{E}_j$, the linearity of computing $x_{U,V}$ implies that each $\mathbf{x}$ is a linear combination of the vertices $\{\mathbf{V}_j\}_{j=1}^{6}$.

Definition 13. *The* representation cube *for three alternatives, denoted by $\mathcal{RC}^3$, is the convex hull defined by the vertices $\{\mathbf{V}_j\}_{j=1}^{6}$.*

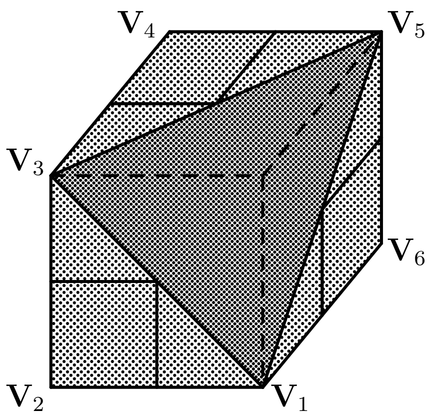


Fig. 8. Representation Cube

There is a representation cube for all $n \geq 4$, but because $\mathcal{RC}^n$ is in a $\binom{n}{2}$ dimensional space (six-dimensional for $n = 4$), indirect arguments are required to analyze its properties. (Readers interested in $\mathcal{RC}^n$ should consult Saari (2000a).) The representation cube $\mathcal{RC}^3$, depicted in Fig. 8, starts with an orthogonal cube. As only six of the eight vertices represent *transitive* unanimity profiles, two of them never arise with probability comparisons. One, $(1, 1, 1)$, represents the impossible cyclic outcome where each of $A \succ B$, $B \succ C$, $C \succ A$ occurs with perfect discrimination. (The other vertex, $(-1, -1, -1)$ represents the reversed cycle.) To convert the orthogonal cube into $\mathcal{RC}^3$, connect the relevant unanimity vertices with lines and flat surfaces; this construction excludes two small tetrahedrons with the cyclic vertices.

Proposition 5. *With a specified linear computational rule for pairs, each $\mathbf{p} \in \mathcal{P}^n$ defines a unique point $\mathbf{x} \in \mathcal{RC}^n$. Conversely, for each $\mathbf{x} \in \mathcal{RC}^n$ and with the standard computational approach, there exist $\mathbf{p} \in \mathcal{P}^n$ that define $\mathbf{x}$. If no coordinate of $\mathbf{x} \in \mathcal{RC}^n$ is ± 1 (i.e., no perfect discrimination), then with the standard approach, $\mathbf{x}$ is supported by a $(n! - 1 - \binom{n}{2})$-dimensional linear subspace of $\mathcal{P}^n$.*

For $n = 3$, Prop. 5 asserts that if $\mathbf{x} \in \mathcal{RC}^3$ is not on a face of the orthogonal cube (which implies perfect discrimination for some pair), then $\mathbf{x}$ is supported by a two-dimensional continuum of profiles. To illustrate, because $\mathbf{p}_1 = (\frac{1}{4}, \frac{1}{5}, 0, \frac{1}{5}, \frac{7}{20}, 0)$, $\mathbf{p}_2 = (\frac{1}{5}, \frac{3}{20}, \frac{1}{10}, \frac{3}{20}, \frac{3}{10}, \frac{1}{10})$, and $\mathbf{p}_3 = (\frac{1}{4}, \frac{1}{20}, \frac{3}{20}, \frac{1}{5}, \frac{1}{5}, \frac{3}{20})$ are linearly independent and each (with the standard pairwise computation) defines $\mathbf{x} = (-\frac{1}{10}, \frac{1}{5}, \frac{1}{10})$, the two-dimensional subspace of profiles

$$\{\mathbf{p} \in \mathcal{P}^3 \,|\, \mathbf{p} = (1-s-t)\mathbf{p}_1 + s\mathbf{p}_2 + t\mathbf{p}_3, \text{ for all } s, t \text{ where } \mathbf{p} \in \mathcal{P}^3\}$$

also defines $\mathbf{x}$. This $n = 3$ two-dimensional space is spanned by the *reversal* vectors defined in Sect. 2.

Recall that $\mathbf{R}_X$ (Def. 6) adds zero points to any candidate's pairwise tally, so the reversal terms do not affect pairwise tallies. Since the sum of the $\mathbf{R}_X$ components is zero, $\mathbf{p}^* = \mathbf{p} + \gamma\mathbf{R}_X$ defines another probability profile if γ is selected so that all $\mathbf{p}^*$ components are non-negative. As already described (Sect. 2), these $\mathbf{R}_X$ vectors span a two-dimensional space and form the kernel of the standard pairwise computations for $n = 3$ alternatives.

Corollary 3. *For $n = 3$, let $\mathbf{p}$ be a specified profile and let $\gamma_A, \gamma_B, \gamma_C$ be scalars so that*

$$\mathbf{p}^* = \mathbf{p} + \sum_{X=A,B,C} \gamma_X \mathbf{R}_X \tag{7.3}$$

also is in $\mathcal{CB}^3$. For the standard computation of pairwise votes $\mathbf{p}$ and $\mathbf{p}^*$ have the same pairwise tallies for all pairs. Conversely, if $\mathbf{p}^*$ and $\mathbf{p}$ have the same pairwise tallies, then they are related in the Eq. 7.3 form. Thus, $\mathbf{p} \in \mathcal{CB}^3$ if and only if $\mathbf{p}^* \in \mathcal{CB}^3$.

This pairwise voting kernel is two dimensional for three alternatives, but its dimension rapidly increases with the number of alternatives; e.g., each $\mathcal{RC}^5$ outcome (that is, probabilities for each of the ten pairs coming from five alternatives) is supported by a 109-dimensional space of profiles. This dimensional jump introduces enormous potential flexibility for modeling.

Combining Prop. 5 with Thm. 12 allows $\mathcal{CB}^n$ to be thought of as the product of an $(n-1)$ dimensional manifold with the $(n! - 1 - \binom{n}{2})$-dimensional pairwise voting kernel. So, to understand the $\mathcal{CB}^n$ structure, it suffices to identify the pairwise voting kernel (Cor. 3) and a particular $(n-1)$-dimensional portion of $\mathcal{CB}^n$.

To demonstrate that $\mathcal{CB}^3$ is very nonlinear, note from Def. 12 and Eq. 7.1 that the binary outcomes in $\mathcal{RC}^3$ allowed by $\mathcal{CB}^3$ profiles define the surface

$$\mathbf{x} = (\frac{a-b}{a+b}, \frac{b-c}{b+c}, \frac{c-a}{c+a}) \in \mathcal{RC}^3.$$

By using $a + b + c = 1$, the parametric representation of the two-dimensional surface of pairwise outcomes in $\mathcal{RC}^3$ generated by $\mathcal{CB}^3$ is

$$\mathbf{x} = (\frac{a-b}{a+b}, \frac{2b+a-1}{1-a}, \frac{1-2a-b}{1-b}), \quad a \geq 0,\, b \geq 0,\, a+b \leq 1. \tag{7.4}$$

Theorem 13. *The pairwise outcomes of $\mathcal{CB}^3$ define a smooth, two-dimensional nonlinear surface in $\mathcal{RC}^3$. Using (x, y, z) as the $\mathcal{RC}^3$ coordinates, the hyperbolic surface* *is*

$$xyz + x + y + z = 0. \tag{7.5}$$

Proof. As Eq. 7.4 is a parametric representation of the $\mathcal{CB}^3$ pairwise outcomes, Eq. 7.5 follows from elementary algebra. □

A convenient way to visualize the geometry of the $\mathcal{CB}^3$ pairwise outcomes is to hold one Eq. 7.5 variable fixed and then describe the resulting line (which is an hyperbola). As Eq. 7.5 is symmetric in the three variables, the analysis is the same for whichever variable is held fixed. Figure 9 displays three sections of this surface.

The symmetry of the $\mathcal{CB}^3$ pairwise outcomes is displayed in Fig. 9. For instance, when $x = 0$ (meaning a $A \sim B$ pairwise tie, or indifference), the section is the straight line $y + z = 0$. Choosing x values that differ only by sign, as in Figs. 9b, c, the resulting sections of $\mathcal{CB}^3$ pairwise outcomes differ by a reflection about the $y + z = 0$ diagonal line. (Similar symmetries hold for any number of alternatives.)

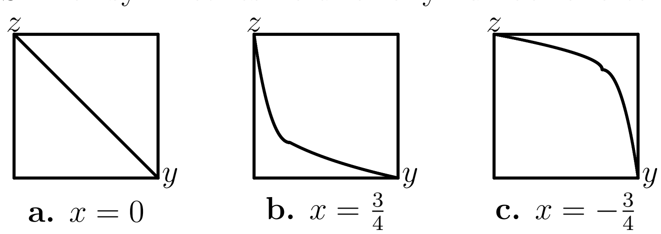


Fig. 9. Sections of the $\mathcal{L}^3$ pairwise outcome space

When $x \to 1$ (that is, when perfect discrimination of A over B is being approached), the curved hyperbola bends to become the union of two particular edges of the $x = 1$ face of $\mathcal{RC}^3$. Similarly, as $x \to -1$ the limit is two of the edges of the $x = -1$ face of $\mathcal{RC}^3$. (These are the two edges that can be drawn without using the cyclic vertex.)

Corollary 4. *The intersection of the closure of the surface of* $\mathcal{CB}^3$ *pairwise outcomes with the* $\mathcal{RC}^3$ *surface is the union of the six* $\mathcal{RC}^3$ *edges connecting vertex* $\mathbf{V}_1$ *to* $\mathbf{V}_2$ *to* … *to* $\mathbf{V}_6$ *to* $\mathbf{V}_1$.

(For $n > 3$ alternatives, the edges are replaced with higher dimensional $\mathcal{RC}^n$ surfaces.)

Proof. On the $x = 1$ face, Eq. 7.5 becomes $yz + y + z + 1 = (y + 1)(z + 1) = 0$. This equation is satisfied if $y, z = -1$. This is the equation of the two edges. □

According to Cor. 4, part 3 of the choice axiom is a boundary condition to ensure continuity for part 2 of the axiom. Namely with perfect $P(A, B) = 1$ discrimination given by $x = 1$, rather than allowing any outcome from the full $x = 1$ face, only those from two edges are allowed. Using Fig. 8 to envision the $\mathcal{CB}^3$ pairwise outcomes, the surface starts as the union of two edges from the $x = 1$ $\mathcal{RC}^3$ face. As x decreases in value, the corresponding section is a hyperbola with ends on the $\mathbf{V}_3$-$\mathbf{V}_4$ edge and $\mathbf{V}_1$-$\mathbf{V}_6$ edges of $\mathcal{RC}^3$. When $x = 0$, the section is a straight line. Then, the surface for negative x values is a reflection of what occurred for $|x|$ until it ends at $x = -1$ as the two edges on the $x = -1$ face (representing $P(B, A) = 1$).

7.3. **Consequences.** If a subject doubts her original evaluation $\mathbf{p}_1$ of the alternatives because evidence suggests that $\mathbf{p}_2$ may be a more accurate assessment, we might expect her to settle on some average. Stated in mathematical terms, it is reasonable to expect the four dimensions of $\mathcal{CB}^3$ to admit an appropriate straight line of profiles. But, $\mathcal{CB}^3$ is so nonlinear that it admits statements of the following kind. (This assertion extends to any number of alternatives.)

Corollary 5. *With any linear computation of pairwise probabilities, the set $\mathcal{CB}^3$ is not convex. Indeed, for any two profiles $\mathbf{p}_1 \neq \mathbf{p}_2$ that satisfy the choice axiom with different pairwise probabilities and without ties, then, with the possible exception of one $\lambda \in (0, 1)$ value, the profile*

$$\mathbf{p}_3 = \lambda \mathbf{p}_1 + (1 - \lambda)\mathbf{p}_2$$

does not satisfy the choice axiom.

Proof. Since $\mathbf{p}_1$ and $\mathbf{p}_2$ define different points on the $\mathcal{CB}^3$ surface in $\mathcal{RC}^3$ without ties, the outcomes is not on a linear portion of the surface. But a line connecting any two distinct points of this type on the surface can meet the surface in at most one other point. As the pairwise computations are linear, the pairwise outcome for $\mathbf{p}_3$ is in the interior of this line segment. This completes the proof. □

According to Cor. 5, if a subject adheres to the choice axiom, she must exhibit *curvilinear,* rather than rectilinear, changes in profiles.[4] The curvature, and this curvilinear comment, are direct consequences of using the standard way to compute pairwise probabilities. In contrast, the P_{Borda} (Def. 10) approach does allow rectilinear choices.

A way to understand this curvature is with the Fig. 8 representation cube. In this cube, the basic profile outcomes (see Sect. 2) lie in what is called the *transitivity plane* (Saari 1999) given by $x + y + z = 0$. It turns out (Saari 1999) that if a profile's outcome is not in the transitivity plane, then the profile has a Condorcet component. As such, Fig. 9 proves that almost all $\mathcal{CB}^3$ outcomes are out of the transitivity plane; e.g., the general case is for a $\mathcal{CB}^3$ profile to have Condorcet components.

[4]As Luce reminded me, the (component wise) geometric mean $p_1^{\lambda} p_2^{1-\lambda}$ *does* satisfy the axiom. This product nature nicely captures the above curvilinear assertion about the nature of profiles that is required by the choice axiom. Furthermore, it underscores a difference with other ways to compute probabilities that are discussed in Sect. 4.

Now that the binary requirements of the choice axiom have been discussed, the next step is to identify the restrictions that are imposed on profiles with three alternative subsets. We already have done so in Sect. 3, but I now provide a more general approach. (For $n = 3$, this completes the description of the profiles because the probability choices of each alternative being top-ranked defines the a, b, c values.)

7.4. **Other probability decision rules.** Luce's claim that his definition of $P_T(X)$ is logically independent of the choice axiom suggests that there must exist other approaches. Already I have shown a continuum of specific choices (Sect. 3), so it remains to find a general result.

To explain in mathematical terms, express Eq. 1.8 as a smooth mapping

$$F : \{(a, b) \,|\, 0 \leq a, b,\ a + b\ \leq 1\} \to \mathcal{P}^3, \tag{7.6}$$

For Eq. 1.8, $F(a, b) = \mathbf{p}_L(a, b)$ is a point in the two-dimensional set of profiles where the $\{A, B\}$ and $\{A, C\}$ respective pairwise probability values are $(\frac{a}{a+b}, \frac{b}{a+b})$, $(\frac{a}{a+c}, \frac{c}{a+c})$. As shown next, most mappings in an Eq. 7.6 format define a way to compute the $P_T(X)$ probability that is compatible with the choice axiom.

Theorem 14. *For $T = \{A, B, C\}$, almost all choices of smooth mappings Eq. 7.6, where $F(a, b) = \mathbf{p}$ is a probability profile with $\{A, B\}$, $\{A, C\}$, and $\{B, C\}$ respective values of $(\frac{a}{a+b}, \frac{b}{a+b})$, $(\frac{a}{a+c}, \frac{c}{a+c})$, and $(\frac{b}{b+c}, \frac{c}{b+c})$ where $c = 1 - a - b$, admit at least a local way to compute the probabilities $P_T(A_j) = a_j$ in a manner that satisfies the choice axiom.*

Proof. Rather than using the implicit function theorem, the proof uses an explicit representation. To ensure that the pairwise computations hold, combine Cor. 3 with $\mathbf{p}_L(a, b)$ to express $F(a, b) = (p_1, \dots, p_6)$ as

$$F(a, b) = \mathbf{p}_L(a, b) + \alpha(a, b)\mathbf{R}_A + \beta(a, b)\mathbf{R}_B. \tag{7.7}$$

To compute $P_T(A)$, add the ranking probabilities where A is top ranked and a weighted "s" multiple of the ranking probabilities where A is second ranked. (In Fig. 6c, add an s multiple of sum of the probabilities from the two heavier shaded regions to the sum of the probabilities from the lightly shaded regions.) The computational scheme for $P_T(A)$ requires choosing the value of s so that

$$a = a + \{2\alpha - \beta + s[-2(2\alpha - \beta) + \frac{ca}{1 - c} + \frac{ba}{1 - b}]\}. \tag{7.8}$$

As long as the term in the square bracket is not zero, and this is true (at least locally) for almost all F, an s value can be found with the desired properties. □

Notice from the proof that these ways to compute $P_T(X)$ use information beyond the rankings where X is top-ranked to include information (through the "s" value) about when X is second ranked. While Thm. 14 proves there are an uncountable number of ways to assign ranking probabilities with an associated $P_T(X)$ computation, it is unsatisfying from an esthetic and maybe a practical perspective. This is because the computational method, the value of s, may change with the a and b values and when computing $P_T(B)$ and $P_T(C)$. This blemish, of course, is avoided in Sect. 4 where a fixed s weight is required for each $P_T(X)$ computation and all a and b.

Theorem 14 holds for any linear pairwise computation method adopted to define $\mathcal{P}^3$. To illustrate by using the $\mathcal{P}^3$ space defined by the pairwise computational method of Eq. 1.14, instead of the standard Eq. 3.1, $F(a,b)$ assigns $p_1 = \frac{3}{2}[\frac{ab}{a+b} - \frac{1}{18}]$ for the ranking $A \succ B \succ C$. To satisfy the choice axiom, the associated $P'_T(A)$ is

$$P'_T(A) = \frac{7}{9}[p_1 + p_2] + \frac{1}{9}[p_3 + p_4 + p_5 + p_6]. \tag{7.9}$$

More generally, $P'_T(X)$, $X = A, B, C$, is the $\frac{2}{3}$ multiple the Luce-Plurality outcome *of these particular ranking probabilities* plus $\frac{1}{9}$. The point is not whether the resulting $P'_T(X)$ is realistic, but rather to demonstrate that the choice axiom holds with a surprisingly rich class of ways to define probabilities.

8. Concluding comment

The natural sense captured by Luce's axiom, where a subject endows each alternatives with a certain level of intensity, probably explains a continued interest in the choice axiom nearly a half century after it was first introduced. But by adopting the perspective that the axiom implicitly defines all ways to compute probabilities along with the associated form of the ranking probabilities, it turns out that many structures and potential extensions of this axiom remain to be discovered. Indeed, while the geometric approach introduced here leads to the creation of a richer selection of alternative computational approaches where the subject uses more information and a significant relaxation on the choice of ranking probabilities, it is only a small indication of what is possible. Also, by examining the information being used to make decisions or compute probabilities, a different explanation of the reversal problem is obtained and, maybe, doubt is cast on the standard way to compute $P(A, B)$ and other probabilities.

9. Proofs

Proof. Thm. 12. The proof uses the codimension argument of the standard implicit function theorem. The general setting, used in singularity theory (e.g., see Golubitsky and Guillemin (1973)) has a smooth mapping $H : R^m \to R^n$ where for a smooth submanifold $\Sigma \subset R^n$ we wish to find the dimension of $H^{-1}(\Sigma)$. If H has a transverse intersection with Σ (that is, where the image of H meets Σ, the tangent space for Σ and $DH(R^m)$ span R^n), then the codimension of $H^{-1}(\Sigma)$ in R^m equals the codimension of Σ in R^n.

Our image space is $R^{\binom{n}{2}}$, the space of all binary outcomes, while H is the mapping computing these outcomes. We know from Saari (1995) that the image has an open set and that the transversality condition is satisfied. If Σ is the set of choice binary outcomes, it is of dimension $n-1$ or codimension $\binom{n}{2}-(n-1)$. The conclusion now follows. □

Proof. Prop. 5. We need to prove that with the standard pairwise computation, each $\mathbf{x} \in \mathcal{RC}^n$ is supported by a $\mathcal{P}^n$ subspace of codimension $\binom{n}{2}$. The argument is the same as the above proof for Thm. 12 where Σ now is the point $\mathbf{x}$ that has codimension $\binom{n}{2}$. The transversality condition follows from the construction of $\mathcal{RC}^n$. □

Proof. Cor. 3. That the reversal terms are in the kernel follows from a direct computation. The proof that this is the total kernel follows from Saari (1999) and the profile decomposition described in Thm. 3. □

Proof. Thm. 4. As the assertion about the dimension of the kernel follows from Saari (2000a, b), it remains to show that the Eq. 3.4 profiles satisfy Luce's conditions; i.e., that $P_T(A_j) = a_j$ and $P_S(A_j) = \frac{P_T(A_j)}{\sum_{A_k \in S} P_T(A_k)}$. To prove that $P_T(A_j) = a_j$, it suffices to prove that $P_T(A_1) = a_1$, and it suffices to consider $A_1 \succ A_2 \succ \cdots \succ A_n$ as a representative ranking. What simplifies the validation is that the form of $\mathcal{R}$ is

$$\begin{aligned} \mathcal{R}(A_1 \succ \cdots \succ A_n) &= a_1 \times \tfrac{a_2}{1-a_1} \times \ldots \times \tfrac{a_{n-1}}{1-\sum_1^{n-2} a_j} \\ &= P_T(A_1) P_{S_{n-1}}(A_2) \ldots P_{S_{n-(n-3)}}(A_{n-2}) P(A_{n-1}, A_n) \end{aligned} \tag{9.1}$$

where $S_{n-k} = \{A_{k+1}, \ldots, A_n\}$. Hence, properties of the probabilities simplify the computations. This is indicated with the computations carried out next.

To verify this assertion $P_T(A_1) = a_1$, I show that if $n-k$ of the n alternatives are selected and ranked in the first $n-k$ position, say $A_1 \succ A_2 \succ \cdots \succ A_{n-k}$, then the sum of the ranking probabilities for the $k!$ ways to complete the ranking with all alternatives, that is $\sum_\sigma \mathcal{R}(A_1 \succ \cdots \succ A_{n-k} \succ \sigma)$ where σ is a ranking of the alternatives $\{A_{n-k+1}, \ldots, A_n\}$, has the form

$$\sum_\sigma \mathcal{R}(A_1 \succ \cdots \succ A_{n-k} \succ \sigma) = a_1 \frac{a_2}{1-a_1} \cdots \frac{a_{n-k}}{1-\sum_{j=1}^{n-(k+1)} a_j}. \tag{9.2}$$

Since Eq. 9.2 is of the $P_T(A_1) P_{S_{n-1}}(A_2) \ldots P_{S_{n-(n-(k+1))}}(A_{n-k})$ form, to change Eq. 9.1 into Eq. 9.2 form, the approach is to sum over all alternatives in $S_{n-(n-(k+1))} =$

$\{A_{n-k}, \dots, A_n\}$ to cancel probability terms at the end of this expression. This is what is done.

For $k = 2$, there are two ways to rank the last two alternatives A_{n-1} and A_n. The sum of the two ranking probabilities has the common factor $a_1 \frac{a_2}{1-a_1} \dots \frac{a_{n-2}}{1-\sum_{j=1}^{n-3} a_j}$, defined by the common ranking of the first $n-2$ alternatives, times $(a_{n-1} + a_n)/(1 - \sum_{j=1}^{n-2} a_j)$. As $\sum a_j = 1$, this last term equals unity, so the sum of the ranking probabilities is the common factor that is the $k = 2$ version of Eq. 9.2. In the Eq. 9.1 terms, we have $\mathcal{R}(A_1 \succ \dots \succ A_{n-1} \succ A_n) + \mathcal{R}(A_1 \succ \dots \succ A_n \succ A_{n-1}) = P_T(A_1)P_{S_{n-1}}(A_2)\dots P_{S_{n-(n-3)}}(A_{n-2})[P(A_{n-1}, A_n) + P(A_n, A_{n-1})]$, where the term in the brackets must equal unity. The same approach is used in an iterative fashion.

Assume Eq. 9.2 holds for $k = i$; we show that it holds for $k = i+1$. Since it holds for $k = i$, for each ranking of the first $n-(i+1)$ alternatives, there are precisely $i+1$ ways to add the next alternative A_α. According to the induction hypothesis, the sum of the ranking probabilities for each way to complete the ranking is $[a_1 \frac{a_2}{1-a_1} \dots \frac{a_{n-(i+1)}}{1-\sum_{j=1}^{n-((i+1)+1)} a_j}] \frac{a_\alpha}{1-\sum_{j=1}^{n-(i+1)} a_j}$. The sum over all of these values has the common factor in the brackets times $\sum_{\alpha=n-i}^{n} a_\alpha/[1-\sum_{j=1}^{n-(i+1)} a_j]$. Again, since $\sum a_j = 1$, the last term equals unity, so the final value is the common factor; this is Eq. 9.2 for $k = i+1$, which completes the induction hypothesis. The $P(A_1) = a_1$ assertion follows from Eq. 9.2 for $k = n-1$.

Again, using the Eq. 9.1 formulation to explain this computation, Eq. 9.2 for $k = i$ becomes $P_T(A_1)P_{S_{n-1}}(A_2)\dots P_{S_{n-(n-(i+2))}}(A_{n-(i+1)})P_{S_{n-(n-(i+1))}}(A_{n-i})$. The goal is to replace the last A_{n-i} with all choices from $S_{n-(n-(i+1))} = \{A_{n-i}, \dots, A_n\}$ and take the sum. This leads to $P_T(A_1)P_{S_{n-1}}(A_2)\dots P_{S_{n-(n-(i+2))}}(A_{n-(i+1)})$ times a sum of terms that must add to unity.

To verify that $P_S(A_j) = \frac{P_T(A_j)}{\sum_{A_k \in S} P_T(A_k)}$, it suffices to prove that

$$P_{S^*_{n-k}}(A_1) = \frac{a_1}{\sum_{j=1}^{n-k} a_j}, \quad S^*_{n-k} = \{A_1, \dots, A_{n-k}\}. \tag{9.3}$$

To explain the difference between Eq. 9.2 and 9.3 by using Fig. 1a, notice that $P_T(A)$ is the sum of the ranking probabilities of regions 1 and 2 while $P(A, B)$ is the sum of the ranking probabilities of regions 1, 2, and 3; hence, with Eq. 9.1,

$$P(A, B) = P_T(A) + P_T(C)P(A, B). \tag{9.4}$$

What is needed is to establish that this equation holds.

For $k = 1$, $P_{S^*_{n-1}}(A_1)$ is the sum of all ranking probabilities where A_1 is top ranked plus those where A_1 is second ranked and A_n is top ranked; notice, this last sum is $P_T(A_n)P_{S^*_{n-1}}(A_1)$ for whatever form $P_{S^*_{n-1}}(A_1)$ assumes. As established above, the first sum is a_1. The sum of ranking probabilities where A_1 is second ranked and A_n is top ranked is, according to Eq. 9.2, $a_n\frac{a_1}{1-a_n}$. Thus $P_{S^*_{n-1}}(A_1) = a_1 + a_n\frac{a_1}{1-a_n} = \frac{a_1}{1-a_n}$, which is Eq. 9.3 for $k = 1$. Restating this computation in Eq. 9.4 terms, we must establish that the specified form of $P_{S^*_{n-1}}(A_1)$ equals $P_T(A_1) + P_T(A_n)P_{S^*_{n-1}}(A_1)$, or that $P_T(A_1)/[1 - P_T(A_n)]$ has the predicted $P_{S^*_{n-1}}(A_1)$ form; a quick calculation proves that it does.

Assuming that Eq. 9.3 holds for $k = i$, we show that it holds for $k = i + 1$. Here, $P_{S^*_{n-(i+1)}}(A_1) = P_{S^*_{n-i}}(A_1) + KP_{S^*_{n-i}}(A_{n-i})$ where $KP_{S^*_{n-i}}(A_{n-i})$ denotes the partial computation of $P_{S^*_{n-i}}(A_{n-i})$, which excludes those ranking probabilities where alternatives from $S^*_{n-(i+1)}$ are ranked above A_1. By use of Eq. 9.1, whatever the form of $P_{S^*_{n-(i+1)}}(A_1)$, we have that $KP_{S^*_{n-i}}(A_{n-i}) = P_{S^*_{n-i}}(A_{n-1})P_{S^*_{n-(i+1)}}(A_1)$. Thus, we need to determine whether $P_{S^*_{n-i)}}(A_1)/[1 - P_{S^*_{n-i}}(A_{n-1})]$ has the predicted representation for $P_{S^*_{n-(i+1)}}(A_1)$. According to the induction hypothesis

$$\frac{P_{S^*_{n-i}}(A_1)}{1 - P_{S^*_{n-i}}(A_{n-1})} = \frac{a_1/\sum_{j=1}^{n-i} a_j,}{1 - [a_{n-i}/\sum_{j=1}^{n-i} a_j]} = \frac{a_1}{a_1 + \cdots + a_{n-(i+1)}}$$

is the desired expression. □

Proof. Cor. 1. Because Eq. 4.4 can be expressed in an Eq. 9.1 format, the proof is essentially the same as the proof of Thm. 4 given above. □

References

[1] Arrow, K.J. (1963).*Social Choice and Individual Values*, 2nd ed., Wiley, New York.
[2] Black, D. (1958). *The Theory of Committees and Elections,* Cambridge University Press, London.
[3] Block, H. D., and J. Marschak (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, and H. Mann (Eds.), *Contributions to Probability and Statistics* (pp. 97-132). Stanford University Press.
[4] Debreu, G. (1960). Review of 'Individual Choice Behavior: A Theoretical Analysis,' **50** *American Economic Review* 186-188.
[5] Estes, W. K. (1997), Some reflections on the role of the choice model in theories of categorization, identification, and learning, pp 321-346 in Marley, A. A. J., *Choice, Decision, and Measurement; Essays in Honor of R. Duncan Luce*, Erlbaum Assoc. New Jersey
[6] Fishburn, P. (1994), On choice probabilities derived from ranking probabilities, *Jour. Math. Psychol.*. **29**, 274-285.
[7] Fishburn, P. (2002), Stochastic Utility, in *Handbook of Utility*, ed. S. Barbera, J. Hammon, and C. Seidl. Dordrecht/Boston: Kluwer Academic Publishers, 275-319.
[8] Golubitsky, M., and V. Guillemin (1973), *Stable Mappings and Their Singularities*, Springer-Verlag.
[9] Luce, R. D. (1959). *Individual Choice Behavior*, Wiley, New York.
[10] Luce, R. D. (1961), A choice theory analysis of similarity judgments, *Psychometrika* **26**, 151-163
[11] Luce, R. D. (1962), An observable property equivalent to a choice model for discrimination experiments, *Psychometrika* **27**, 163-167.
[12] Luce, R. D. (1977), The choice axiom after twenty years, *Jour. Math Psychology* **15**, 215-233
[13] Luce, R. D., R. R. Bush, and E. Galanter 1965, *Handbook of Mathematical Psychology* Vol. III, Wiley, New York.
[14] Marley, A. A. J. (1968), Some probabilistic models of simple choice and ranking, *Jour. Math. Psychol.* **5**, 311-332.

[15] Marley, A. A. J. (1982), Random utility models with all choice probabilities expressible as 'functions" of the binary choice probabilities, *Math. Soc. Sci.* **3**, 39-56.
[16] Marley, A. A. J. (1997), *Choice, Decision, and Measurement; Essays in Honor of R. Duncan Luce*, Erlbaum Assoc. New Jersey.
[17] Regenwetter, M., A. a. J. Marley, and H. Joe, (1998), Random utility threshold models of subset choice, *Australian Journal of Psychology* **50**, 175-186.
[18] Saari, D. G. (1989) A dictionary for voting paradoxes, *Jour. Economic Theory* **48**, 443-475.
[19] Saari, D. G. (1995). *Basic Geometry of Voting*, Springer-Verlag, New York.
[20] Saari, D. G. (1992), Millions of election rankings from a single profile,*Soc. Choice & Welfare* **9**, 277-306.
[21] Saari, D. G. (1998) Connecting and resolving Arrow's and Sen's Theorems, *Social Choice & Welfare* **15**, 239-261.
[22] Saari, D. G. (1999) Explaining all three alternative voting outcomes, *Jour. Econ. Theory* **87**, 313-355.
[23] Saari, D. G. (2000a) Explaining positional voting paradoxes I; The binary case, *Economic Theory,* **15**, 1-53.
[24] Saari, D. G. (2000b), Explaining positional voting paradoxes II; The general case, *Economic Theory,* **15**, 54-103.
[25] Saari, D. G., (2001a), *Chaotic Electioins! A Mathematician Looks at Voting.* American Math Society, Providence, RI
[26] Saari, D. G. (2001b), *Decisions and Elections; Explaining the Unexpected*, Cambridge University Press.
[27] Saari, D. G., (2002), Adopting a plurality vote perspective; *Math of Operations Research* **27**, 45-64.
[28] Saari, D. G., and K. Sieberg (2001), the sum of the parts can violate the whole, *Amer. Pol. Science Rev.* **95**, 415-433.
[29] Saari, D. G. and F. Valgones (1999), The geometry of Black's single peakedness condition, *Jour. Math. Econ.* **32**, 429-456.
[30] Tversky, A. (1972a). Elimination by aspects: A theory of choice. *Psychological Review,* **79**, 281-299.
[31] Tversky, A. (1972b), Choice by elimination. *Journal of Mathematical Psychology,* **9**, 341-367.
[32] Yellott Jr., J. I. (1977), The relationship between Luce's Choice Axiom, Thurston's theory of comparative judgement, and the double exponential distribution. *Journal of Math Psychology* **15**, 109-144.
[33] Yellott Jr., J. I. (1997), Preference models and irreversibility, pp 131-151 in Marley, A. A. J., *Choice, Decision, and Measurement; Essays in Honor of R. Duncan Luce*, Erlbaum Assoc. New Jersey

Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA 92698-5100

E-mail address: `dsaari@uci.edu`