Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Mathematical Psychology 48 (2004) 107–134

Journal of Mathematical Psychology

http://www.elsevier.com/locate/jmp

Binary choice, subset choice, random utility, and ranking: A unified perspective using the permutahedron

Jun Zhang*

Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA

Received 3 December 2002; revised 8 December 2003

Abstract

The d-permutahedron $\Pi_{d-1} \subset R^d$ is defined as the convex hull of all d-dimensional permutation vectors, namely, vectors whose components are distinct values of a d-element set of integers $[d] \equiv \{1, 2, \ldots, d\}$. By construction, Π_{d-1} is a convex polytope with $d!$ vertices, each representing a linear order (ranking) on $[d]$, and has dimension $\dim(\Pi_{d-1}) = d - 1$. This paper provides a review of some well-known properties of a permutahedron, applies the geometric-combinatoric insights to the investigation of the various popular choice paradigms and models by emphasizing their inter-connections, and presents a few new results along this line.

Permutahedron provides a natural representation of ranking probability; in fact it is shown here to be the space of all Borda scores on ranking probabilities (also called "voters profiles" in the social choice literature). The following relations are immediate consequences of this identification. First, as all $d!$ vertices of Π_{d-1} are equidistant to its barycenter, Π_{d-1} is circumscribed by a sphere S^{d-2} in a $(d-1)$-dimensional space, with each spherical point representing an equivalent class of vectors whose components are defined on an interval scale. This property provides a natural expression of the random utility model of ranking probabilities, including the condition of Block and Marschak. Second, Π_{d-1} can be realized as the image of an affine projection from the unit cube $\mathscr{C}_{d(d-1)/2}$ of dimension $d(d-1)/2$. As the latter is the space of all binary choice vectors describing probabilities of pairwise comparisons within d objects, Borda scores can be defined on binary choice probabilities through this projective mapping. The result is the Young's formula, now applicable to any arbitrary binary choice vector. Third, Π_{d-1} can be realized as a "monotone path polytope" as induced from the lift-up of the projection of the cube $\mathscr{C}_d \subset R^d$ onto the line segment $[0, d] \subset R^1$. As the 2^d vertices of the d-cube $\mathscr{C}_d$ are in one-to-one correspondence to all subsets of $[d]$, a connection between the subset choice paradigm and ranking probability is established. Specifically, it is shown here that, in the case of approval voting (AV) with the standard tally procedure (Amer. Pol. Sci. Rev. 72 (1978) 831), under the assumption that the choice of a subset indicates an approval (with equal probability) of all linear orders consistent with that chosen subset, the Brams–Fishburn score is then equivalent to the Borda score on the induced profile. Requiring this induced profile (ranking probability) to be also consistent with the size-independent model of subset choice (J. Math. Psychol. 40 (1996) 15) defines the "core" of the AV Polytope. Finally, Π_{d-1} can be realized as a canonical projection from the so-called Birkhoff polytope, the space of rank-position probabilities arising out of the rank-matching paradigm; thus Borda scores can be defined on rank-position probabilities. To summarize, the many realizations of a permutahedron afford a unified framework for describing and relating various ranking and choice paradigms.

1. Background

In the general context of decision-making through choice over d distinct objects (political candidates, consumer products, job options, etc.), denote the master choice set $[d] := \{1, 2, \ldots, d\}$, with objects conveniently labelled as natural numbers 1 through d ($d > 2$ is assumed in this paper unless explicitly noted otherwise). Common choice paradigms which arise from a wide variety of political, social–economical and psychological situations include:

1. Linear ordering/ranking, in which all d objects are rank-ordered[1] according to their levels of desirability, assuming no ties.

*Tel.: +1-734-763-6161; fax: +1-734-763-7480.
E-mail address: junz@umich.edu.

[1] For simplicity, only linear/total order and sometimes weak order of objects are considered; no considerations for semi-order or interval order will be given in this paper.

0022-2496/$ - see front matter
doi:10.1016/j.jmp.2003.12.002

2. Binary choice, in which two of the d objects (called a "binary subset") are selected at a time to be compared against each other for their desirability, and all such pairwise comparisons of objects in $[d]$ are performed.
3. Subset choice, in which only one distinct subset of $[d]$ of any number of elements (including the null-set $\emptyset$) is selected or approved of, while the desirability of all the selected elements is considered to be "equal" or comparable (their precise meanings to be clarified later).
4. Rank assignment, in which each rank-position (1 through d, in *ascending* order of desirability) is to be assigned, as a partition of unity, to objects in $[d]$.

When there are more than one individuals ("voters") involved in performing such a binary choice, subset choice, or linear ranking task, or when such choice is non-deterministic, a probability distribution will be induced over the set of, respectively, all linear orders over $[d]$, all binary subsets of $[d]$, or all subsets of $[d]$. For instance, the probability distribution over the $d!$ linear orders is termed the "voters profile" in the social choice literature or "ranking probability" elsewhere, the probability distribution over the 2^d subsets of $[d]$ is called "subset choice probability", while the more common "binary choice probability" refers to a vector with $d(d-1)/2$ components, each representing the probability of choosing one alternative over another in a two-element subset of $[d]$.

There have been attempts in the past to link the preference structure revealed by these various choice paradigms using some unifying constructs. One common approach is to assume the existence of an underlying interval-scaled utility associated with each choice option (a candidate in $[d]$). These utilities, represented by a d-dimensional vector $\mathbf{v} = [v_1, v_2, \ldots, v_d]^T$ indeed, may vary randomly, though often not necessarily independently. The distribution of such random utility (RU) values is denoted by a non-negative function $f(v_1, v_2, \ldots, v_d)$ that satisfies

$$\int f(v_1, v_2, \ldots, v_d)\, dv_1\, dv_2 \cdots dv_d = 1,$$

with f yet to be interpreted (see next paragraph). This so-called (non-parametric) "Random Utility (RU) model", as a probabilistic generalization of deterministic utility theory, has been argued to provide the much-needed tool for combining algebraic and probabilistic representations of choice and, therefore, to serve as a unifying framework to reconcile normative and descriptive approaches to modelling and measurement of choice in social sciences (Regenwetter & Marley, 2001).

One important consequence of RU representation of choice is its immediate connection to the ranking paradigm. Denote a total (linear) ranking of all candidates in $[d]$ by $\pi = \langle \pi(1)\pi(2)\cdots\pi(d) \rangle$, where $\pi(i)$ is the rank (also called "rank-position") given to candidate $i, i \in [d]$. As a convention adopted in this paper, the rank-positions are natural numbers 1 through d, forming a set denoted by $[d]$ as well, with larger values indicating higher desirability. For convenience, π^{-1} denotes the mapping from rank-position to candidate identity, so $\pi^{-1}(k)$ represents the candidate who occupies rank-position k according to ranking π, while $\pi^{-1}(d)$ returns the most-desirable candidate. In the literature as well as in this paper, $\langle \pi(1)\pi(2)\cdots\pi(d) \rangle$ is referred to as "ranking", while $(\pi^{-1}(1)\pi^{-1}(2)\cdots\pi^{-1}(d))$ is referred to as "ordering" (in increasing desirability). Denote the set of all rankings of d candidates as $\mathscr{L}_d$, with set size $d!$. The "ranking probability" P_π can be defined on the set of linear orders $\mathscr{L}_d$:

$$P_\pi \geqslant 0, \quad \sum_{\pi \in \mathscr{L}_d} P_\pi = 1. \tag{1}$$

Block and Marschak (1960) showed that under certain conditions, the ranking probability P_π is naturally connected to the density $f(v_1, v_2, \ldots, v_d)$ of jointly distributed RU variable $\mathbf{v} = [v_1, v_2, \ldots, v_d]^T$ through the relationship

$$P_\pi = \int_{\Delta_\pi} f(v_1, v_2, \ldots, v_d)\, dv_1\, dv_2 \cdots dv_d, \tag{2}$$

where the region of integration Δ_π is the connected point-set

$$\Delta_\pi = \{\mathbf{v} \in R^d : v_{\pi^{-1}(d)} > v_{\pi^{-1}(d-1)} > \cdots > v_{\pi^{-1}(1)}\}. \tag{3}$$

The necessary and sufficient condition for (2) to hold is the "non-coincidence" condition, roughly stated as

$$f(v_1, v_2, \ldots, v_d) = 0 \quad \text{whenever } v_i = v_j \ (i \neq j). \tag{4}$$

This is to say that the set of points where any two random variables assume the same value has measure zero; probability density is not concentrated on the boundaries of Δ_π. Block and Marschak (1960) demonstrated that any such density function $f(v_1, v_2, \ldots, v_d)$ induced a ranking probability P_π according to (2); conversely a density function could always be constructed to satisfy (2) given an arbitrary ranking probability P_π.

This result is intuitively important: assuming the non-coincidence condition, a characterization of RU representation is equivalent to a characterization of ranking probability P_π. As much as non-coincident RU representation may become a unifying description for choice behavior (argued by M. Regenwetter), ranking probability would play a central role in relating choice probabilities associated with the various paradigms mentioned earlier.

1.1. Ranking probability and binary choice

The binary choice paradigm is perhaps one of the most studied and best axiomatized paradigm of choice. Let a_{ij} be the relative frequency (or probability) of choosing object "i" in a two-element set $\{i,j\}$. There can be many axiomatizations and representations of the probability of binary choice. For instance, the "Weak Utility Model" postulates that $a_{ij} > a_{ji}$ if and only if there exists a real-valued function $g(\cdot)$ such that $g(i) > g(j)$. The "Strong Utility Model" postulates that $a_{ij} > a_{kl}$ if and only if there exists a real-valued function g such that $g(i) - g(j) > g(k) - g(l)$ for all i,j,k,l. The "Strict Utility Model" (Bradley–Terry–Luce or BTL Model) further requires the existence of a positive-valued function $g(\cdot)$ such that $a_{ij} = \frac{g(i)}{g(i)+g(j)}$. The "Fechnerian Utility Model" assumes the existence of a real-valued function g and a strictly monotone increasing function ϕ, such that the binary choice probability takes the form $a_{ij} = \phi[g(i) - g(j)]$, where ϕ is often taken as the cumulative distribution function of a normal random variable, or at other times the logistic function. Finally, the "RU model" (see (2), with $d = 2$) assumes the existence of random variables v_i and v_j associated with i and j such that $a_{ij} = \mathrm{Prob}[v_i > v_j]$.

When more than two candidates are involved, binary choice can be performed on any two-element subset of $[d]$, $d > 2$. In this case, the choice probabilities become a $d(d-1)/2$-dimensional vector $\mathbf{a} = [a_{12}, a_{13}, \ldots, a_{(d-1)d}]^T$, called the "binary choice vector", whose range is the unit cube

$$\mathscr{C}_{d(d-1)/2} := \{[a_{12}, a_{13}, \ldots, a_{(d-1)d}]^T \in R^{d(d-1)/2} : 0 \leqslant a_{ij} \leqslant 1,\ i,j \in [d],\ i < j\}$$

representing all possible pairwise comparisons of candidates in the master set $[d]$. Note that any constraint placed on the binary-choice probability (such as transitivity) will necessarily limit the feasible region in the cube $\mathscr{C}_{d(d-1)/2}$. For example, for any triplets i,j,k which are elements of $[d]$, weak stochastic transitivity (WST) assumes that if $a_{ij} \geqslant \frac{1}{2}$ and $a_{jk} \geqslant \frac{1}{2}$ then $a_{ik} \geqslant \frac{1}{2}$, moderate stochastic transitivity (MST) assumes that if $a_{ij} \geqslant \frac{1}{2}$ and $a_{jk} \geqslant \frac{1}{2}$ then $a_{ik} \geqslant \min(a_{ij}, a_{jk})$, while strong stochastic transitivity (SST) assumes that if $a_{ij} \geqslant \frac{1}{2}$ and $a_{jk} \geqslant \frac{1}{2}$ then $a_{ik} \geqslant \max(a_{ij}, a_{jk})$, see Luce and Suppes (1965). These increasingly stronger assumptions about the underlying process of binary comparison will constrain the realizable region to some increasingly smaller core in the cube $\mathscr{C}_{d(d-1)/2}$. While important for various axiomatizations, in this paper no such constraints will be placed on the binary choice vector $\mathbf{a}$ unless explicitly stated.

The binary choice vector $\mathbf{a}$ can be related to the ranking probability P_π defined for ranking $\pi \in \mathscr{L}_d$ of elements of $[d]$, where $\mathscr{L}_d$ denotes the set of all linear orders. A sensible relationship between binary choice and ranking is

$$a_{ij} = \sum_{\{\pi \in \mathscr{L}_d \,:\, \pi(i) > \pi(j)\}} P_\pi, \tag{5}$$

where the summation is over all π's that rank candidate i as more desirable than candidate j. Given P_π, one obtains a_{ij} through marginalization, so each ranking probability gives rise to a unique binary choice vector. But the inference in the opposite direction, i.e., finding a ranking probability that is consistent with (in the sense of (5)) an arbitrarily given binary choice vector, is far less straightforward[2]—this problem has been referred to as the characterization problem (see Marley, 1992). To begin with, not all $d(d-1)/2$-dimensional vectors confined to the cube $\mathscr{C}_{d(d-1)/2}$ are realizable by P_π via (5); this is because, among the $2^{\frac{d(d-1)}{2}}$ vertices of $\mathscr{C}_{d(d-1)/2}$, only $d!$ of those are "valid" ones, namely, ones that map one-to-one to the set of linear orders $\mathscr{L}_d$. The number of valid vertices represents only a small proportion; this fraction $d!/2^{\frac{d(d-1)}{2}}$ equals 0.75, 0.375, 0.117, 0.022 for $d = 3, 4, 5, 6$, respectively, and approaches zero rapidly as d increases. Any $d(d-1)/2$-dimensional vector which cannot be represented as a convex combination of valid vertices necessarily cannot be represented as a convex combination of those representing linear orders; therefore no ranking probability P_π exists that will give rise to such a binary choice vector via (5). Hence, the necessary (and sufficient) condition for the existence of P_π is that $\mathbf{a} = [a_{12}, a_{13}, \ldots, a_{(d-1)d}]^T$ can be expressed as a convex combination of the valid vertices of $\mathscr{C}_{d(d-1)/2}$. The convex hull of these $d!$ vertices, that is, the fractional mixture of the corresponding vectors, forms a geometric object contained within $\mathscr{C}_{d(d-1)/2}$. It is called the "Binary Choice Polytope" or "Linear Ordering Polytope",[3] which normatively describes the solution to the characterization problem of binary choice (Cohen & Falmagne, 1990, Fishburn, 1992, Koppen, 1991, 1995, Suck, 1992). See Appendix A for the mathematical background on polytopes.

The Binary Choice Polytope belongs to the general class of 0/1-polytopes (see Ziegler, 1999 for a recent introduction). Its facets (faces of maximal dimension) are defined by a system of inequalities/equalities on a_{ij}'s, the components of a binary choice vector $\mathbf{a}$. However, it turns out that the problem of finding a complete

[2] In Babington Smith's (1950) ranking model

$$P_\pi = \mathrm{const} \prod_{\{(i,j) \,:\, \pi(i) > \pi(j)\}} a_{ij},$$

one readily obtains a ranking probability from the binary choice probability, but it does not satisfy Eq. (5).

[3] Actually, the Binary Choice Polytope is defined in the space $R^{d(d-1)}$ (instead of $R^{d(d-1)/2}$) where the relations $a_{ij} + a_{ji} = 1$ are treated as one of the facet-defining equalities/inequalities.

description of these facets is a deceptively simple but challenging one, as not only all of its facet-defining inequalities are not yet found (and new inequalities are constantly emerging, including the more recent ones by Bolotashvili, Kovalev, & Girlich, 1999), but there is some indication that the effort of exhausting all such inequalities to obtain a complete linear description is perhaps futile because such a description may not itself be "described" following arguments in computation complexity theory (Pekeč, 2000).

1.2. Ranking probability and subset choice

An alternative to the binary choice paradigm is the "subset choice" paradigm (Brams & Fishburn, 1978, 1983, Falmagne & Regenwetter, 1996, Doignon & Regenwetter, 1997, Regenwetter, Marley, & Joe, 1998, Doignon & Regenwetter, 2002, Doignon & Fiorini, to appear. From the d-element master set $[d]$, consider the set formed by all subsets of $[d]$, i.e., $\mathscr{S}_d = \{S : S \subseteq [d]\}$. The set $\mathscr{S}_d$ is commonly referred to as the power set of the original master set $[d]$ and denoted as $2^{[d]}$; it has 2^d elements, including the null-set $\emptyset$ and $[d]$ itself. Under the subset choice paradigm, any element $S \in \mathscr{S}_d$ of this power set, i.e., any subset of $[d]$, becomes the basic choice primitive. A probability distribution defined on the 2^d alternative elements of $\mathscr{S}_d$, called the "subset choice probability", can be introduced in the same way that the ranking probability over the $d!$ elements of $\mathscr{L}_d$ has been introduced in (1)

$$P_S \geqslant 0, \quad \sum_{S \subseteq [d]} P_S = 1.$$

Subset choice paradigms are natural extensions to a popular voting mechanism introduced in the social choice and voting literature, namely approval voting (AV) (Brams & Fishburn, 1978, 1983). In recent years, the size-independent (SI) model of AV is proposed by Falmagne and Regenwetter (1996). The SI model envisions a two-stage process in voters' choices of subsets, a first stage concerning the size of the subset, followed by a second stage concerning the particular subset (among all subsets of equal size) being chosen. The subset choice probability P_S is then related to the ranking probability P_π through

$$P_S = f(|S|) \cdot \sum_{\{\pi \in \mathscr{L}_d \,:\, \{\pi^{-1}(d), \ldots, \pi^{-1}(d-|S|+1)\} = S\}} P_\pi. \tag{6}$$

Here, for a fixed S with set-size $|S| = k$, the summation is over all rankings π whose top k entries $\pi^{-1}(d), \ldots, \pi^{-1}(d-k+1)$ are exactly those elements of S; there are a total of $k!\,(d-k)!$ rankings for each S. The set-size probability $f(k)$, $k = 1, \ldots, d-1$ is given by

$$f(k) = \sum_{\{S \subseteq [d] \,:\, |S| = k\}} P_S, \tag{7}$$

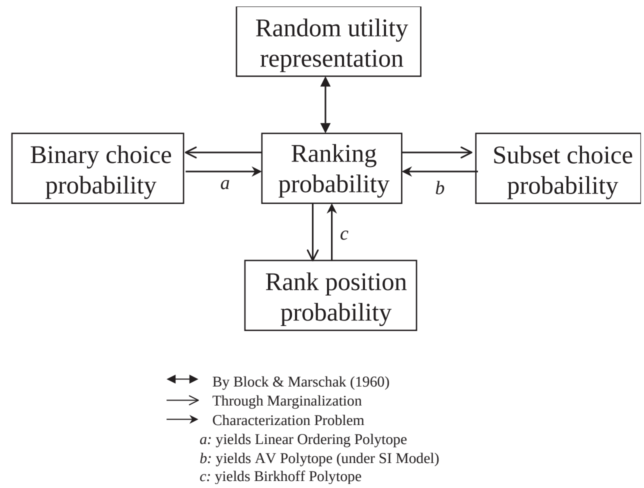

Fig. 1. Connections between common choice paradigms (ranking, binary choice, subset choice) and the random utility representation. Note the central role of ranking probability (the direction of an arrow indicates an inference or specification from one representation/paradigm to another).

with $f(0) = P_\emptyset$, $f(d) = P_{[d]}$ and that $\sum_{k=0}^{d} f(k) = 1$. Given P_π and $f(k)$, one readily obtains P_S. As with the binary choice case, the inference in the opposite direction is much harder. Given P_S (and hence $f(|S|)$), conditions for inducing a probability distribution P_π on all rankings in accordance with (6) have been called the characterization problem for AV. Its solution defines a polytope called the AV Polytope (Doignon & Regenwetter, 1997), and all of its facets have been characterized recently (Doignon & Fiorini, to appear).

1.3. Imposing structures on linear orders

The above discussion shows that the probability distribution on linear orders P_π plays a pivotal role in formulating the characterization problems for binary choice and for subset choice paradigms. This is schematically summarized as Fig 1. The conditions for the existence of P_π under (5) for a binary choice vector, or under (6) for a subset choice probability amount to linear descriptions of the Binary Choice Polytope or the AV Polytope, respectively. These polytopes are, unfortunately, quite complex. One hopes that further restrictions on how to combine and compare linear orders may give rise to a simpler geometric–combinatoric structure.

There have been at least two approaches in the literature for endowing structures on linear orders (which turn out to be much related). The first approach relies on the introduction of a metric on $\mathscr{L}_d$ to describe the difference between two linear orders. Familiar metrics include Spearman's D_ρ (arising from ρ, the correlation coefficient of differences in ranks), Kendall's D_τ (arising from τ, the correlation coefficient of signs of rank comparisons), Cayley's distance (minimum number of transpositions between two rankings), among many others (see Kendall, 1962). These metrics are then used for parametric models of ranking probability whereby P_π is inversely related to the distance $D(\pi, \pi_0)$ from an arbitrary ranking π to a modal ranking π_0 (Fligner & Verducci, 1986). A prominent metric-based ranking model is Mallows' (1957) ϕ model, an exponential family with the assumption of unimodality in the distribution of rankings. When endowed with the Spearman's ρ for measuring the distance between two rankings π_1 and π_2

$$D_\rho(\pi_1, \pi_2) = \sum_{i=1}^{d} (\pi_1(i) - \pi_2(i))^2, \tag{8}$$

Mallow's ϕ model yields a particular probability distribution on the sphere S^{d-2}, known as the von-Mises Fisher distribution in the modelling of directional data (McCullagh, 1993). Other types of distance-based models may incorporate the dynamics of the ranking process, such as a multi-stage process where a voter chooses the most preferred candidate first, and then the next preferred one from the remaining pool, and so forth, to produce a linear order (see Critchlow, Fligner, & Verducci, 1991).

An alternative approach to endow structures on $\mathscr{L}_d$ is based on the notion of a "scoring function" (Young, 1975) in the voting literature. Given P_π (called a "voters profile" there for obvious reasons), the idea is to construct a score for each candidate $i \in [d]$ to reflect the collective or aggregated preference of the voting population. Each candidate's score thus obtained will be compared against others' for their desirability. One such scoring method is the well-acclaimed "Borda rule", namely, candidate i accrues some point(s) (or "marks") equal to[4] the rank-position $\pi(i)$ in a particular ranking π; the total points v_i accrued by each candidate from the voting population is then properly weighted by the voters profile P_π to yield an "order of merit". Formally, the Borda score of a ranking probability P_π is defined as a d-dimensional vector $\mathbf{v}^{\mathrm{Bd}}$:

$$\mathbf{v}^{\mathrm{Bd}} = \begin{bmatrix} \sum_{\pi \in \mathscr{L}_d} P_\pi \cdot \pi(1) \\ \sum_{\pi \in \mathscr{L}_d} P_\pi \cdot \pi(2) \\ \vdots \\ \sum_{\pi \in \mathscr{L}_d} P_\pi \cdot \pi(d) \end{bmatrix}, \tag{9}$$

i.e.,

$$\mathbf{v}^{\mathrm{Bd}} = \sum_{\pi \in \mathscr{L}_d} P_\pi \begin{bmatrix} \pi(1) \\ \pi(2) \\ \vdots \\ \pi(d) \end{bmatrix}. \tag{10}$$

The notions of the Borda rule and the Borda score are well-known in the social choice literature; in fact they have been systematically axiomatized (Young, 1974, 1975, Nitzan & Rubinstein, 1981). Among its axioms (Young, 1974) is a cancellation condition effectively stating that two rankings π_1 and π_2 will cancel each other's effect if one is the "reverse" of the other, defined as $\pi_1(i) = d - \pi_2(i)$ for any $i \in [d]$. The cancellation in the net consequence of reverse rankings is a crucial axiom; it is closely related to the construction of "net ranking probability" (Regenwetter & Grofman, 1998a).

The above two approaches of imposing structures on linear orders, one involving a metric on rankings and the other involving a scoring function on ranking probabilities, turn out to be closely related; indeed they often parallel one another. Of special interest here is a result by Cook and Seiford (1982)[5] showing that the Borda vector $\mathbf{v}^{\mathrm{Bd}}$ in (10) actually achieves the minimum over the distance from a point $\mathbf{v} = [v_1, v_2, \ldots, v_d]^T \in R^d$ to all rank vectors when weighted by P_π and using Spearman's ρ metric (8)

$$D_\rho(\mathbf{v}, \pi) = \sum_{\pi \in \mathscr{L}_d} \sum_{i=1}^{d} P_\pi (v_i - \pi(i))^2. \tag{11}$$

The Borda vector is thus construed as an "average" rank vector, in the least-mean-square sense and consistent with the generalized average on ordered sets (Ovchinnikov, 1996), obtained from the d rankings associated with a voters profile P_π. Note that the geometries arising out of Borda scores, and more generally the position voting schemes, were discussed in Saari (1990, 1992, 1993); however, these papers fell short of identifying the space of all Borda vectors as forming a well-defined polytope under investigation here.

As we shall see, construction (10) of the Borda score (vector) gives rise to a combinatoric–geometric object known as the *permutahedron* (alternatively spelt as *permutohedron*). The permutahedron is a very simple and special kind of polytope that has been well studied by mathematicians, see the recent treatise by Ziegler

[4] Given a linear order π, the points assigned to a candidate i, in the common Borda rule, is the number of other candidates that are ordered below the focal candidate i. The points assigned to a candidate with rank-position $\pi(i)$ is thus $\pi(i) - 1$, and ranges between 0 (i is least preferred) to $d - 1$ (i is most preferred). Here, for convenience, we add 1 to this mark, so the Borda score ranges between 1 (least preferred) to d (most preferred).

[5] These authors apparently mis-attributed Spearman's ρ to Kendall in this and a number of subsequent publications. It is clear from Kendall (1962) that the metric associated with his τ is

$$D_\tau(\pi_1, \pi_2) = \sum_{\text{all } i<j} 1 - (\operatorname{sgn}(\pi_1(i) - \pi_1(j)) \cdot \operatorname{sgn}(\pi_2(i) - \pi_2(j)))$$

in which $\operatorname{sgn}(t)$ denotes the sign of t, whereas the metric associated with Spearman's ρ is given by (8).

(1995) and its excellent review by Suck (1997). As its vertices represent the set of permutations of d objects, a permutahedron is a kind of permutation polyhedra (Bowman, 1972) that geometrically represent permutations. It can be realized in various ways, as projections or lift-ups of projections of other polytopes. Its nice geometric properties allow a unified treatment of various choice paradigms (binary choice, subset choice, etc.) and the related RU theory.

Although permutahedron is a familiar object to combinatorial mathematicians, it has so far rarely been used in models of choice (binary, subset, RU, Borda scoring) where the polytope theory has demonstrated to be relevant. In particular, the potential power of permutahedron in connecting the aforementioned choice paradigms has not been systematically investigated. The paper provides a comprehensive review of some well-known mathematical properties of a permutahedron, points out their meanings within the context of, as well as their applications to various choice paradigms and models, along with a few new results. The goal of this paper is to bring to the awareness of mathematical psychologists such a geometric-combinatorial object and the intuitions it provides. The main new results include the demonstration of: (i) the surjective mapping of the space of binary choice vectors to the permutahedron, the space of Borda scores (Theorem 2.3); though this was presented and proven as an exercise in Ziegler's book, the projection will be explicated in a matrix form allowing direct appreciation of the connection between the permutahedron and *any* binary choice vector (not only the ones that lie within the Binary Choice Polytope); (ii) the monotonicity and homogenizing property associated with this projective map if one further assumes BTL representation of the choice probability (Corollary 2.4); (iii) the embedding of the permutahedron in a hypersphere (Proposition 2.5); this enables a geometric description of the RU representation of ranking probabilities and the Block and Marschak (1960) condition; (iv) the connection between the Brams–Fishburn score of the AV and the Borda score under the equal-probability interpretation of subset choice (Theorem 2.7) and the latent voters profile consistent with this interpretation (Corollary 2.8); and (v) the core that defines the AV Polytope under the SI model of the underlying preference structure (Theorem 2.9).

The remaining of the paper is organized as follows: Section 2.1 introduces the formal definition of permutahedron, both as a convex combination of permutation vectors and as an intersection of half-planes with a system of inequalities defining its facets, the dual definitions of a convex polytope that are of fundamental importance in the polytope theory. We then elaborate various alternative constructions of a permutahedron, i.e., as a zonotope (which is a projection of the cube) in Section 2.2, as a geometric object circumscribing a hypersphere in Section 2.3, as a monotone path polytope (which arises from the lift-up of the projective map from the cube onto a line segment) in Section 2.4, and through the canonical projection from the Birkhoff Polytope in Section 2.5. Each of these constructions provides a natural link to the binary choice paradigm, the RU representation, the subset choice paradigm, and the rank-assignment paradigm, respectively. Section 3 summarizes the central role of permutahedron in connecting these choice paradigms in view of its rich geometric–combinatorial properties, and discusses a connection to the topological approach (Chichilnisky & Heal, 1983) to preference and choice.

To help readers who might be unfamiliar with the polytope theory, a review of the mathematical background is provided as Appendix A. This includes dualistic characterizations of a polytope either by its vertices or by its facets, the face lattice of polytopes, the projection of polytopes (including the projection of cubes), and the fiber polytope as lift-ups of polytope projection (including the monotone path polytope). These materials provide the necessary mathematical knowledge that our current exposition draws upon.

2. Choice paradigms and the permutahedron

This section will review various choice paradigms and discuss their inter-connections under the common framework of the permutahedron. Special attention will be paid to the geometric and the combinatorial view of the permutahedron, the various realizations of a permutahedron as they relate to the various choice paradigms, and the projections as well as the lift-ups of projections as the inter-connections among the choice paradigms.

First, from its construction in Section 2.1, a permutahedron over d objects, or d-permutahedron Π_{d-1} for short,[6] will be shown to be a natural candidate for representing the probability distribution P_π of linear orders π over a d-element set $[d]$, with each of the $d\,!$ vertices of Π_{d-1} representing a possible ranking $\pi \in \mathscr{L}_d$. Since a point within Π_{d-1} represents the Borda score associated with P_π, the various geometric properties of a permutahedron reviewed here will naturally link the ranking probability and its Borda score to other choice paradigms, including the binary choice paradigm, the subset choice paradigm, and the rank-assignment paradigm, and to other representations of choice

[6] Strictly speaking, one should refer to Π_{d-1} as a $(d-1)$-permutahedron, since the prefix usually refers to the dimensionality of a polytope in the polytope literature. In this paper, we call it a d-permutahedron to emphasize the fact that it arises as permutations over d objects. We retain the notation of Π_{d-1} to indicate that it is a $(d-1)$-dimensional object.

probability, including the BTL representation and the RU representation. Then, in Section 2.2, the d-permutahedron Π_{d-1} will be realized as the affine projection of a $d(d-1)/2$-dimensional cube $\mathscr{C}_{d(d-1)/2}$; the latter is the space that defines binary choice probabilities. Explicating the projection matrix yields a formula, Young's formula, to associate an induced score for *any* binary choice vector (probability), not necessarily the ones that conform to a given ranking probability through (5), i.e., the ones that lie within the Binary Choice Polytope. It will be shown that this induced score is consistent with the representation of the binary choice probability, along with certain desired properties. Next, in Section 2.3, the d-permutahedron Π_{d-1}, as a $(d-1)$-dimensional object, is shown to be circumscribed by a unit sphere S^{d-2}; this property will be used to construct a geometric representation of the RU model and to derive its relation to ranking probabilities (including the Block–Marschak condition). There, it will be shown that any interval-scale vector (a vector whose components are subject to the affine freedom in the choice of a reference zero and a scaling factor) can be mapped to a point in the $(d-2)$-dimensional unit sphere S^{d-2} osculating Π_{d-1}. Next, in Section 2.4, Π_{d-1} will be realized as a special fiber polytope, namely the monotone path polytope, arising from projecting a d-cube $\mathscr{C}_d \subset R^d$ to a line segment $[0,d] \subset R^1$; this property will be shown to relate subset choice to ranking. There, all subsets of the d-element set $[d]$ are naturally mapped to the 2^d vertices of $\mathscr{C}_d$, and AV under the Brams–Fishburn tally procedure is shown to amount to an equal-probabilistic assignment of all linear orders that are compatible with the chosen subset. The SI model of approval voting (Falmagne & Regenwetter, 1996) is also analyzed using this framework. Finally, in Section 2.5, Π_{d-1} will be related to the rank-position probability that arises from rank-matching paradigm. That a Birkhoff Polytope B_d for $d \times d$ bistochastic matrices has a canonical projection to Π_{d-1} will be utilized to derive an induced score for the d candidates.

2.1. Permutahedron

2.1.1. Construction of a permutahedron

Permutahedron is a convex polytope associated with permutations on a set of given objects labelled by natural numbers $[d] = \{1, \dots, d\}$. To fix the notation, let $\pi = \langle \pi(1)\pi(2)\cdots\pi(d) \rangle$ denote as a mapping from $[d]$ as the set of candidates to $[d]$ as the set of ranks, with $\pi(i)$ denoting the rank (also called "rank-position") associated with candidate i, and $\pi^{-1}(j)$ denoting the candidate occupying the rank j. Our convention is that higher ranks correspond to larger integer values, so $\pi = \langle 4132 \rangle$ means that candidate "1" has the highest rank or is most preferred ($\pi(1) = 4$), that candidate "2" has lowest rank or is least preferred ($\pi(2) = 1$), etc. For d candidates, there are a total of $d\,! = d \cdot (d-1) \cdots 2 \cdot 1$ different permutations, denoted as $\pi_k \in \mathscr{L}_d$ ($k = 1, 2, \dots, d!$), where $\mathscr{L}_d$ is the set of all permutations over d elements.

To understand how a permutahedron arises, introduce the notion of a *permutation vector*, namely, a d-dimensional vector whose components are rank-positions of a linear order π:

$$\mathbf{x}_\pi = [\pi(1), \pi(2), \dots, \pi(d)]^T, \tag{12}$$

with T denoting vector transpose. For instance, when $\pi = \langle 4132 \rangle$, the corresponding permutation vector is $\mathbf{x}_\pi = [4, 1, 3, 2]^T$. There are $d!$ permutation vectors $\mathbf{x}_{\pi_k} \in R^d$, $k = 1, \dots, d!$, all defined in R^d.

A d-permutahedron, denoted Π_{d-1},[7] is the convex hull of all of the $d!$ permutation vectors $\mathbf{P} = \{\mathbf{x}_{\pi_1}, \mathbf{x}_{\pi_2}, \dots, \mathbf{x}_{\pi_{d!}}\}$

$$\Pi_{d-1} = \mathrm{conv}(\mathbf{P}) = \left\{ \lambda_1 \mathbf{x}_{\pi_1} + \lambda_2 \mathbf{x}_{\pi_2} + \cdots + \lambda_{d!} \mathbf{x}_{\pi_{d!}} : \pi_k \in \mathscr{L}_d,\ \lambda_k \geqslant 0;\ \sum_{k=1}^{d!} \lambda_k = 1 \right\}. \tag{13}$$

Its barycenter $\mathbf{b}$, a d-dimensional vector, is obtained by setting $\lambda_1 = \lambda_2 = \cdots = \lambda_{d!} = 1/d\,!$ in (13):

$$\mathbf{b} = \sum_{k=1}^{d!} \frac{1}{d\,!} \mathbf{x}_{\pi_k} = \frac{(d-1)!(1+2+\cdots+d)}{d\,!} \mathbf{1} = \frac{d+1}{2} \mathbf{1},$$

where $\mathbf{1} = [1, 1, \dots, 1]^T$.

This construction of a permutahedron is closely related to the calculation of Borda scores in the voting and the social choice literature (cf. Section 1.3). There, the collection of λ_k (denoted P_π) is rightfully called the *voters profile*, and formula (10) for calculating Borda scores $\mathbf{v}^{\mathrm{Bd}}$ gives exactly the same form of (13), once the definition of a permutation vector (12) is understood. Though not widely picked up by mathematical psychologist until recently, permutahedra have been explicitly applied to consensus ranking (Cook & Seiford, 1982, 1990) and to the description of statistical ranking models in general (McCullagh, 1993).

The permutahedron Π_{d-1} characterizes concisely both combinatorial and geometric properties of permutations over d objects, apparently first investigated by Schoute (1911). The actual shapes of some low-dimensional permutahedra are graphically illustrated in Fig. 2 for $d = 3, 4$. Instead of using the permutation vectors $[\pi(1), \pi(2), \dots, \pi(d)]^T$, define the π-corresponding vertex as $\mathbf{x}_\pi = [c_{\pi(1)}, c_{\pi(2)}, \dots, c_{\pi(d)}]^T$ for any base vector $[c_1, c_2, \dots, c_d]^T$ (for later convenience $c_1 > c_2 > \cdots > c_d$ is

[7] The reason that a d-permutahedron is subscripted $d-1$ rather than d will become clear in Proposition 2.5. See also footnote 6.

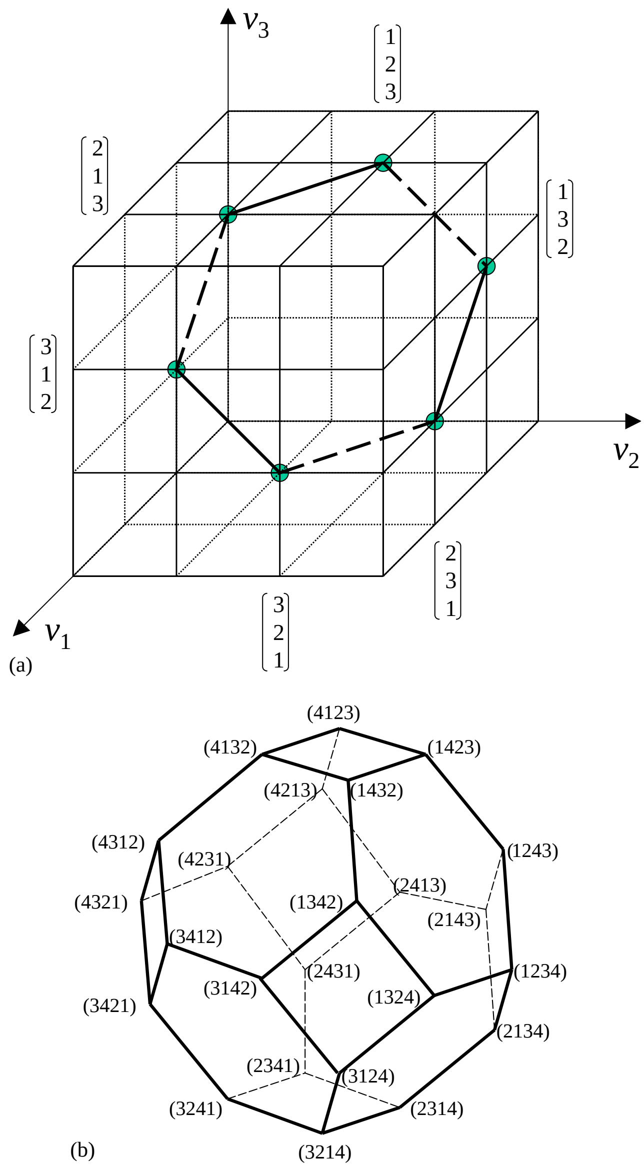

Fig. 2. Permutahedra in two and three dimensions, in the case of (a) $d = 3$ (note its construction as the convex hull of $[1,2,3]^T, [1,3,2]^T, \ldots, [3,2,1]^T$); and in the case of (b) $d = 4$. Note that in (b), consistent with our convention of interpreting $\pi(i) = j$ as "item i's rank-position is j", the vertices are labelled as ordering $(\pi^{-1}(1)\pi^{-1}(2)\pi^{-1}(3)\pi^{-1}(4))$, such that the neighboring vertices differ only in the transposition of adjacent ranks between two objects (for instance the exchange of the lowest ranks between the first and the third object $(1342) \leftrightarrow (3142)$). This ensures that the permutation vectors associated with adjacent vertices differ with only minimum distance. This is the labelling scheme adopted in Ziegler (1995, p. 18) and in Yemelichev (1984, p. 227). Note that the permutation vector associated with, say (1342), is $[1,4,2,3]^T$.

required without loss of generality). The resulting geometric object

$$\mathrm{conv}\{[c_{\pi(1)}, c_{\pi(2)}, \ldots, c_{\pi(d)}]^T : \pi \in \mathscr{L}_d\} = \sum_{k=1}^{d!} \lambda_k \begin{bmatrix} c_{\pi_k(1)} \\ c_{\pi_k(2)} \\ \vdots \\ c_{\pi_k(d)} \end{bmatrix}$$

is called a *permutation polytope*, which generalizes permutahedron (13) in a trivial way. Such mathematical generalization is in obvious correspondence with the general position methods for rank aggregation (Young, 1975, Cook & Kress, 1992). Another related combinatoric–geometric object is the permuto-associahedron constructed for signed, bracketed permutations, see Ziegler (1995) for further details.

2.1.2. Facet-defining inequalities of a permutahedron

Permutahedra are known to have relatively simple configurations in space. They are simple polytopes. In fact, although a d-permutahedron has $d!$ vertices defined in R^d, it is actually a geometric object of dimension $d-1$ (a proof of this observation will be given in Proposition 2.5 in the next subsection). For this reason, the d-permutahedron is denoted as Π_{d-1}.

As an alternative to defining a permutahedron as the convex combination of permutation vectors, we may also define it as the bounded intersection of half-spaces given by a system of hyperplanes, in accordance with the basic, dualistic view of a convex polytope (see Appendix A). As any facet of a permutahedron is passed through by one and exactly one of the hyperplanes (which assumes an equality sign in the system of facet-defining inequalities/equalities), the geometric arrangement of these hyperplanes in space defines the same Π_{d-1}. Fortunately, this system of facet-defining inequalities/equalities has been worked out and is to be reproduced below—they followed a strong result on vector majorization by the combinatoric mathematician Richard Rado in a paper published in 1952, and were given in its explicit form in a book on polytope and optimization by Yemelichev, Kovalev, and Kravtsov (1984). The facet-defining equality/inequalities were also independently derived by Gaiha and Gupta (1977).

Theorem 2.1 (Rado, 1952; Gaiha & Gupta, 1977, Theorem 2; Yemelichev, Kovalev, & Kravtsov, 1984). *The d-permutahedron Π_{d-1} is given by the following system of constraints on its coordinates $\mathbf{v} = [v_1, v_2, \ldots, v_d]^T$*

$$\sum_{i \in M} v_i \leqslant \frac{|M|(2d - |M| + 1)}{2} \quad \textit{for all } M \subset [d], \tag{14}$$

$$\sum_{i \in [d]} v_i = \frac{d(d+1)}{2}, \tag{15}$$

where M is a non-trivial (proper and non-empty) subset of $[d]$ having integer set-size $|M|$.

Proof. See Yemelichev et al. (1984, pp. 228–230), using Rado's result regarding necessary and sufficient conditions for majorization of one vector by another. An alternative proof is given in Gaiha and Gupta (1977), using a well-known inequality from Hardy, Litteg, and

Polya (1934). A third proof is given in Ziegler (1995, p. 304). □

Remark 1. Recall the generalization of a permutahedron to a permutation polytope, in which the base vector becomes $[c_1, c_2, \ldots, c_d]^T$ (with $c_1 > c_2 > \cdots > c_d$). The corresponding system of facet-defining inequalities/equalities are

$$\sum_{i \in M} v_i \leqslant \sum_{i=1}^{|M|} c_i \quad \text{for all } M \subset [d],$$

$$\sum_{i=1}^{d} v_i = \sum_{i=1}^{d} c_i.$$

The case for the d-permutahedron is obtained by simply setting $c_i = d - i + 1$.

Remark 2. The proof of this proposition, as well as Corollary 2.2 below involves the notion of vector majorization and the Birkhoff/von Neumann Theorem on bistochastic matrices (see Section 2.5). Ziegler's (1995) proof was based on the notion of fiber polytopes. These proofs are omitted here due to their high technicality.

Remark 3. Because of equality (15), the system of inequalities (14) can be equivalently cast as

$$\sum_{i \in M} v_i \geqslant \frac{|M|(|M|+1)}{2} \quad \text{for all } M \subset [d].$$

Since a permutahedron can be described as the intersection of a collection of half-spaces (each given by a facet-defining hyperplane), its individual faces are characterized by letting certain inequality to assume equal sign. Yemelichev et al. (1984) provided a particularly powerful result in elucidating all faces of a permutahedron, as reproduced below without proof.

Corollary 2.2 (Yemelichev et al., 1984, Theorem 3.4). *A set of solutions of system* (14), (15) *is a j-face* ($0 \leqslant j \leqslant d-2$) *of the d-permutahedron if and only if for each such solution, inequalities* (14) *are satisfied as equalities only for a nested sequence of subsets* $M_1, M_2, \ldots, M_{d-j-1}$, *where*

$$M_1 \subset M_2 \subset \cdots \subset M_{d-j-1} \subset \{1, 2, \ldots, d\}.$$

In particular, each of the hyperplanes that define the boundary of the permutahedron (i.e. that pass through the facets) "converts" only one of the inequalities in (14) *into an equality (since* $d-j-1$ *equals 1 for* $j = d-2$*). Each of the* $2^d - 2$ *ways of selecting a non-trivial subset* $M \subset [d]$ *corresponds to a distinct facet.*

Proof. See Yemelichev et al. (1984, p. 231). □

Remark. The faces of Π_{d-1} have interesting properties: each of its j-dimensional faces or *j-face* ($j = 0, 1, \ldots, d-2$) corresponds to an ordered partition of the set $\{1, 2, \ldots, d\}$ into $d - j$ non-empty parts. Thus vertices ($j = 0$) are permutations, and edges ($j = 1$) connect pairs of rankings that differ only by a single transposition of adjacent rank-positions for two elements of $[d]$ (cf., Fig. 2). It can be shown that every vertex of Π_{d-1} belongs exactly to $d-1$ edges (Gaiha & Gupta, 1977). Each facet ($j = d-2$), on the other hand, is a partition of the set $[d]$ into two subsets with non-overlapping elements $(S, [d] \backslash S)$, where S is any non-empty and proper subset of $[d]$. The number of j-faces of a permutahedron $f_j(\Pi_{d-1})$ has the expression

$$f_j = \sum_{\substack{\{i_1, i_2, \ldots, i_{d-j} \\ i_1 + i_2 + \cdots + i_{d-j} = d\}}} \frac{d!}{i_1! i_2! \cdots i_{d-j}!} \tag{16}$$

(the sum is taken across all positive integral solutions of $i_1 + i_2 + \cdots + i_{d-j} = d$). The number of facets can be calculated as $2^d - 2$.

Besides being characterized as the convex hull of permutation vectors (13), or as the intersection of a system of half-spaces (14) and (15), a permutahedron can also arise as a projection, or a lift-up of a certain projection, between polytopes. More specifically, a d-permutahedron can be realized as a zonotope, the projection of a unit cube of dimension $d(d-1)/2$, as a monotone path polytope associated with projecting a cube of dimension d onto a line segment $[0, d]$, or as the canonical projection of the Birkhoff Polytope which is associated with $d \times d$ bistochastic matrices; see below.

2.2. Connection to binary choice vectors via zonotope projection

2.2.1. Projection of a binary choice vector

Permutahedron is a special kind of polytope, namely a zonotope, defined as the image of a cube under affine projection. In other words, a permutahedron can be realized as the projection of a cube of an appropriate (and higher) dimension. This observation was made in Ziegler (1995, p. 200), along with a sketch of proof. Since this fact plays a major role in linking binary choice probabilities and ranking probabilities, such zonotope projection is explicated here, along with a constructive proof. A graphic illustration is shown as Fig. 3, for $d = 3$.

Theorem 2.3. *The d-permutahedron* Π_{d-1} *is realizable as an affine projection of a cube* $\mathscr{C}_{d(d-1)/2}$, *the domain on which the binary choice probability vector*

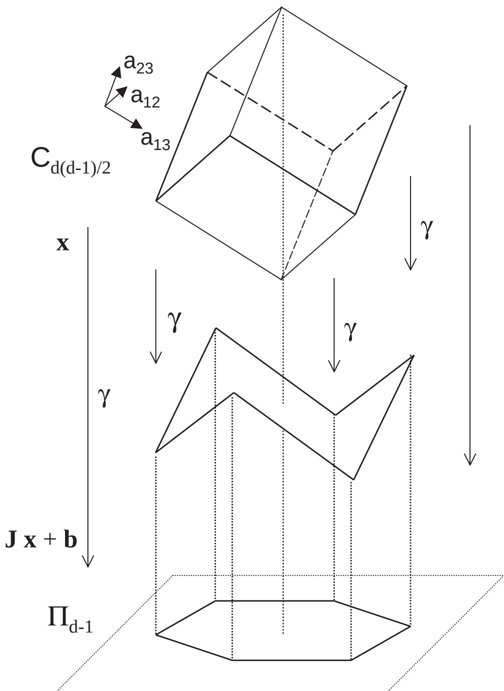

Fig. 3. The d-permutahedron as a zonotope, i.e., resulting from a suitably chosen projection of a cube of dimension $d(d-1)/2$ (here $d=3$). Note that projecting to the center of the hexagon (the 3-permutahedron) are two vertices $[a_{12}, a_{13}, a_{23}]^T = [1, 0, 1]^T$ and $[a_{12}, a_{13}, a_{23}]^T = [0, 1, 0]^T$, which are "invalid" vertices that lie outside of the Binary Choice Polytope for $d=3$ (cf. Section 1.1).

$\mathbf{a} = [a_{11}, a_{12}, \ldots, a_{(d-1)d}]^T$ *is defined. In other words, there exists a surjective map*: $\mathscr{C}_{d(d-1)/2} \ni \mathbf{a} \mapsto \mathbf{v} = [v_1, v_2, \ldots, v_d]^T \in \Pi_{d-1}$ *given by*

$$\begin{cases} v_1 = a_{11} + a_{12} + \cdots + a_{1d} + \dfrac{1}{2} = \sum\limits_{j=1, j\neq 1}^{d} a_{1j} + 1 \\ v_2 = a_{21} + a_{22} + \cdots + a_{2d} + \dfrac{1}{2} = \sum\limits_{j=1, j\neq 2^2}^{d} a_{2j} + 1 \\ \quad\cdots\cdots \\ v_d = a_{d1} + a_{d2} + \cdots + a_{dd} + \dfrac{1}{2} = \sum\limits_{j=1, j\neq d}^{d} a_{dj} + 1 \end{cases}. \tag{17}$$

Proof. Recall (see Appendix A) that a p-dimensional unit cube $\mathscr{C}_p$, centered at origin, can be expressed as $\mathbf{x} = \sum_{l=1}^{p} t_l\, \boldsymbol{\varepsilon}_l$ with $\boldsymbol{\varepsilon}_l$'s as base vectors and $-\frac{1}{2} \leqslant t_l \leqslant \frac{1}{2}$ as coordinates $(l = 1, 2, \ldots, p)$, see (A.1). Its projection $\mathbf{x} \mapsto \mathbf{z} = \mathbf{J}\mathbf{x} + \mathbf{b}$ into a lower-dimensional space $R^q \ni \mathbf{z}$ $(q < p)$ can be expressed as $\mathbf{z} = \sum_{l=1}^{p} t_l \mathbf{j}_l + \mathbf{b}$ where $\mathbf{j}_l$ is the lth column vector of the $q \times p$ matrix $\mathbf{J}$, and $\mathbf{b}$ is the image of the p-cube's center, see (A.2). In the present case, $p = d(d-1)/2$ and t_l represents the pairwise binary choice probability $t_l = a_{ij} - \frac{1}{2}$, $0 \leqslant a_{ij} \leqslant 1$. For convenience, the ij subscript $(i = 1, 2, \ldots, d;\ j = i+1, i+2, \ldots, d)$ is used in place of l $(l = 1, 2, \ldots, d(d-1)/2)$, so long as it is understood that $\sum_{l=1}^{d(d-1)/2} = \sum_{i=1}^{d} \sum_{j=i+1}^{d}$, with

$$l = (i-1)d - \frac{i(i+1)}{2} + j. \tag{18}$$

To ensure that the center of the cube is properly mapped to the center of the d-permutahedron, we choose $\mathbf{b} = \frac{d+1}{2}\mathbf{1} = \frac{d+1}{2}[1, 1, \ldots, 1]^T$.

We now proceed to construct the particular matrix $\mathbf{J} = [\mathbf{j}_1, \mathbf{j}_2, \ldots, \mathbf{j}_{d(d-1)/2}]$ by setting

$$\mathbf{j}_l = \mathbf{e}_i - \mathbf{e}_j, \quad i = 1, 2, \ldots, d,\ \ j = i+1, i+2, \ldots, d,$$

where $\mathbf{e}_i = [\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{d-i}]^T$ is just a base vector of R^d (with 1 in its ith coordinate and 0 elsewhere). More explicitly, the (k,l)th entry of the matrix $\mathbf{J}$ is $\delta_{ik} - \delta_{jk}$, with l linked to i,j through (18), and the Kronecker δ-function given by

$$\delta_{ki} = \begin{cases} 1 & \text{when } k = i, \\ 0 & \text{otherwise.} \end{cases}$$

So the (k,l)th entry of the projection matrix is 1 (if $k=i$), -1 (if $k=j$), or 0 (otherwise). When we line up the components of the binary choice vector (a column vector) in the following order $\mathbf{a} = [a_{12}, a_{13}, \ldots, a_{1d}, a_{23}, \ldots, a_{2d}, \ldots, a_{(d-1)d}]^T$, the projection matrix has the following explicit form

$$\mathbf{J} = \left[\begin{array}{ccccccccccccc} 1 & 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 \\ -1 & 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 & -1 & \cdots & 0 & \cdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & 1 & 1 & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 & 0 & \vdots & \cdots & 0 & \cdots & -1 & 0 & 1 \\ 0 & 0 & 0 & \cdots & -1 & 0 & 0 & \cdots & -1 & \cdots & 0 & -1 & -1 \\ \multicolumn{5}{c}{\underbrace{\hspace{8em}}_{d-1}} & \multicolumn{4}{c}{\underbrace{\hspace{6em}}_{d-2}} & \cdots & \multicolumn{2}{c}{\underbrace{\hspace{3em}}_{2}} & \underbrace{\hspace{1em}}_{1} \end{array}\right]. \tag{19}$$

With this $\mathbf{J}$, the resulting zonotope is, according to (41),

$$\mathbf{z} = \frac{d+1}{2}\mathbf{1} + \sum_{i=1}^{d} \sum_{j=i+1}^{d} \left(a_{ij} - \frac{1}{2}\right)(\mathbf{e}_i - \mathbf{e}_j).$$

The terms for summation evaluate to

$$\begin{aligned}
&\sum_{i=1}^{d} \sum_{j=i+1}^{d} \left(a_{ij} - \frac{1}{2}\right)\mathbf{e}_i - \sum_{i=1}^{d} \sum_{j=i+1}^{d} \left(a_{ij} - \frac{1}{2}\right)\mathbf{e}_j \\
&\quad = \sum_{i=1}^{d} \sum_{j=i+1}^{d} \left(a_{ij} - \frac{1}{2}\right)\mathbf{e}_i \\
&\quad - \sum_{j=1}^{d} \sum_{i=j+1}^{d} \left(a_{ji} - \frac{1}{2}\right)\mathbf{e}_i \quad \text{(exchange of indices } i,j) \\
&\quad = \sum_{i=1}^{d} \sum_{j=i+1}^{d} \left(a_{ij} - \frac{1}{2}\right)\mathbf{e}_i \\
&\quad + \sum_{i=1}^{d} \sum_{j=1}^{i-1} \left(\frac{1}{2} - a_{ji}\right)\mathbf{e}_i \quad \left(\sum_{j=1}^{d} \sum_{i=j+1}^{d} = \sum_{i=1}^{d} \sum_{j=1}^{i-1}\right) \\
&\quad = \sum_{i=1}^{d} \sum_{j=i+1}^{d} \left(a_{ij} - \frac{1}{2}\right)\mathbf{e}_i \\
&\quad + \sum_{i=1}^{d} \sum_{j=1}^{i-1} \left(a_{ij} - \frac{1}{2}\right)\mathbf{e}_i \quad \left(a_{ij} - \frac{1}{2} = \frac{1}{2} - a_{ji}\right) \\
&\quad = \sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} \left(a_{ij} - \frac{1}{2}\right)\mathbf{e}_i \\
&\quad = \sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} a_{ij}\, \mathbf{e}_i - \frac{d-1}{2}\mathbf{1} \quad \left(\sum_{i=1}^{d} \sum_{j=1, j\neq i}^{d} \mathbf{e}_i \right.\\
&\qquad \left. = \sum_{i=1}^{d} (d-1)\mathbf{e}_i = (d-1)\mathbf{1}\right).
\end{aligned}$$

Therefore,

$$\mathbf{z} = \sum_{i=1}^{d} \left(\sum_{j=1, j\neq i}^{d} a_{ij} \right)\mathbf{e}_i + \mathbf{1} = \sum_{i=1}^{d} \left(\sum_{j=1, j\neq i}^{d} a_{ij} + 1 \right)\mathbf{e}_i.$$

To verify that the zonotope described by the above $\mathbf{z}$ vector is indeed the d-permutahedron, we proceed to show that its components, hereafter denoted v_i with an abuse of notation

$$v_i = 1 + \sum_{j=1, j\neq i}^{d} a_{ij} = \frac{1}{2} + \sum_{j=1}^{d} a_{ij},$$

in fact satisfy the system of constraints defining a permutahedron. First, consider equality (15):

$$\begin{aligned}
\sum_{i=1}^{d} v_i &= \sum_{i=1}^{d} \left(1 + \sum_{j=1, j\neq i}^{d} a_{ij}\right) \\
&= d + \frac{d(d-1)}{2} \quad (\text{since } a_{ij} + a_{ji} = 1) \\
&= \frac{d(d+1)}{2},
\end{aligned}$$

which is the right-hand side of (15). Second, for any subset $M \subset \{1, 2, \ldots, d\}$, denote its complementary as $\bar{M}$, and the number of elements in M and $\bar{M}$ as $|M|$ and $d - |M|$,

$$\begin{aligned}
\sum_{i\in M} v_i &= \sum_{i\in M} \left(1 + \sum_{j=1, j\neq i}^{d} a_{ij}\right) \\
&= |M| + \sum_{i\in M} \left(\sum_{j\in M, j\neq i} a_{ij} + \sum_{j\in \bar{M}} a_{ij} \right) \\
&= |M| + \frac{|M|(|M|-1)}{2} + \sum_{i\in M} \sum_{j\in \bar{M}} a_{ij} \\
&\leqslant |M| + \frac{|M|(|M|-1)}{2} + |M|(d - |M|) \\
&\quad \text{since } (a_{ij} \leqslant 1 \text{ and } |\bar{M}| = d - |M|) \\
&= \frac{|M|(2d - |M| + 1)}{2},
\end{aligned}$$

which is the right-hand side of (14). Third, we show that for any $\pi \in \mathscr{L}_d$, there is a one-to-one mapping between a π-representing vertex of Π_{d-1} and the corresponding vertex of $\mathscr{C}_{d(d-1)/2}$ consistent with π. For ranking π, the inequality $\pi(i) > \pi(j)$ means that candidate i ranks higher than candidate j according to π. The particular vertex of $\mathscr{C}_{d(d-1)/2}$ consistent with this π is represented by the binary choice vector $\boldsymbol{a}^\pi$ whose components are

$$a_{ij}^\pi = \begin{cases} 1 & \text{if } \pi(i) > \pi(j), \\ 0 & \text{if } \pi(i) < \pi(j). \end{cases}$$

After projection,

$$\begin{aligned}
v_i &= 1 + \sum_{j=1, j\neq i}^{d} a_{ij}^\pi \\
&= 1 + \sum_{\{j\,:\,\pi(j) < \pi(i)\}} a_{ij}^\pi \quad (\text{since } a_{ij}^\pi = 0 \text{ if } \pi(i) < \pi(j)) \\
&= 1 + (\pi(i) - 1) \quad (\text{since } a_{ij}^\pi = 1 \text{ for each } j : \pi(j) < \pi(i)) \\
&= \pi(i).
\end{aligned}$$

This is to say, the coordinates of the projected point $[v_1, v_2, \ldots, v_d]^T = [\pi(1), \pi(2), \ldots, \pi(d)]^T$ is the π-representing vertex in the d-permutahedron. $\square$

Remark 1. Mnemonically, projection (17) can be written

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{bmatrix} = \begin{bmatrix} a_{12} & a_{22} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

or

$$\mathbf{v} = \mathbf{A1} + \frac{1}{2}\mathbf{1}.$$

It conforms[8] to the formula in Young (1974) for calculating Borda scores given a binary choice probability (see also Regenwetter & Grofman, 1998a), but there is an important difference. Young's formula applies to the case where (i) the ranking probability (voters profile) P_π is known; and (ii) the binary choice probability is calculated from P_π according to (5). By applying Young's formula, the resulting scores are consistent with Borda's idea of assigning a weight of k $(0 \leqslant k \leqslant d-1)$ in accordance with the number of objects (k) ranked below the focal object. Here, *no* constraint has been placed on the binary choice probability vector. Even when it may not be compatible with *any* ranking probability/voters profile, i.e., when it lies *outside* of the Binary Choice Polytope, a binary choice vector still maps to a unique point in the permutahedron and thereby defines a Borda score. Formula (17) can also be viewed as constructing the graph-theoretic "out-degree" of each candidate, and reflects the essence of the "net probability" constructed by Regenwetter, Marley, and Grofman (2002) for arbitrary binary relations.

Remark 2. Formula (17) provides some useful insights into the property of the so-called "Condorcet winner" in the voting literature. A Condorcet winner is a candidate who is preferred to any other candidate in pairwise contests. Translated into the current notation, a Condorcet winner k satisfies $a_{kj} \geqslant \frac{1}{2}$ for all $j \in [d]$ (and for at least one such j, strict inequality holds). Therefore, the associated Borda score satisfies $v_k > (d+1)/2$. This is to say, a Condorcet winner k cannot have the lowest Borda score, for otherwise $v_k \leqslant v_i$, $(i \in [d], i \neq k)$, $\sum_{i=1}^d v_i \geqslant \sum_{i=1}^d v_k = v_k \cdot d > d(d+1)/2$, contradicting (15). This is a well-known result in the voting community.

Remark 3. Formula (17) provides a connection to another important notion in voting, namely, that of Copeland score (Copeland, 1951, Saari & Merlin, 1996, Merlin & Saari, 1997). A Copeland scoring rule first assigns, in pairwise contests between a focal candidate i and each other candidates $k \in [d]$, $k \neq i$, a mark of $1, \frac{1}{2}, 0$ if candidate i beats, ties with, or loses to candidate k. The marks are then summed to yield a Copeland score v_i^{CP} for each candidate i, $\forall i \in [d]$. Translated into the current notation, a_{ik} is allowed to only take the value of $1, \frac{1}{2}, 0$—formula (17) in this case exactly yields the Copeland score.

Remark 4. Note that mapping (17) is surjective (onto): in general there are uncountably many points of $\mathscr{C}_{d(d-1)/2}$ (i.e., points in the Binary Choice Polytope) that map to the same point in the d-permutahedron Π_{d-1}; this is understandable since the former is of much higher dimension than the latter. The center of the cube $\mathscr{C}_{d(d-1)/2}$, with coordinates $a_{ij} = \frac{1}{2}$, $(i = 1, 2, \ldots d,\ j = i+1, i+2, \ldots, d)$ projects to the barycenter $\mathbf{b} = \frac{d+1}{2}\mathbf{1}$ of Π_{d-1}. The "strength" of this induced ranking can be defined as

$$\frac{1}{d} \sum_{i,j \in [d]} (v_i - v_j)^2.$$

When $v_i = \frac{d+1}{2}$, $\forall i \in [d]$, there is null strength in the induced ranking. For $d = 3$, null ranking strength occurs if and only if the following tri-cyclic condition is satisfied—this is when the binary choice vector coincides with the axis of projection (see Fig. 3):

$$a_{12} = a_{23} = a_{31}.$$

For $d > 3$, cyclic preference is neither a necessary nor a sufficient condition for null ranking strength. Statistical tests on the strength of a rank vector have been devised (Feigin & Cohen, 1978, Alvo, Cabilio, & Feigin, 1982; Feigin & Alvo, 1986, Alvo & Cabilio, 1993).

2.2.2. Compatibility with the BTL choice model

The previous subsection shows that any binary choice vector induces a (generalized Borda) score for each candidate in $[d]$: the induced score equals the sum of probabilities in which a candidate is (pairwise) chosen over other candidates. Here we further investigate whether this induced score is compatible with other representations for binary choice, most notably the BTL representation (Bradley & Terry, 1952, Luce, 1959). If the score $\mathbf{v} = [v_1, v_2, \ldots, v_d]^T (v_i > 0,\ i = 1, 2, \ldots, d)$ is compatible with the BTL representation of the binary choice probability, then

$$a_{ij} = \frac{v_i}{v_i + v_j}.$$

If our induced score (17) is compatibility with the BTL representation, then a mapping arises $\mathscr{G} : \Pi_{d-1} \to \Pi_{d-1}$. In vector components $\mathscr{G} : \mathbf{v} \mapsto [g_1, g_2, \ldots, g_d]^T$ through

$$g_k(v_1, v_2, \ldots, v_d) = \sum_{j=1}^{d} \frac{v_k}{v_k + v_j} + \frac{1}{2}, \quad k \in [d]. \tag{20}$$

The properties of such a map are given in the next corollary.

Corollary 2.4. (i) *The map $\mathscr{G}$ is "rank-preserving", namely, it preserves the ranks of the components of the vector $\mathbf{v} = [v_1, v_2, \ldots, v_d]^T$, in that $g_i \leqslant g_j$ if and only if $v_i \leqslant v_j$.* (ii) *The map $\mathscr{G}$ is "homogenizing", namely, it makes pairwise utility ratios closer to unity*:

$$\left| \frac{g_i}{g_j} - 1 \right| \leqslant \left| \frac{v_i}{v_j} - 1 \right|.$$

[8] The additional term 1 in (17) is the result of our assigning a mark of $\pi(i)$, as opposed to the more conventional $\pi(i) - 1$, for candidate i under π. See footnote 4.

(iii) *The only fixed point of the map $\mathscr{G}$ is the barycenter of the Π_{d-1}.*

Proof. From (20), we have

$$
\begin{aligned}
g_i - g_k &= \sum_{j=1}^{d} \frac{v_i}{v_i+v_j} - \sum_{j=1}^{d} \frac{v_k}{v_k+v_j} \\
&= \sum_{j=1}^{d} \frac{(v_i-v_k)v_j}{(v_i+v_j)(v_k+v_j)} \\
&= (v_i - v_k) \sum_{j=1}^{d} \frac{v_j}{(v_i+v_j)(v_k+v_j)}.
\end{aligned}
$$

Since all the terms under summation are strictly positive, we conclude that $g_i - g_k$ and $v_i - v_k$ have the same sign, and that $v_i = v_k$ iff $g_i = g_k$. Therefore the transformation $\mathscr{G}$ is "rank preserving": $g_i \geqslant g_k$ iff $v_i \geqslant v_k$.

To prove the homogenizing property of the map $\mathscr{G}$, we calculate

$$
\begin{aligned}
\frac{g_i}{g_k} - \frac{v_i}{v_k} &= \left(\frac{g_i}{v_i} - \frac{g_k}{v_k}\right)\frac{v_i}{g_k} \\
&= \left(\sum_{j=1}^{d} \frac{v_k - v_i}{(v_i+v_j)(v_k+v_j)}\right)\frac{v_i}{g_k} \\
&= \left(1 - \frac{v_i}{v_k}\right)\frac{v_i v_k}{g_k}\left(\sum_{j=1}^{d} \frac{1}{(v_i+v_j)(v_k+v_j)}\right). \qquad (21)
\end{aligned}
$$

Therefore, $g_i/g_k - v_i/v_k$ and $1 - v_i/v_k$ have the same sign. Suppose $v_i \leqslant v_k$, or $v_i/v_k \leqslant 1$, then $g_i/g_k \geqslant v_i/v_k$. On the other hand, since $\mathscr{G}$ has just been proven to be rank-preserving, the assumption $v_i \leqslant v_k$ also leads to $g_i \leqslant g_k$ or $g_i/g_k \leqslant 1$. That is to say

$$\frac{v_i}{v_k} \leqslant \frac{g_i}{g_k} \leqslant 1$$

holds whenever $v_i \leqslant v_k$. Similarly we can show that when $v_i \geqslant v_k$,

$$\frac{v_i}{v_k} \geqslant \frac{g_i}{g_k} \geqslant 1.$$

Therefore

$$\left|\frac{g_i}{g_k} - 1\right| \leqslant \left|\frac{v_i}{v_k} - 1\right|.$$

The mapping $\mathscr{G}$ is making the scores more and more homogenous.

Finally, $v_i = \frac{d+1}{2}$, $\forall i \in [d]$ is easily verified to be a fixed point of the map $\mathscr{G}$. To show that it is the only fixed point, i.e., to show that $g_i = v_i$, $\forall i \in [d]$ leads to $v_i = \frac{d+1}{2}$, we note from (21) that requiring $g_i = v_i$ and $g_k = v_k$ for two distinct i, k leads to $v_i = v_k$. Therefore all v_i's must be equal (and equal to $\frac{d+1}{2}$) for any fixed point of $\mathscr{G}$. □

Remark. Given the arbitrary binary choice probability a_{ij}'s, one often seeks to recover a scale, namely the v_i's under the BTL representation, on individual candidates in $[d]$. Part (iii) of Corollary 2.4 implies that the g_i's may not be such $v_i'S$ in general, but are only compatible with the BTL representation, in the sense of (i) and (ii).

2.3. Connection to the RU representation via spherical embedding

2.3.1. Affine equivalence and the representation of RU on the unit sphere

As briefly discussed in Section 1.3, the non-parametric RU representation assumes that each candidate in the choice set $[d] \equiv \{1, 2, \ldots, d\}$ is associated with a real-valued random variable ("utility") v_i, $i = 1, 2, \ldots, d$. The joint distribution of the d random variables $\mathbf{v} = [v_1, v_2, \ldots, v_d]^T$ is given by the probability density $f(v_1, v_2, \ldots, v_d)$. Under (4), the non-coincidence condition of Block and Marschak (1960), the ranking probability and the density function of the jointly distributed RU variables are related through Eq. (2).

Let us now investigate the properties of the region of integration $\Delta_\pi = \{\mathbf{v} = [v_1, v_2, \ldots, v_d]^T : v_{\pi^{-1}(d)} > v_{\pi^{-1}(d-1)} > \cdots > v_{\pi^{-1}(1)}\}$ associated with a given π. Clearly, if $\mathbf{v} \in \Delta_\pi$ is a point in this region, then after a positive affine transformation $\mathbf{v}' = a\mathbf{v} + b\mathbf{1}$, with any $a > 0$ and b, the new point also belongs to this region $\mathbf{v}' \in \Delta_\pi$; this is because $v_{\pi^{-1}(d)} > v_{\pi^{-1}(1)} > \cdots > v_{\pi^{-1}(1)} \rightarrow av_{\pi^{-1}(d)} + b > av_{\pi^{-1}(d-1)} + b > \cdots > av_{\pi^{-1}(1)} + b$ for all $a > 0$ and b. This observation motivates us to introduce the notion of "affine equivalency" $\sim$ over the point-set R^d in which the utility vectors $\mathbf{v}$ are defined. Specifically, $\mathbf{v} \sim \mathbf{v}'$ if and only it there exists a unique $a > 0$ and b such that $\mathbf{v}' = a\mathbf{v} + b\mathbf{1}$.[9] The affine parameters a, b are defined on $\mathscr{A} = (0, \infty) \times (-\infty, \infty)$. We can thus define a quotient space,[10] $R^d/\mathscr{A}$, to represent all equivalence classes of utility vectors $\mathbf{v} \in R^d$.

The idea of affine equivalency of utility vectors is actually rooted in the interval-scale nature of utility measurement. The affine freedom (a, b) is a natural consequence of the arbitrariness in choosing the origin and the unit in assigning utility values (here all utility dimensions are interchangeable therefore necessarily have the same reference zero and measuring unit). So we may impose, as a normative requirement of the RU model, the following representational constraint on the

[9] The affine equivalence relationship is easily seen to be: (i) reflexive: $\mathbf{v} \sim \mathbf{v}$; (ii) symmetric: $\mathbf{v} \sim \mathbf{v}' \rightarrow \mathbf{v}' \sim \mathbf{v}$; this is because $\mathbf{v}' = a\mathbf{v} + b\mathbf{1} \rightarrow \mathbf{v} = (1/a)\mathbf{v}' + (-b/a)\mathbf{1}$; and (iii) transitive: $\mathbf{v} \sim \mathbf{v}', \mathbf{v}' \sim \mathbf{v}'' \rightarrow \mathbf{v} \sim \mathbf{v}''$; this is because $\mathbf{v}' = a\mathbf{v} + b\mathbf{1}, \mathbf{v}'' = a'\mathbf{v}' + b'\mathbf{1} \rightarrow \mathbf{v}'' = (aa')\mathbf{v} + (a'b + b')\mathbf{1}$.

[10] Let X be a set and let $\sim$ be an equivalence relation on X. The equivalence classes due to $\sim$ form a partition of X into pairwise disjoint subsets whose union is X. The set of all equivalence classes will be denoted $X/\sim$, and is called the quotient set of X modulo the equivalence relation. There is a natural projection

$$\gamma : X \rightarrow X/\sim$$

such that $\gamma(x)$ is the equivalence class of x.

probability density function:

$$f(v_1, v_2, \dots, v_d) = g(a, b)\, h(u_1, u_2, \dots, u_d)$$

with

$$u_i = \frac{v_i - b}{a},$$

along with the restrictions

$$\sum_{i=1}^{d} u_i = 0, \qquad \sum_{i=1}^{d} u_i^2 = 1$$

so that

$$b = \frac{\sum_{i=1}^{d} v_i}{d}, \qquad a = \sqrt{\sum_{i=1}^{d} (v_i - b)^2}.$$

Note that the probability density $h(u_1, u_2, \dots, u_d)$ is defined on the unit sphere $\mathbf{u} = [u_1, u_2, \dots, u_d]^T \in S^{d-2}$, which can be divided into $d!$ patches

$$\delta_\pi = \Big\{ \mathbf{u} = [u_1, u_2, \dots, u_d]^T : u_{\pi^{-1}(d)} > u_{\pi^{-1}(d-1)} > \cdots > u_{\pi^{-1}(1)}, \sum_{i=1}^{d} u_i = 0, \ \sum_{i=1}^{d} u_i^2 = 1 \Big\}.$$

Each point $\mathbf{u} \in \delta_\pi$ defines an affine-equivalent class of points $\mathbf{v} \in \Delta_\pi$. Therefore

$$P_\pi = \int_{\Delta_\pi} f(v_1, v_2, \dots, v_d)\, dv_1\, dv_2 \cdots dv_d = \int_{\delta_\pi} h(u_1, u_2, \dots, u_d)\, d\omega,$$

where the a, b variables have been integrated out, with the remaining spherical variables to be integrated on the unit sphere (here $d\omega$ denotes the surface element of S^{d-2}). Parametric models of $h(u_1, u_2, \dots, u_d)$ on S^{d-2} can be prescribed—a prominent one being the von-Mises–Fisher distribution, which belongs to an exponential family.

To summarize, based on considerations of the interval-scale nature of utility vectors, we converted the RU model into an affine-equivalent representation on the unit sphere.

2.3.2. Vertices of a permutahedron as spherical landmarks

As the next step, we follow the approach of McCullagh (1993) to relate the unit sphere S^{d-2} to the permutahedron Π_{d-1}. To do so, the following property of a permutahedron is reviewed.

Proposition 2.5. *A d-permutahedron Π_{d-1} is a $(d-1)$-dimensional object. All of its $d!$ vertices are equidistant to its barycenter $\frac{d+1}{2}[1, 1, \dots, 1]^T$.*

Proof. Recall the definition of a permutahedron (13), using permutation vectors as its vertices. For any permutation vector $\mathbf{x}_{\pi_k} \in R^d$ as given by (12), its projection along the axis $\mathbf{1} \equiv [1, 1, \dots, 1]^T$ equals

$$[1, 1, \dots, 1] \cdot \mathbf{x}_{\pi_k} = \pi_k(1) + \pi_k(2) + \cdots + \pi_k(d) = \frac{d(d+1)}{2},$$

which is a constant (i.e., independent of the particular k). Therefore, for any point $\mathbf{v}$ in the permutahedron, its projection along $[1, 1, \dots, 1]^T$

$$[1, 1, \dots, 1] \cdot \mathbf{v} = [1, 1, \dots, 1] \cdot \left(\sum_{k=1}^{d!} \lambda_k \mathbf{x}_{\pi_k} \right) = \sum_{k=1}^{d!} \lambda_k \frac{d(d+1)}{2} = \frac{d(d+1)}{2}$$

is a constant, since $\sum_{k=1}^{d!} \lambda_k = 1$. This proves that d-permutahedron lies in a $(d-1)$-dimensional subspace, and the distance of any vertex to the barycenter $\frac{d+1}{2}[1, 1, \dots, 1]^T$ is

$$\sum_{j=1}^{d} \left(\pi(j) - \frac{d+1}{2} \right)^2 = \sum_{j=1}^{d} \left(j - \frac{d+1}{2} \right)^2 = \frac{(d+1)d(d-1)}{12},$$

a constant. To prove that it is of dimensionality $d-1$ exactly (and not less), we only need to show that if its projection along a vector $\mathbf{c} = [c_1, c_2, \dots, c_d]^T$ is a constant, then that vector must be proportional to $\mathbf{1}$, i.e., $c_1 = c_2 = \cdots = c_d$. To show, for instance, $c_1 = c_2$, consider the projection of two vertices $\mathbf{x}_1 = [1, 2, 3, \dots, d]^T$ and $\mathbf{x}_2 = [2, 1, 3, \dots, d]^T$. That $\mathbf{x}_1 \cdot \mathbf{c} = \mathbf{x}_2 \cdot \mathbf{c}$ immediately leads to $c_1 = c_2$. Therefore, all c_i's must be equal. □

Remark 1. Proposition 2.5 shows that the $d!$ vertices of a permutahedron are equal-distant to its barycenter. This suggests that a properly scaled d-permutehedron can be circumscribed by the unit sphere S^{d-2}. For example, the hexagon (in Fig. 2a) is circumscribed by a circle (S^1), whereas the 4-permutahedron (in Fig. 2b) is circumscribed by a sphere (S^2). Therefore, the $d!$ vertices of the permutehedron serve as landmarks of S^{d-2} on which the probability density $h(u_1, u_2, \dots, u_d)$ is defined. The vertex $\pi = \langle \pi(1)\pi(2)\cdots\pi(d) \rangle$ corresponds to equally spaced u-values:

$$u_{\pi^{-1}(d)} - u_{\pi^{-1}(d-1)} = u_{\pi^{-1}(d-1)} - u_{\pi^{-1}(d-2)} = \cdots = u_{\pi^{-1}(2)} - u_{\pi^{-1}(1)}. \tag{22}$$

Solving for the u_i's yields

$$u_{\pi^{-1}(j)} = \frac{\sqrt{3}(2j - (d+1))}{\sqrt{d(d+1)(d-1)}}.$$

With this set of vertices, most points on S^{d-2} can be classified, using the spherical distance metric, as belonging to a certain closest vertex—each vertex $\pi = \langle \pi(1)\pi(2)\cdots\pi(d) \rangle$ has a region of "attraction" δ_π, i.e., a neighborhood within which all points are closer to this vertex than to any other vertex. The only exceptions are the points on the boundaries of such regions—they occur when at least two (but not all d) candidates have same utility values, i.e., are tied in a weak order. Each region of attraction turns out to be the intersection of an unbounded, polyhedral cone (whose apex is at the center of the sphere) with S^{d-2}; the sphere S^{d-2} itself is then divided, by these $d!$ polyhedral cones, into $d!$ patches δ_π, $\pi \in \mathscr{L}_d$.

Remark 2. Three observations of this spherical representation can be made. First, the Block and Marschak (1960) condition (4) is equivalent to stating that the boundaries of δ's have measure zero in probability. Second, ranking probabilities can be viewed as a special kind of RU model in which the probability measure is concentrated on the $d!$ isolated points on the sphere S^{d-2}; this is exactly Block and Marschak's (1960) construction of RU density function from a given ranking probability. Third, for any RU distribution, the ranking probability, induced according to (2), can be viewed as individual vertices (linear orders) of the permutahedron "pulling" the probability measure that are "spread" in respective neighborhoods towards their centers; the boundary of attraction is given by setting $u_{\pi^{-1}(k+1)} = u_{\pi^{-1}(k)}$ for one of the $k = 1, 2, \ldots, d-1$; this is in accord with the geometric description of Borda's rule of converting Borda scores to a ranking assuming no ties (Cook & Seiford, 1982).

Remark 3. It is worth mentioning that the spherical representation of the preference space derived here is in the same spirit as the topological approach to preference aggregation in the social choice and welfare context (Chichilnisky, 1980, Chichilnisky & Heal, 1983, Baryshnikov, 1993, Heal, 1997). However, since the utility vectors associated with the null-preference $v_1 = v_2 = \cdots = v_d$ map to the barycenter of the d-permutahedron, the space of utility vectors (modulo the affine equivalence) should be, strictly speaking, $S^{d-2} \cup \{0\}$. The inclusion of the null-preference point introduces nontrivial topological consequences for the existence of a proper social choice function. An elaboration on this topic will be given by another paper (Jones, Zhang, & Simpson, 2003).

2.4. Connection to AV via monotone path polytope

2.4.1. Approval voting

AV is a mechanism of social choice through which each voter selects or approves of, from a master set of candidates $[d]$, a subset of individuals $S \subseteq [d]$ who presumably are above a voter's "threshold" of acceptability. First informally suggested by Robert T. Weber, it was popularized as a viable alternative voting mechanism in a number of articles by Brams and Fishburn (e.g., 1978, 1983). To date, AV mechanism was or has been adopted by several professional societies in their electoral process, and its effectiveness was analyzed using the outcomes of 10 real-life elections (Regenwetter & Grofman, 1998b). Once AV ballots are collected, a probability distribution P_S (in terms of relative frequencies) over all subsets $S \subseteq [d]$ is specified. In the common tally procedure of AV (Brams & Fishburn, 1978), the accumulation of votes (across the voting population) in favor of each candidate is according to the total number of times a candidate has appeared in the subsets of which he is a member/element. The votes (AV score) for candidate i is counted by summing across the probability distribution P_S over all subsets S

$$v_i = \sum_{S \subseteq [d]} P_S \, m_i(S),$$

where $m_i(S)$ is the membership function

$$m_i(S) = \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, each subset S (approved by a voter) contributes to the score vector by an amount $[m_1(S), m_2(S), \ldots, m_d(S)]^T$. We call the total score accumulated according to this "vanilla" AV method the *Brams–Fishburn score*, $\mathbf{v}^{\mathrm{BF}} = [v_1, v_2, \ldots, v_d]^T$:

$$\mathbf{v}^{\mathrm{BF}} = \sum_{S \subseteq [d]} P_S \begin{bmatrix} m_1(S) \\ m_2(S) \\ \vdots \\ m_d(S) \end{bmatrix}. \tag{23}$$

One could, of course, implement other tally procedures. An obvious alternative is to inversely weigh the importance of each subset by the percentage of the total voters choosing subsets equal to its size:

$$\mathbf{v}^{\mathrm{SI}} = \sum_{S \subseteq [d]} \frac{P_S}{f(|S|)} \begin{bmatrix} m_1(S) \\ m_2(S) \\ \vdots \\ m_d(S) \end{bmatrix}, \tag{24}$$

where $f(k)$ is given by (7). As we shall see (Section 2.4.6), this tally procedure is related to the SI model of AV, and hence hereby referred to as the *SI score*. In essence, it encourages a voter to choose an unpopular set-size in order to increase the importance of his/her vote.

Apart from these different tally procedures, researchers have also been interested in underlying models of voters' preferences that generate the AV ballots. For instance, one may imagine that approving a particular subset S amounts to equal-probabilistically choosing all linear orders whose top $|S|$ elements are precisely the elements of S (hereafter referred to as the "equal-probability model"). On the other hand, one may hypothesize a two-stage process, as in the SI model, and seek to uncover, from the given data P_S, an underlying probability distribution P_π over all linear orders $\pi \in \mathscr{L}_d$ that is "consistent" with P_S, in the sense of (6). While in the former case there is a unique probability distribution P_π over linear orders that can be inferred from ballots, in the latter case, there is no guarantee that such a P_π will ever be found for an arbitrary P_S. The condition under which this is possible is called the characterization problem for AV, whose solutions, in the space of all subset choice probabilities P_S, form a polytope, the so-called AV Polytope (Doignon & Regenwetter, 1997; Doignon & Regenwetter, 2002; Doignon & Fiorini, to appear).

One interesting question arises: how are AV scores (the Brams–Fishburn score versus the SI score) related to the AV models (the "equal-probability" model versus the SI model)? Though the former is a matter of tally procedures while the latter latent preference models, it will be shown (Sections 2.4.4 and 2.4.6) that AV scores exactly equal Borda scores of the latent probability distribution of linear orders under the corresponding models. This again demonstrates the important role played by Borda scoring method (and the permutahedron) in linking a tally procedure with a voter preference model in the case of AV.

2.4.2. Permutahedron as a monotone path polytope

It is known (Ziegler, 1995) that permutahedron can be realized as a "monotone path polytope" arising from the lift-up of the projection from a cube to a line segment. The mathematical details of monotone path polytopes (and fiber polytopes in general) are given in Appendix A. Here we give the motivation of considering such a projection, and then reproduce this conclusion (Theorem 2.6) and its proof due to Ziegler. Since the construction is crucial in linking ranking probability to a subset choice model, the necessary mathematical facts are reviewed here using a language familiar to researchers of probabilistic choice.

Let $\mathscr{C}_d = [0,1]^d \subset R^d$ be the d-dimensional unit cube with the lower left corner aligned with the origin. There is a natural correspondence between the vertices of $\mathscr{C}_d$ and the subsets of $[d]$ to be constructed as follows. Denote the base vectors of R^d as $\mathbf{e}_i = \Big[\underbrace{0,\dots,0}_{i-1},1,\underbrace{0,\dots,0}_{d-i}\Big]^T$ for $i \in [d]$. A vertex of the cube $\mathscr{C}_d$, represented as a vector of the form $\sum_{i\in S}\mathbf{e}_i$, is identified with the subset $S \subseteq [d]$ itself. Recall that any ranking (ordering) π induces a sequence of nested subsets

$$\emptyset \subset \{\pi^{-1}(d)\} \subset \{\pi^{-1}(d), \pi^{-1}(d-1)\} \subset \cdots \subset \{\pi^{-1}(d), \pi^{-1}(d-1), \dots, \pi^{-1}(1)\} \equiv [d].$$

We denote $L(\pi,k)$, namely the kth "leading set" of π, as the set of candidates at the k top ranks associated with a given π

$$L(\pi,k) = \{\pi^{-1}(d), \pi^{-1}(d-1), \dots, \pi^{-1}(d-k+1)\} \tag{25}$$

for $1 \leqslant k \leqslant d$ with $|L(\pi,k)| = k$, and

$L(\pi,0) = \emptyset$.

These nested subsets induced by π simply obey

$L(\pi,0) \subset L(\pi,1) \cdots \subset L(\pi,d)$

as more and more candidates are included, based on their desirability as determined by π. Geometrically, i.e., using vertices to represent subsets, this sequence can be seen as defining a "path" associated with π, denoted σ^π, that starts from the lower-left vertex of the cube $\mathscr{C}_d$ and travels along its edges all the way up to its upper-right vertex:

$$\sigma^\pi:\ \mathbf{0} \equiv \sigma^\pi(0) \to \sigma^\pi(1) \to \sigma^\pi(2) \to \cdots \to \sigma^\pi(d) \equiv \mathbf{1}, \tag{26}$$

where each $\sigma^\pi(k)$ is the vertex corresponding to the k-element set $L(\pi,k)$

$$\sigma^\pi(k) = \sum_{i=1}^{k} \mathbf{e}_{\pi^{-1}(d-i+1)}, \quad k = 1,2,\dots,d.$$

To formalize this intuition under the framework of monotone path polytope, construct the following special projection γ from $\mathscr{C}_d$ to R^1 along the diagonal axis $[1,1,\dots,1]^T$:

$$\gamma: [0,1]^d \to [0,d], \quad \mathbf{x} \mapsto [1,1,\dots,1]\cdot\mathbf{x} = \sum_{i=1}^{d} x_i. \tag{27}$$

Here the range $[0,d] \subset R^1$ of the projection is actually a line segment $[0,d]$ (see Fig. 4)[11]. When the aforementioned path σ^π is projected via γ onto line segment (see Fig. 4), points that are increasingly farther away from the origin map onto increasingly larger values between 0 and d—the vertex corresponding to the set $L(\pi,k)$ maps to the point $k \in [d] \cup \{0\}$. Fig. 4 shows an example for $d=3$, where the darker line indicates the path corresponding to $\pi = \langle 213 \rangle$, so that $L(\pi,1) = \{3\}, L(\pi,2) = \{1,3\}, L(\pi,3) = \{1,2,3\}$, and $\sigma^\pi(1) = \mathbf{e}_3, \sigma^\pi(2) = \mathbf{e}_1 + \mathbf{e}_3, \sigma^\pi(3) = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$.

The projection $\gamma: R^d \ni \mathscr{C}_d \to [0,d] \in R^1$ induces a so-called fiber polytope (see Appendix A), which necessarily is of dimension $d-1$, and which lives in the fiber orthogonal to the projected dimension and originating

[11] Note that in the graph, the cube's sides have lengths $1/d$ and the projection is along the vector $\mathbf{c} = 1/\sqrt{d}[1,1,\dots,1]$, but the resulting algebraic relationship (27) still holds.

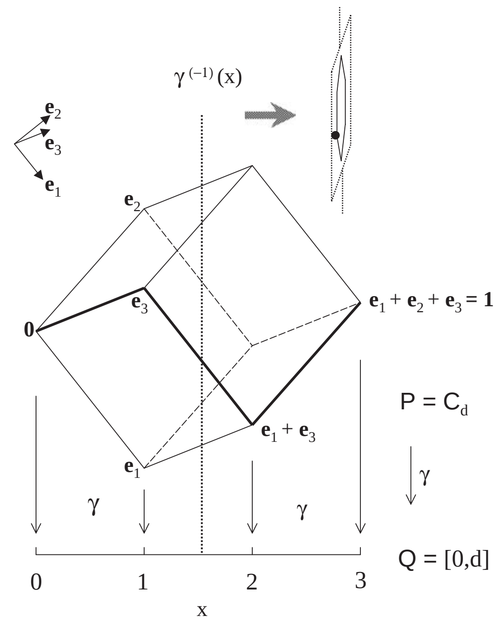

Fig. 4. The d-permutahedron as a fiber polytope, i.e., resulting from the construction of monotone paths associated with the projection of a cube of dimension d to a line segment $[0, d]$. Here, for $d = 3$, the thicker line represents a particular linear-order path (corresponding to $\langle 213 \rangle$, represented as a dot in the inset), while the vertical plane cutting across the cube in the middle is shown in the inset to contain the proper fiber polytope—a hexagon.

from $\gamma(\mathbf{r_0})$ where $\mathbf{r_0}$ is the barycenter of $\mathscr{C}_d$. It turns out that this resulting fiber polytope is in fact a d-permutahedron.

Theorem 2.6 (Ziegler, 1995, p.303). *The fiber polytope associated with the diagonal projection* (27) *from a d-cube* $\mathscr{C}_d \subset R^d$ *to a line segment* $[0, d] \subset R^1$ *is a d-permutahedron.*

Proof. According to its definition, the fiber polytope is the collection of points whose coordinates are given by integral (A3) for all possible (horizontal) sections σ of $\mathscr{C}_d$ with respect to the projected line segment. To prove that the monotone path polytope resulting from the specific projection (27) is indeed a permutahedron, one only needs to calculate the coordinates corresponding to the monotone paths that consist of extreme points of the cube, i.e., its edges. Given a ranking π, an explicit calculation of the coordinates associated with the path σ^π (26) is given below, which is adapted from Ziegler (1995, p. 303):

$$
\begin{aligned}
&\int_0^d \sigma^\pi(x)\, dx \\
&= \sum_{k=1}^{d} \frac{1}{2}(\sigma^\pi(k-1) + \sigma^\pi(k)) \\
&= \frac{1}{2}\sigma^\pi(0) + \sigma^\pi(1) + \sigma^\pi(2) + \cdots + \sigma^\pi(d-1) \\
&\quad + \frac{1}{2}\sigma^\pi(d) \\
&= d\,\mathbf{e}_{\pi^{-1}(d)} + (d-1)\mathbf{e}_{\pi^{-1}(d-1)} + \cdots + 2\,\mathbf{e}_{\pi^{-1}(2)} \\
&\quad + \mathbf{e}_{\pi^{-1}(1)} - \frac{1}{2}\mathbf{1} \\
&= \sum_{k=1}^{d} k\mathbf{e}_{\pi^{-1}(k)} - \frac{1}{2}\mathbf{1} \\
&= \sum_{k=1}^{d} \pi(k)\,\mathbf{e}_k - \frac{1}{2}\mathbf{1} \\
&= \begin{bmatrix} \pi(1) \\ \pi(2) \\ \vdots \\ \pi(d) \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.
\end{aligned}
$$

This established a one-to-one correspondence of a π-defined path in $\mathscr{C}_d$ to a π-representing vertex of Π_{d-1}—the fiber polytope associated with projection (27) is indeed a d-permutahedron. $\square$

2.4.3. *Subset choice probability and facets of a permutahedron*

The fact (Theorem 2.6) that the fiber polytope induced by projection (27) turns out to be the d-permutahedron is of special significance in linking the subset choice paradigm to ranking probability. First, as mentioned earlier, the 2^d vertices of a d-cube $\mathscr{C}_d$ are in one-to-one correspondence with all subsets of $[d]$; it is easy to envision a d-dimensional vector, $[m_1(S), m_2(S), \ldots, m_d(S)]^T$, whose binary-valued vector component indicates whether or not an element of $[d]$ has been included in the subset $S \subseteq [d]$. Therefore $\mathscr{C}_d$ is a natural representation for the subset choice paradigm, where the subset choice probability P_S can be defined on the subsets $S \subseteq [d]$, which are vertices of $\mathscr{C}_d$. Second, the surjective projection (27) of $\mathscr{C}_d$ onto the line segment $[0, d]$ collapses subsets according to their set-sizes—all subsets of a given size $|S| = k$ project to the same (integer-valued) point $k \in [d] \cup \{0\}$, so the probability distribution over set-size $f(k)$, $k = 0, 1, \ldots, d$, can be defined, on the integer support along the line-segment $[0, d]$. Third, a monotone path σ^π on the boundary of the cube $\mathscr{C}_d$, which starts from the vertex $[0, 0, \ldots, 0]^T$ and ending at the vertex $[1, 1, \ldots, 1]^T$, generates a sequence of leading sets $L(\pi, k)$ of π:

$$L(\pi, k) = L(\pi, k-1) \cup \{\pi^{-1}(d-k+1)\}.$$

The gradual inclusion of elements into $L(\pi, k)$ as $k = |L(\pi, k)|$ increases depends on the ordering π^{-1}; it always starts with the most desired object $\pi^{-1}(d)$, followed by the second most desired one $\pi^{-1}(d-1)$, so on. Since each such path $\sigma^\pi(1) \to \sigma^\pi(2) \to \cdots \to \sigma^\pi(d)$ uniquely represents a linear order, it is called a *linear-order path*.

Ranking probability P_π can be defined on the $d!$ linear-order paths.

These considerations provide all necessary ingredients for a concise description and characterization of subset choice models (in connection with ranking models). Consider a particular vertex $V_S \in \mathscr{C}_d$ corresponding to a given subset $S \subset [d]$ with set-size $|S| = k$, $(1 \leqslant k \leqslant d-1)$. The vertex V_S is *en route* of many linear-order paths of $\mathscr{C}_d$, in fact those paths that correspond to all such linear orders that rank elements of S ahead of elements of $[d] \backslash S$, i.e., those π's whose kth leading set $L(\pi, k)$ equals S. This vertex (of the cube) defines a bipartition of the set $[d]$ into two disjoint subsets that are set-wise ordered for desirability; hence, from Section 2.1.2, it corresponds to a unique facet of the permutahedron Π_{d-1}, hereby denoted $\mathscr{F}_S$. There are $d!/((d-k)!k!)$ vertices (of Π_{d-1}) on this facet, each corresponding to one of the linear-order paths on the cube $\mathscr{C}_d$ constrained to pass through V_S. All such linear orders $\pi \in \mathscr{L}_d$ satisfy $S = L(\pi, k)$, or

$$S = \{\pi^{-1}(d), \pi^{-1}(d-1), \ldots, \pi^{-1}(d-k+1)\}.$$

Following Doignon and Regenwetter (1997), denote the set of all rankings/permutations whose leading set (as defined by (25)) equals a given S as Π_S:

$$\Pi_S = \{\pi \in \mathscr{L}_d : L(\pi, k) = S, k = |S|\}.$$

In other words, Π_S, which is a subset of $\mathscr{L}_d$, contains all those π's that rank any candidate belonging to S more favorably than any one not belonging to S. Each element of Π_S, which is a ranking and which at the same time represents a linear-order path, corresponds to a certain vertex of the facet $\mathscr{F}_S$ of the permutahedron. Note that a vertex of the permutahedron Π_{d-1} corresponds to a ranking π, whereas a vertex of the cube $\mathscr{C}_d$ corresponds to a subset S of $[d]$; one should not confuse the two. On the other hand, π corresponds to a path of the d-cube, while S corresponds to a facet of Π_{d-1}—the facet in fact represents the convex hull of all monotone paths of $\mathscr{C}_d$ constrained to pass one of its (the cube's) vertex. All vertices of $\mathscr{C}_d$, organized in terms of the number of steps away from the origin, give a total count of

$$\binom{d}{1} + \binom{d}{2} + \cdots + \binom{d}{d-1} = 2^d - 2.$$

The $2^d - 2$ facets of the permutahedron are organized in the following way: d facets each with $(d-1)!$ edges, $d(d-1)/2$ facets each with $2(d-2)!$ edges, $\ldots$, $d!/(k!(d-k)!)$ facets each with $k!(d-k)!$ edges, $\ldots$, and d facets each with $(d-1)!$ edges. The total number of edges is

$$\frac{1}{2}(d \cdot (d-1)! + d(d-1)/2 \cdot 2(d-2)! + \cdots + d \cdot (d-1)!) = d/2 \cdot d!,$$

which conforms to the formula for f-numbers (Eq. (16), for $j = 1$).

For example, when $d = 4$ (see Fig. 2b), the subset $S = \{1, 3, 4\} \subset \{1, 2, 3, 4\}$ is associated with a particular vertex $\mathbf{e}_1 + \mathbf{e}_3 + \mathbf{e}_4$ of $\mathscr{C}_4$. Each ranking that is compatible with S in placing candidates "1", "3", "4" in front of "2" (i.e., $\pi(2) = 1$) has the general form $\pi = \langle *1** \rangle$ or $\pi^{-1} = (***2)$; the set of permutations $\Pi_S = \{\langle 4\mathbf{1}32 \rangle, \langle 4\mathbf{1}23 \rangle, \langle 3\mathbf{1}24 \rangle, \langle 2\mathbf{1}34 \rangle, \langle 2\mathbf{1}43 \rangle, \langle 3\mathbf{1}42 \rangle\} = \{(134\mathbf{2}), (143\mathbf{2}), (413\mathbf{2}), (431\mathbf{2}), (341\mathbf{2}), (314\mathbf{2})\} \subset \mathscr{L}_d$ corresponds to all paths of $\mathscr{C}_d$ passing through $\mathbf{e}_1 + \mathbf{e}_3 + \mathbf{e}_4 = [1, 0, 1, 1]^T$ with monotonically increasing path-lengths. These paths map onto the vertices (of Π_{d-1}) that define the facet $F_{\{1,3,4\}}$. Unfortunately, it is not possible to visualize the projection γ for $d = 4$ in the same way we did for $d = 3$ in Fig. 4.

2.4.4. Brams–Fishburn score of AV ballots and the equal-probability model

The above combinatorial–geometric considerations hint at some intimate connections between the AV paradigm and the ranking paradigm, along with their scoring rules (Bram–Fishburn score for the former and Borda score for the latter). Specifically, the event of choosing a particular subset (a vertex of the d-cube, a facet of the d-permutahedron) and the event of ranking candidates in a particular linear order (a monotone path of the d-cube, a vertex of the d-permutahedron) can be made equivalent in terms of the scores they generate over the set of candidates.

Theorem 2.7 (Connection of the Brams–Fishburn score and the Borda score). *In an AV paradigm, if the approval of a subset $S \subseteq [d]$ of candidates amounts to assigning equal probability to all rankings $\Pi_S \subseteq \mathscr{L}_d$ consistent with S, i.e., to those rankings which place the approved candidates in front of the non-approved ones, then Brams–Fishburn score* (23) *is equivalent, up to an affine transform, to computing the Borda score for each candidate that is consistent with the latent ranking probabilities of the voters.*

Proof. For any subset $S \subseteq [d]$ with $0 \leqslant |S| = k \leqslant d$, the corresponding set of consistent rankings is $\Pi_S \subset \mathscr{L}_d$. Assuming equal probability among the total of $k!(d-k)!$ rankings in Π_S, the Borda score *due to* S is computed as

$$\mathbf{v}^{\mathrm{Bd}}(S) = \frac{1}{k!(d-k)!} \sum_{\pi \in \Pi_S} \begin{bmatrix} \pi(1) \\ \pi(2) \\ \vdots \\ \pi(d) \end{bmatrix}.$$

Denote the ith component of $\mathbf{v}^{\mathrm{Bd}}(S)$ as $v_i^{\mathrm{Bd}}(S)$. If $i \in S$, then $\pi(i) \in \{d, d-1, \ldots, d-k+1\}$. The sum of the

$k!\,(d-k)!$ terms of $\pi(i)$, as π runs through Π_S, equals

$$(k-1)!\left(\sum_{j=d-k+1}^{d} j\right)\cdot(d-k)! = k!(d-k)!\,\frac{2d-k+1}{2},$$

for $i \in S$.

If $i \notin S$, then $\pi(i) \in \{1,2,\ldots,d-k\}$, and the sum of the $k!\,(d-k)!$ terms of $\pi(i)$ equals

$$(d-k-1)!\left(\sum_{j=1}^{d-k} j\right)\cdot k! = k!(d-k)!\,\frac{d-k+1}{2},$$

for $i \notin S$.

Therefore

$$\begin{aligned} v_i^{\mathrm{Bd}}(S) &= \frac{2d-k+1}{2} m_i(S) + \frac{d-k+1}{2}(1-m_i(S)) \\ &= \frac{d}{2} m_i(S) + \frac{d-k+1}{2}. \end{aligned}$$

Note that the coefficient before $m_i(S)$ is independent of k. Summing over $S \subseteq [d]$

$$v_i^{\mathrm{Bd}} = \sum_{S\subseteq[d]} P_S\, v_i^{\mathrm{Bd}}(S)$$

yields

$$v_i^{\mathrm{Bd}} = \frac{d}{2}\left(\sum_{S\subseteq[d]} P_S\, m_i(S)\right) + \left(\frac{d+1}{2} - \frac{1}{2}\sum_{k=1}^{d} kf(k)\right), \tag{28}$$

where $f(k)$ denotes the probability distribution over subset size k as given by (7), with $\sum_{k=0}^{d} f(k) = 1$. This shows that the Borda score v_i^{Bd} for candidate i is equivalent to the Brams–Fishburn score v_i^{BF} in (23), apart from an affine transform. □

Remark 1. The value $\sum_{k=1}^{d} kf(k) = \sum_{S\subseteq[d]} P_S \cdot |S|$ is the average subset size chosen by the voting population. It is uniformly subtracted from the scores of all candidates, and therefore will not affect their Borda ordering.

Remark 2. Theorem 2.7 says that approving a particular subset (under Brams–Fishburn scoring) is equivalent to several possible rankings (using Borda score) chosen in a way that is consistent with the approved subset and without bias; here "consistency" is taken to mean that the chosen k candidates rank higher than the non-chosen d–k candidates, and "without bias" means equal-probabilistically. This gives the geometric interpretation of the Brams–Fishburn tally procedure, i.e., each choice of a subset (vertex of the d-cube) $S \subseteq [d]$ amounts to an equal-probabilistic assignment of all rankings (linear-order paths on the d-cube) that are compatible with the chosen subset. The subset choice probability P_S will induce a ranking probability P_π, as stated in the next corollary (note that with an abuse of notation, we use the same symbol P for ranking probability and for subset choice probability).

Corollary 2.8 (Latent profile under equal-probabilistic model). The probability distribution over subsets P_S in AV, under equal-probability model, induces a probability distribution over rankings P_π, under Borda scoring, through:

$$\begin{aligned} P_\pi &= \sum_{k=0}^{d} P_{L(\pi,k)} \\ &= \frac{1}{d!} P_\emptyset + \sum_{k=1}^{d} \frac{1}{k!(d-k)!} \\ &\quad\times P_{\{\pi^{-1}(d),\pi^{-1}(d-1),\ldots,\pi^{-1}(d-k+1)\}}. \end{aligned} \tag{29}$$

Proof. The weighting factor for each eligible subset, $1/(k!(d-k)!)$, is according to the multiplicity of all permutations consistent with a given subset of size k. Summing up all contributing $L(\pi,k)$ gives (29). □

Remark. Intuitively, the "support" for a ranking π is contributed to by all subsets S with which π is consistent in the sense that all S's are leading sets of π. The support for π is accumulated along the path of its leading sets with increasing set-size

$$\emptyset \equiv L(\pi,0) \subset L(\pi,1) \subset \cdots \subset L(\pi,d) \equiv [d],$$

where $L(\pi,k)$ is defined in (25). From Corollary 2.2, each $L(\pi,k)$ (for any $\pi \in \mathscr{L}_d$ and $1 \leqslant k \leqslant d$) defines a facet of Π_{d-1} that contain $k!(d-k)!$ vertices. Since the π-representing vertex of Π_{d-1} is the intersection of $d-1$ such facets (corresponding to $k = 1,2,\ldots,d-1$ respectively), each term in summation of (29) is indeed a contribution from those facets to their common vertex.

2.4.5. SI model and the core of the AV polytope

Under the SI model of subset choice (Falmagne & Regenwetter, 1996, Doignon & Regenwetter, 1997), voters are assumed to have (i) an explicit profile or probability distribution P_π of all rankings π over $[d]$; and independently (ii) a probability distribution $f(k)$ over the size of the subset ($k = 0,1,\ldots,d$) that is to be approved of. The subset choice probability P_S is related to the ranking probability P_π and the size probability $f(k)$ through (6).

The SI model proposes to group all subsets according to their sizes; geometrically, this amounts to grouping the vertices of $\mathscr{C}_d$ according to their image of projection on the line-segment $[0,d]$. For each k, any linear-order path σ^π (for $\pi \in \mathscr{L}_d$) has to include one of the

$d!/(k!(d-k)!)$ vertices that project to $k \in [0, d]$. The fundamental assumption of the SI model (6) is that the sum of the probabilities of all rankings compatible with a chosen subset equals, after normalization by the set-size probability $f(k)$, the subset probability P_S.

The SI model for subset choice differs from the equal-probability model of AV in two important aspects: (i) While the latter can always induce a probability distribution of ranking from an empirically observed probability distribution over subsets (namely AV ballots), the former assumes the existence of a probability distribution of ranking from which the empirically observed probability distribution of subsets (AV ballots) are related to; (ii) While the latter attributes equal probability to all paths (rankings) compatible with a given subset, the former does not require (and indeed often violates) the equal-probability interpretation of compatible paths (rankings).

The existence of a latent ranking probability distribution is by no means automatically guaranteed for any AV ballots—the necessary conditions are characterized by the AV polytope (Doignon & Regenwetter 1997, 2002, Doignon & Fiorini, to appear). Here we further investigate necessary conditions on the subset choice probability P_S such that the SI model would yield the same ranking probability as the equal-probability model. The region within the AV Polytope that is compatible with the equal-probability model is called the *core* of the AV Polytope.

Theorem 2.9 (Core of the AV polytope). *The subset probability P_S consistent with both the SI and the equal-probability interpretations of AV satisfies the following simultaneous equations*:

$$P_S = f(|S|)\left(P_S + \sum_{\{S' : S' \subset S\}} \rho_1(S', S) P_{S'} + \sum_{\{S' : S' \supset S\}} \rho_2(S', S)\, P_{S'}\right), \tag{30}$$

where S, S' are subsets of $[d]$,

$$\rho_1(S, S') = \frac{(|S| - |S'|)!(d - |S|)!}{(d - |S'|)!} \quad \textit{for } S' \subset S,$$

$$\rho_2(S, S') = \frac{(|S'| - |S|)!\,(|S|)!}{|S'|!} \quad \textit{for } S \subset S',$$

and $f(k)$ is the probability over set-size given by (7).

Proof. According to Corollary 2.8, under the equal-probability model, the induced ranking probability P_π is related to the probability distribution over subsets P_S through (29). Requiring P_S to also fulfill the condition of the SI model (6) yields

$$\begin{aligned} P_S &= f(|S|) \sum_{\{\pi \in \mathscr{L}_d : L(\pi, |S|) = S\}} \sum_{l=0}^{d} \frac{1}{l!(d-l)!} P_{L(\pi,l)} \\ &= f(|S|) \sum_{l=0}^{d} \frac{1}{l!(d-l)!} \left(\sum_{\{\pi \in \mathscr{L}_d : L(\pi,|S|) = S\}} P_{L(\pi,l)} \right). \end{aligned}$$

With S fixed, the summation over π is performed for those rankings having their top $k \equiv |S|$ elements as those in S, and the summation over set-size l is performed for variable-size leading sets of a given, eligible π. When $l < k$, since $L(\pi, l) \subset L(\pi, k)$ holds for any $\pi \in \mathscr{L}_d$, the summation over π in the parentheses above only includes terms $P_{S'}$ for which $S' \subset L(\pi, k) = S$

$$\sum_{\{\pi \in \mathscr{L}_d : L(\pi,|S|) = S\}} P_{L(\pi,l)} = \sum_{\{S' : S' \subset S, |S'| = l\}} \chi_{S'}\, P_{S'},$$

where $\chi_{S'}$ denotes the multiplicity of an l-element subset $S' \subset S$ that can be produced by an eligible π satisfying (i) $L(\pi, k) = S$ and (ii) $L(\pi, l) = S'$. A straightforward combinatoric consideration gives $\chi_{S'} = l!(k-l)!(d-k)!$. Therefore

$$\sum_{\{\pi \in \mathscr{L}_d : L(\pi,|S|) = S\}} P_{L(\pi,l)} = \sum_{\{S' : S' \subset S, |S'| = l\}} l!(d-l)!(d-k)! P_{S'}.$$

Similarly, for $l > k$, one may derive

$$\sum_{\{\pi \in \mathscr{L}_d : L(\pi,|S|) = S\}} P_{L(\pi,l)} = \sum_{\{S' : S' \supset S, |S'| = l\}} k!(l-k)!(d-l)! P_{S'}.$$

Finally, for $l = k$, the only term that survives the summation is $S' = S$. Putting these results together yields the desired formula (30). □

Remark 1. Note that the system of simultaneous equations (30) is linear in P_S. The total number of independent constraints on P_S, according to (30), is

$$2^d - (d+1) + 1 = 2^d - d;$$

this is because at each $k = 0, 1, \ldots, d$, $\sum_{\{S : |S| = k\}} P_S = f(k)$, while $\sum_{k=0}^{d} f(k) = 1$ (resulting in over-counting of the number of constraints). The total number of independent components of P_S is

$$2^d - 1 - (2^d - d) = d - 1,$$

which is the dimension of the d-permutahedron! Therefore, the region within the AV polytope satisfying (30) forms the core of the probability space associated with the subset choice paradigm.

Remark 2. Note that the reciprocal of $\rho_1(S, S')$ and of $\rho_2(S, S')$

$$\rho_1^{-1} = \binom{d-|S'|}{d-|S|}, \quad \rho_2^{-1} = \binom{|S'|}{|S|}$$

equal the number of ways of choosing $d-|S'|$ items out of $d-|S|$ items ($|S'|<|S|$), or choosing $|S|$ items out of $|S'|$ items ($|S'|>|S|$), respectively. Therefore, $\rho_2(S, S')$ can be interpreted as the probability that those voters choosing S' would also have chosen $S \subset S'$ if they were restricted to choose a smaller subset, while $\rho_1(S, S')$ can be interpreted as the probability that those voters not choosing $\bar{S}' = [d]\backslash S'$ would not have chosen $[d]\backslash S = \bar{S} \subset \bar{S}'$ either even if they were to enlarge the approved subset. Under this interpretation, the constraints (30), after a recast

$$\begin{aligned}\frac{P_S}{\sum_{S' \subset S} \rho_1(S', S) P_{S'} + \sum_{S' \supset S} \rho_2(S', S) P_{S'}} &= \frac{f(|S|)}{\sum_{k=0, k \neq |S|}^{d} f(k)} \\ &= \frac{f|S|}{1 - f(|S|)},\end{aligned} \tag{31}$$

can be viewed as some kind of consistency requirement on subset probability if the set-size is a truly irrelevant factor in determining the aggregated preference (in terms of Borda score over the candidates, see Section 2.4.1). It gives the condition under which the Brams–Fishburn score and the SI score become equivalent.

2.4.6. *Comparing the SI score and the Brams–Fishburn score*

In calculating the Brams–Fishburn score (23), total votes for each candidate are tallied according to how many times a candidate is included among all approved subsets. As shown in Section 2.4.4, this vote counting rule is equivalent, up to an affine transform, to the rule of calculating Borda scores, provided that the choice of a subset is interpreted as an equal-probabilistic approval of all linear orders consistent with the chosen subset. In this sense, AV scores (arising from subset choice paradigm) are said to be equivalent to Borda scores (arising from ranking paradigm).

Under the Brams–Fishburn tally procedure, any chosen subset S contributes towards the entire candidate pool a total vote count of $\sum_{i=1}^{d} v_i^{\mathrm{BF}}(S) = \sum_{i=1}^{d} m_i(S) = |S|$. In Borda scoring, any chosen linear order contributes towards the candidate pool a total count of $\sum_{i=1}^{d} i = d(d+1)/2$. To truly equate vote distribution under the approval voting procedure with that under Borda scoring (of the induced voters profile), one would use a modified scoring rule $\mathbf{v}^{\mathrm{Mod}}(S) = [v_1^{\mathrm{Mod}}(S), \ldots, v_d^{\mathrm{Mod}}(S)]^T$, which may give some sense of balancing the contributions from subsets of different set-size (since a voter may be contented with the idea that the same amount of total effective votes will be cast regardless how many candidate he/she chooses):

$$\mathbf{v}^{\mathrm{Mod}}(S) = \frac{d}{2}\begin{bmatrix} m_1(S) \\ m_2(S) \\ \vdots \\ m_d(S) \end{bmatrix} + \frac{d-|S|+1}{2}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{32}$$

This rule, in comparison to the Brams–Fishburn score

$$\mathbf{v}^{\mathrm{BF}}(S) = \begin{bmatrix} m_1(S) \\ m_2(S) \\ \vdots \\ m_d(S) \end{bmatrix}, \tag{33}$$

makes $\sum_{i=1}^{d} v_i^{\mathrm{Mod}} = d(d+1)/2$, regardless of the number of candidates ($|S| = k$) being approved of. Under this situation, the Brams–Fishburn score on the AV ballots is *identical* to the Borda score on the latent voters profile (under equal-probability interpretation of each chosen subset).

The SI model, on the other hand, may also induce a ranking probability distribution under very stringent conditions (stated as conditions for AV Polytope). When such a ranking probability does exist, the Borda score on each candidate can be calculated using a formula developed in Regenwetter and Grofman (1998a, Theorem 5), which is reproduced here

$$v_i^{\mathrm{SI}} = \sum_{S \subseteq [d]} \frac{P_S}{f(|S|)} m_i(S). \tag{34}$$

The contribution to the Borda scores by $S \subseteq [d]$ is (for $f(|S|) \neq 0$)

$$\mathbf{v}^{\mathrm{SI}}(S) = \frac{1}{f(|S|)}\begin{bmatrix} m_1(S) \\ m_2(S) \\ \vdots \\ m_d(S) \end{bmatrix}, \tag{35}$$

and $\mathbf{v}^{\mathrm{SI}}(S) = 0$ if $f(|S|) = 0$.[12] Compared with (23), it is obvious that the only (and significant) difference between the Brams–Fishburn score (equal-probability model) and the SI model (SI model) is with respect to the weights individual subsets carry. In the former case (33), a candidate's appearing (or non-appearing) in a subset constitutes a vote of 1 count (or 0 count), whereas in the latter case (35) a candidate's appearing (or non-appearing) constitutes $\frac{1}{f(|S|)}$ count (or 0 count). The SI score appears to have given too much weight to those voters who happen to choose an unpopular set-size

[12] In Theorem 5 of Regenwetter and Grofman (1998a), the summation does not include the $S = [d]$ term. However, they define Borda scores using $[d-1, d-2, \ldots, 0]^T$ as weighting function, rather than $[d, d-1, \ldots, 1]^T$ used in this article, see footnote 4. As a result, (35) is a valid formula for calculating Borda scores under the current definition so long as the summation over S includes the master set $[d]$.

(small $f(|S|)$), whereas the Brams–Fishburn score gives equitable contribution for each voter.

When the subset probability P_S falls within the core of the AV polytope, then the latent rank probability distribution P_π (voters profile) is consistent with both the SI interpretation and the equal-probability interpretation. In this case, the Brams–Fishburn tally procedure and the SI tally procedure yield the same score—the Borda score over one and the same P_π. This occurs when the subset size is truly irrelevant, see Eq. (31) under Remark 2 (Section 2.4.5).

2.5. Connection to the rank-position probability

Finally, for completeness of this exposition, the connection between the permutahedron and the rank-matching paradigm will be discussed. A rank-matching (also called rank-assignment) paradigm involves the assignment of rank-positions $1, 2, \ldots, d$ to the d candidates. Let b_{ij} be the amount (fraction) of jth rank-position assigned to candidate i, which satisfies (i) $b_{ij} \geqslant 0$ (positivity of assignment); (ii) $\sum_i b_{ij} = 1$ (each rank-position fully spent); (iii) $\sum_j b_{ij} = 1$ (each candidate fully assigned). The b_{ij}'s (for $i, j \in [d]$) form elements of a $d \times d$ *bistochastic matrix* $\mathbf{B}_d$.

Among all bistochastic matrices, of special interest is the class of *permutation matrices* Ω_π, whose elements assume the value of either 0 or 1, with exactly one 1 in each row and in each column. The permutation matrix Ω_π is in one-to-one correspondence with the linear order π, with its ijth entry

$$(\Omega_\pi)_{ij} = \begin{cases} 1 & \text{if } \pi(i) = j, \\ 0 & \text{otherwise.} \end{cases}$$

The convex hull of all permutation matrices (i.e., when π exhausts the set of linear orders $\mathscr{L}_d$) forms a polytope, known as the *Birkhoff Polytope* (alternatively called *Assignment Polytope* or *Perfect Matching Polytope*):

$$\begin{aligned} \mathscr{B}_d &= \operatorname{conv}\{\Omega_\pi\} \\ &= \left\{ \lambda_1 \Omega_{\pi_1} + \lambda_2 \Omega_{\pi_2} + \cdots + \lambda_{d!} \Omega_{\pi_{d!}} : \pi_k \in \mathscr{L}_d, \lambda_k \geqslant 0, \right. \\ &\qquad \left. \sum_{k=1}^{d!} \lambda_k = 1 \right\}. \end{aligned} \tag{36}$$

Examples of permutation matrices are shown as follows (for $d = 3$):

$$\Omega_{\langle 123 \rangle} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \Omega_{\langle 132 \rangle} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\Omega_{\langle 213 \rangle} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \Omega_{\langle 231 \rangle} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix},$$

$$\Omega_{\langle 312 \rangle} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \Omega_{\langle 321 \rangle} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Since permutation matrices (Ω_π) are bistochastic matrices $\mathbf{B}_d$ in their extreme forms, one might ask whether every bistochastic matrix can be represented in this way (36), i.e., as a convex combination of permutation matrices (and therefore corresponds to a point in the Birkhoff Polytope). Since the ranking probability P_π itself can be viewed as the coefficients for this convex combination, an equivalent statement of this characterization problem, in the language of choice paradigms, is

$$b_{ij} = \sum_{\pi \in \mathscr{L}_d} P_\pi \, (\Omega_\pi)_{ij} = \sum_{\{\pi \in \mathscr{L}_d \,:\, \pi(i) = j\}} P_\pi. \tag{37}$$

Given P_π, one readily obtains b_{ij} through marginalization. Given b_{ij}'s, are there any constraints required of b_{ij} for such P_π to exist (as in the characterization problems for binary choice and for subset choice)?

It turns out that the answer is "no": any bistochastic matrix (rank-position probability b_{ij}) can be represented as a ranking probability P_π where the rankings are represented by permutation matrices; this is commonly known as the Birkhoff/von Neumann Theorem, proven independently by Birkhoff (1946) and von Neumann (1953) under different contexts. Therefore, all bistochastic matrices (rank-position probabilities) are well "characterized"—the set of all bistochastic matrices *is* the Birkhoff Polytope, with permutation matrices (linear orders) as its vertices.[13] Also with $d!$ vertices, the dimensionality of $\mathscr{B}_d$ equals $(d-1)^2$, which is determined as follows:

$$\underbrace{d^2}_{\substack{\text{degree of freedom for a} \\ d \times d \text{ square matrix}}} - \underbrace{2d}_{\substack{\text{constraints on each} \\ \text{row and each column}}} + \underbrace{1}_{\substack{\text{over-counting constraints since} \\ \text{each row/column sums to 1}}}.$$

Properties of $\mathscr{B}_d$ have been studied and documented in great mathematical detail (Brualdi & Gibson, 1977a, b, c, 1976); in particular, all of its d^2 facets have been characterized (Balinski & Russakoff, 1974). The Birkhoff Polytope has been introduced into the choice probability literature by Suck (1992) in connection with the study of the Binary Choice Polytope.

Permutation matrices and permutation vectors (introduced in Section 2.1.1) are two representations of the set of linear orders $\mathscr{L}_d$, with the former resulting in the Birkhoff Polytope $\mathscr{B}_d$ and the latter the permutahedron

[13] Thus, the situation with rank-position probabilities b_{ij} here is quite different from the binary choice case and from the subset choice case, where constraints on the binary choice probability or on the subset choice probability exist for each to induce a ranking probability distribution.

Π_{d-1}. Naturally, the two polytopes have close connections. In fact, it is well known that there exists a canonical projection $\gamma : R^{d^2} \to R^d$ from the Birkhoff Polytope $\mathscr{B}_d$ to the d-permutahedron Π_{d-1}. Denote $\mathbf{b} = [b_{11}, \ldots, b_{1d}, b_{21}, \ldots, b_{dd}]^T \in R^{d^2}$. The so-called *canonical projection*

$$\mathbf{v} = \mathbf{J}\mathbf{b}$$

with $\mathbf{J} \in R^{d^2 \times d}$ given by

$$\begin{bmatrix} 1 & 2 & \cdots & d & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 2 & \cdots & d & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 2 & \cdots & d \end{bmatrix}$$

gives rise to a point $\mathbf{v} \in \Pi_{d-1}$. In other words,

Proposition 2.10 (Permutahedron as canonical projection of the Birkhoff Polytope). *Any $d \times d$ bistochastic matrix $\mathbf{B}_d$ can be surjectively mapped onto the d-permutahedron via*:

$$\mathbf{B}_d \cdot [1, 2, \ldots, d]^T = \mathbf{v}^{\mathrm{Bk}} \in \Pi_{d-1}. \tag{38}$$

Proof. The following proof is adapted from Yemelichev et al. (1984, p. 229). All one needs to show is that the coordinates

$$v_i^{\mathrm{Bk}} = \sum_{j=1}^{d} j b_{ij}, \quad i = 1, 2, \ldots, d, \tag{39}$$

satisfy the system of constraints defining the d-permutahedron (in Theorem 2.1). Clearly,

$$\begin{aligned} \sum_{i=1}^{d} v_i^{\mathrm{Bk}} &= \sum_{i=1}^{d} \sum_{j=1}^{d} j b_{ij} \\ &= \sum_{j=1}^{d} j \cdot 1 \quad \left(\text{since } \sum_{i=1}^{d} b_{ij} = 1\right) \\ &= \frac{d(d+1)}{2}, \end{aligned}$$

hence proving (15). Now, for any subset $M \subset [d]$ and any integer $j \in [d]$, denote $h_j^M = \sum_{i \in M} b_{ij}$ as a partial sum of the elements in the jth column. Clearly $0 \leqslant h_j^M \leqslant 1$ and $0 \leqslant \sum_{j=1}^{d} h_j^M = \sum_{i \in M} \sum_{j=1}^{d} b_{ij} = |M|$. Therefore,

$$\begin{aligned} &\sum_{i \in M} v_i^{\mathrm{Bk}} - \sum_{j=1}^{|M|} j \\ &\quad = \sum_{j=1}^{d} j h_j^M - \sum_{j=1}^{|M|} j \\ &\quad = \sum_{j=|M|+1}^{d} j h_j^M - \sum_{j=1}^{|M|} j(1 - h_j^M) \\ &\quad \geqslant \sum_{j=|M|+1}^{d} |M| \cdot h_j^M - \sum_{j=1}^{|M|} |M| \cdot (1 - h_j^M) \\ &\quad = |M| \cdot \left(\sum_{j=1}^{d} h_j^M - |M| \right) \\ &\quad = 0. \end{aligned}$$

Therefore

$$\sum_{i \in M} v_i^{\mathrm{Bk}} \geqslant \sum_{j=1}^{|M|} j = \frac{|M|(|M|+1)}{2},$$

which is the system of inequalities (14) defining the permutahedron. □

Remark 1. This connection between the Birkhoff Polytope $\mathscr{B}_d$ and the permutahedron Π_{d-1} indicates that Borda scoring of a ranking probability P_π can be conducted in two separable steps: converting P_π into a rank-position probability b_{ij} (while still preserving the structure of linear orders), and then using Borda's point-assignment system, i.e., assigning j points to rank-position j, to construct Borda scores via (38). This is seen in the following identities

$$\begin{aligned} v_i^{\mathrm{Bd}} &= \sum_{\pi \in \mathscr{L}_d} P_\pi \, \pi(i) \quad \text{from (10)} \\ &= \sum_{j=1}^{d} \sum_{\{\pi : \pi(i) = j\}} P_\pi \cdot j \\ &= \sum_{j=1}^{d} j \cdot \left(\sum_{\{\pi : \pi(i) = j\}} P_\pi \right) \\ &= \sum_{j=1}^{d} j \cdot b_{ij} \quad \text{from (37)} \\ &= v_i^{\mathrm{Bk}}. \end{aligned}$$

In other words, the contribution of the jth rank-position is counted as having strength j, while the induced voting score for each candidate i is the sum of strengths of all rank-positions weighted by the fraction assigned for each rank-position. See Regenwetter and Grofman (1998a, p. 43). Just as $\mathbf{v}^{\mathrm{Bd}} = [v_1^{\mathrm{Bd}}, v_2^{\mathrm{Bd}}, \ldots, v_d^{\mathrm{Bd}}]^T$ minimizes (11), $\mathbf{v}^{\mathrm{Bk}} = [v_1^{\mathrm{Bk}}, v_2^{\mathrm{Bk}}, \ldots, v_d^{\mathrm{Bk}}]^T$ would, as pointed out in Cook and Seiford (1982), minimize the expression

$$\sum_{i=1}^{d} \sum_{j=1}^{d} b_{ij} (v_i - \pi(j))^2$$

among all $\mathbf{v} = [v_1, v_2, \ldots, v_d]^T \in R^d$.

3. Conclusions and discussions

Permutahedron, as the space of Borda scores of a probability distribution on the set of rankings (linear orders) over d candidates, is a useful tool in linking various choice paradigms: (i) It can be realized as a "zonotope", i.e., a projection from a cube, of dimension

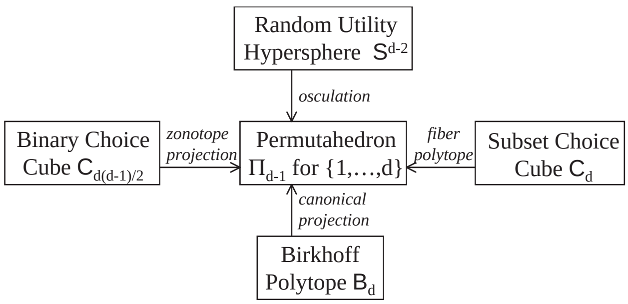

Fig. 5. Connection of a d-permutahedron, the space of Borda scores of candidates given a ranking probability, to the $d(d-1)/2$-cube, the d-cube, and the $(d-2)$-hypersphere, representing, respectively, the binary choice probability, the subset choice probability, and the random distribution of utility values of the d candidates (cf. Fig. 1).

$d(d-1)/2$, that describes the binary choice probability; (ii) It can be realized as a "monotone path polytope" associated with projecting a cube, of dimension d, that describes subset choice probability, onto a line segment that describes the subset size; (iii) It can be realized as a canonical projection from the Birkhoff Polytope that describes bistochastic matrices, with $d \times d$ elements, that relates to the rank-position probability; (iv) Finally, all of its vertices are contained in a hypersphere, with dimension $d-2$, which is the quotient space of all equivalent classes of interval-scaled d-dimensional utility vectors. See Fig. 5 for a summary.

These properties of a permutahedron supply us with a wealth of connections to various choice paradigms, namely, those of binary choice, subset choice, ranking, and rank-matching, that are commonly employed in psychology, economics, political science, consumer behavior, etc. Important insights gained from this approach include: the extension of Young's formula of Borda scores to all cases of binary choice probabilities (from only those induced from a ranking probability), as well as its compatibility with the BTL representation of choice probability; the connection between the AV tally procedure à la Brams and Fishburn (1983) and the equal-probability model of the induced voters profile; the condition for the compatibility of the Brams–Fishburn tally procedure with the size-independent model à la Falmagne and Regenwetter (1996); and the spherical representation of the random utility model of choice, in a way that illuminates the intimate connection between interval-valued utility vectors and linear orders. Most importantly, the notion of Borda scores can be extended from the scoring of a ranking probability (Eq. (10)) to the scoring of a binary choice probability (Eq. (17)), the scoring of a subset choice probability (Eq. (28)) under the equal-probability model under the SI model (Eq. (34)) and the scoring of a rank-position probability (Eq. (39)).

The permutahedron-based approach may also prove to be useful in aggregation of preference or social choice. The pivotal result in social choice and welfare is Arrow's impossibility theorem, which sets fundamental limits on any well-behaved social welfare ordering. In recent years, the essence of this impossibility of "proper" (i.e., unanimous, anonymous, and continuous) aggregation of preference, as well as the homotopic equivalence from the Pareto rule to a dictator rule, has been understood from a deep, topological perspective (Chichilnisky, 1980, Chichilnisky & Heal, 1983, Baryshnikov, 1993, Heal, 1997, Baryshnikov, 2000, Lauwers, 2000). Central to these arguments is the topological notion of non-contractibility of the space of preference, an example being the space of non-coincidental random utilities under the spherical representation (see Section 2.3.1). When null preference is allowed, the space of affinely equivalent utility vectors also includes the origin (the barycenter of the permutahedron). Jones, Zhang, & Simpson (2003) have investigated the natural topology (which turns out to be non-Hausdorff) that includes this null-preference point as a connected component in the space of preference. There, it is shown that proper aggregation is possible if and only if the null-preference is allowed for the society (output of the aggregation map) but not for the individual voters (input to the aggregation map). A natural question is that whether this conclusion can be generalized to other preference spaces (of binary choice, subset choice, etc). It is known (Baryshnikov, 1993) that the nerve of the covering of the set of linear orders on $[d]$ by pairwise (binary) comparisons, which is a simplicial complex of dimension $(d+1)(d-2)/2$, is homotopically equivalent to the sphere S^{d-2}. It is unclear whether a covering by variable-sized subsets enjoys similar properties. Future research will illuminate the mathematical structure of and connections between the various spaces of preference and choice.

Appendix A. Mathematical background on polytopes

This appendix contains basic materials about the convex polytope, including its dualistic characterizations either using the set of its vertices or by the set of its facets. Special attention will be paid to the projection and to the lift-up of projections of polytopes, where permutahedra may arise. For details, see Ziegler (1995).

A.1. Polytopes

A *convex polytope* ("polytope") is the convex hull of a finite set $\mathbf{P}$ of points (vectors) $\mathbf{P} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in

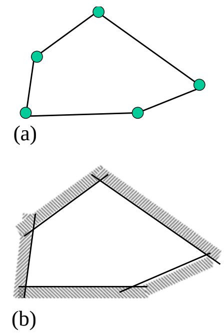

Fig. 6. A polytope defined as (a) the convex hull of a finite set of points; or equivalently as (b) the enclosure of finite number of intersecting half-spaces.

some real vector space $R^d \ni \boldsymbol{x}_i (i = 1, 2, \ldots, n)$:

$$\mathscr{P} = \mathrm{conv}(\mathbf{P}) = \left\{ \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \cdots + \lambda_n \mathbf{x}_n : \text{ all } \lambda_i \geqslant 0; \sum_{i=1}^{n} \lambda_i = 1 \right\}.$$

Assuming that none of the $\mathbf{x}_i$'s can be written as a convex combination of the other points (otherwise, they can always be excluded from $\mathbf{P}$ to yield a reduced set of points), such a set of n distinct points or vectors form the *vertices* of the polytope $\mathscr{P}$. When all vectors in $\mathbf{P}$ are in general positions in R^d (i.e., no d of them lie on a common affine hyperplane), then the polytope is called a simplicial polytope.

On the other hand, one can also define a polytope as an enclosure (i.e., intersection that is bounded in space) of finitely many closed half-spaces in R^d:

$$\mathscr{P} = \{\mathbf{x} \in R^d : \mathbf{J}\,\mathbf{x} \leqslant \mathbf{b}, \mathbf{J} \in R^{m \times d}, \mathbf{b} \in R^m\}.$$

Each half-space is governed by one linear inequality (with equality sign corresponding to a hyperplane), and a finite collection of them (indicated by the matrix J above), if unbounded, would form a polyhedron. When all the defining hyperplanes are in general positions, the polytope is called a simple polytope. These two definitions of a polytope, as a linear combination of points and as the bounded intersection of half-spaces (see Fig. 6), reflect the fundamental duality between a linear space and its dual space induced by the inner-product operation.

Familiar examples of polytopes include: (i) polygons, which are polytopes with dimension $d = 2$ (e.g., trapezoid, pallellogram); (ii) the d-dimensional hypercube ("d-cube"), which is the d-dimensional extension of a cube; (iii) the d-dimensional simplex, with $d + 1$ vertices; here one of the vertices $\mathbf{x}_{d+1} = [0, 0, \ldots, 0]^T$ is the origin, and the other d vertices are $\mathbf{x}_k = [\underbrace{0, \ldots, 0}_{k-1}, 1, \underbrace{0, \ldots, 0}_{d-k}]^T$, $k = 1, 2, \ldots, d$.

Several polytopes have received considerable interest in recent years in the choice literature. They are (i) the Binary Choice Polytope, resulting from the characterization of binary choice vectors compatible with a ranking probability; (ii) the AV polytope, resulting from the characterization of the subset choice probability (of the SI model) as compatible with a ranking probability; and (iii) the Birkhoff Polytope, arising from the rank-assignment paradigm and bistochastic matrices.

A.2. Faces of a polytope

The dimension, $\dim \mathscr{P}$, of a polytope $\mathscr{P}$ is the dimension of its affine span. A polytope contains faces which form its boundaries. The face of a d-dimensional polytope can take the form of: (i) a vertex, which is of zero-dimension (i.e., a point); (ii) an edge, which is one-dimensional (i.e., a line segment); (iii) a facet, which has dimension $d - 1$ (i.e., the maximal proper face); or (iv) any dimension in between. In fact, any proper face $\mathscr{F} \subset \mathscr{P}$ can be expressed as the convex combination of a subset $\mathbf{F}$ of the original set $\mathbf{P}$ of vectors making up $\mathscr{P}$

$$\mathscr{F} = \mathrm{conv}\{\mathbf{x}_i : \mathbf{x}_i \in \mathbf{F} \subset \mathbf{P}\}.$$

The collection of all faces of a polytope, ordered by inclusion, forms a lattice, called the *face lattice*. The length of the maximal chain of this lattice plus one equals the dimension of the polytope. For a d-dimensional simple polytope, every vertex belongs to d facets, whereas for a d-dimensional simplicial polytope, every facet has d vertices and hence are all simplices. Simple and simplicial polytopes are dual to each other. Any d-dimensional polytope ($d \geqslant 3$) that is both simple and simplicial must be a simplex.

From the combinatoric point of view, a polytope is completely characterized by its face-lattice—two polytopes are said to be combinatorially equivalent if their face lattices are isomorphic. For a d-dimensional polytope $\mathscr{P}$, denote the total number of its j-dimensional faces as $f_j(\mathscr{P}), (j = d-1, d-2, \ldots, 1, 0)$, and define $f_d(\mathscr{P}) = 1$ for convenience. These numbers collectively form a $(d+1)$-dimensional vector called the f-vector (some authors also define $f_{-1}(\mathscr{P}) = 1$, thus making the f-vector $(d+2)$-dimensional). The components of the f-vector of a polytope satisfy some numerical relationships. The most prominent one is the Euler–Poincaré relation

$$\sum_{j=0}^{d} (-1)^j f_j(\mathscr{P}) = 1,$$

which reduces, as a special case for $d = 3$, to the famous Euler formula between the number of vertices (f_0), the number of edges (f_1), and the number of facets (f_2) for any convex polytope in three dimensions:

$$f_0 - f_1 + f_2 = 2.$$

Interestingly, for simplicial polytopes, the Euler–Poincaré relation itself turns out to be a special case ($k = d$) of the so-called Dehn–Sommerville relations:

$$\sum_{j=0}^{k} (-1)^j \binom{d-j}{d-k} f_j(\mathscr{P}) = f_k(\mathscr{P}),$$

which holds for $k = 0, 1, 2, \ldots, d$ (only half of these relations are linearly independent). Moreover, components of the f-vector may be constrained within certain upper and lower bounds (e.g., the Upper Bound Theorem, McMullen, 1971). More details can be found elsewhere (see Bayer & Lee, 1993, Ziegler, 1995).

A.3. Projection of polytopes

An interesting property of a polytope that follows immediately from its definition concerns its affine projection. A projection $\gamma : \mathscr{P} \to \mathscr{Q}$ of a polytope is defined as an affine map $\gamma : R^p \to R^q$,

$$\mathbf{x} \mapsto \mathbf{J}\mathbf{x} + \mathbf{b},$$

where $\mathscr{P} \subset R^p$ is a p-dimensional polytope, $\mathscr{Q} \subset R^q$ is a q-dimensional polytope, and $\gamma(\mathscr{P}) = \mathscr{Q}$. Here $\mathbf{J} \in R^{q \times p}$ specifies a $q \times p$ matrix, $\mathbf{b} \in R^q$ is a vector of dimension q, and $\mathbf{x} \in R^p$ is a vector of dimension p. Clearly, $\dim(\mathscr{Q}) \leqslant \dim(\mathscr{P}) = p$, with equality holding when the rank of $\mathbf{J}$ is p. From the linear algebra behind the affine map, for any face $\mathscr{F} \subset \mathscr{Q}$, its pre-image

$$\gamma^{-1}(\mathscr{F}) = \{y \in \mathscr{P} : \gamma(y) \in \mathscr{F}\}$$

is a face of $\mathscr{P}$. Furthermore, if $\mathscr{F}_1$, $\mathscr{F}_2$ are faces of $\mathscr{Q}$, then $\mathscr{F}_1 \subseteq \mathscr{F}_2$ holds if and only if $\gamma^{-1}(\mathscr{F}_1) \subseteq \gamma^{-1}(\mathscr{F}_2)$. That is, the inclusion relation between faces in the projected polytope (of a lower dimension) can be "traced" back to their inclusion relation in the original polytope (of a higher dimension). See Fig. 7.

An interesting case is the projection of a cube, which yields a special kind of polytope called a *zonotope*. A zonotope $\mathscr{Z}$ is the image of a cube under an affine projection $\gamma : R^p \supset \mathscr{C}_p \to \mathscr{Z} \subset R^q$

$$\mathscr{Z} = \{\mathbf{z} : \mathbf{z} = \mathbf{J}\mathbf{x} + \mathbf{b}, \mathbf{x} \in \mathscr{C}_p\}.$$

An example is given by Fig. 3 where the two-dimensional zonotope arises as the projection of $\mathscr{C}_3$. Expressing the p-cube $\mathscr{C}_p$ in 0-centered coordinates, one has

$$\mathscr{C}_p = \left\{ \mathbf{x} : \mathbf{x} = \sum_{l=1}^{p} t_l \boldsymbol{\varepsilon}_l ; -\frac{1}{2} \leqslant t_l \leqslant \frac{1}{2}, \quad l = 1, 2, \ldots, p \right\} \tag{A.1}$$

where $\boldsymbol{\varepsilon}_l$'s $(l = 1, 2, \ldots, p)$ are a set of p Cartesian base vectors in R^p. The matrix $\mathbf{J}$ can be expressed in terms of its column vectors: $\mathbf{J} = [\mathbf{j}_1, \mathbf{j}_2, \ldots, \mathbf{j}_p]$ where each of the $\mathbf{j}_l$ $(l = 1, 2, \ldots, p)$ is a q-dimensional vector. Since $\boldsymbol{\varepsilon}_l = [\underbrace{0, \ldots, 0}_{l-1}, 1, \underbrace{0, \ldots, 0}_{p-l}]^T$ has a 1 only in its lth position and has 0 elsewhere, $\mathbf{J}\,\boldsymbol{\varepsilon}_l = \mathbf{j}_l$, so

$$\mathscr{Z} = \left\{ \mathbf{z} : \mathbf{z} = \mathbf{b} + \sum_{l=1}^{p} t_l \mathbf{j}_l ; -\frac{1}{2} \leqslant t_l \leqslant \frac{1}{2}, \ l = 1, 2, \ldots, p \right\}. \tag{A.2}$$

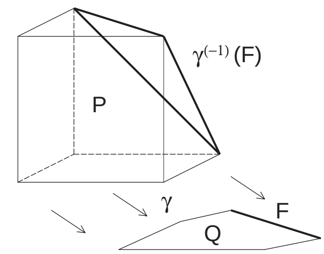

Fig. 7. Affine projection between polytopes (note the tracing of a face of a lower-dimensional polytope to that of a higher-dimensional one).

Being a zonotope is a geometric property rather than a combinatorial one. Zonotopes are centrally symmetric with respect to its barycenter. Affine projection of a zonotope is still a zonotope; faces of a zonotope is again a zonotope. Further properties of zonotopes are reviewed in McMullen (1971).

A.4. Fiber polytopes

Two concepts are intimately associated with projections, namely, "sections" and "fibers". Let $\gamma : \mathscr{P} \to \mathscr{Q}$ be a projection of polytopes with $\mathscr{P} \subset R^p$, $\mathscr{Q} \subset R^q$, such that $\gamma(\mathscr{P}) = \mathscr{Q}$, with $\dim(\mathscr{Q}) < \dim(\mathscr{P})$. A *section* is a (continuous) map $\sigma : \mathscr{Q} \to \mathscr{P}$ that satisfies $\gamma(\sigma(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathscr{Q}$. Informally, since any projection maps many points of a higher dimensional space $\mathbf{y} \in R^p$ to one single point of a lower dimensional space $\mathbf{x} \in R^q$ $(p > q)$, the set of pre-image points (i.e., $\mathbf{y}$'s) that project to the same point $\mathbf{x} \in R^q$ form a *fiber* $\gamma^{-1}(\mathbf{x}) = \{\mathbf{y} \in R^p : \gamma(\mathbf{y}) = \mathbf{x}\}$ that is attached, or growing out of $\mathbf{x}$ on R^q, the base manifold. A (horizontal) section σ is then an inverse map (not unique, of course) obtained by selecting one point in each fiber on R^q. The collection of fibers on a base manifold is called the *fiber bundle*; it allows various horizontal sectioning.

For any given section σ, one can construct the integral

$$\int_{\mathscr{Q}} \sigma(\mathbf{x})\, d\mathbf{x},$$

where the vector-valued integral is with respect to points of the base manifold $\mathbf{x} \in \mathscr{Q}$; thus the integral itself is a point in R^{p-q}. The fiber polytope $\Sigma(\mathscr{P}, \mathscr{Q})$ is defined as the set of all such points, each obtained through a sectioning σ:

$$\Sigma(\mathscr{P},\mathscr{Q}) = \left\{\frac{1}{\mathrm{vol}(\mathscr{Q})}\int_{\mathscr{Q}} \sigma(\mathbf{x})\,d\mathbf{x}\ :\sigma \text{ is a section associated with } \gamma\right\}, \tag{A.3}$$

where the volume of $\mathscr{Q}$ is defined as

$$\mathrm{vol}(\mathscr{Q}) = \int_{\mathscr{Q}} d\mathbf{x}.$$

It can be shown that $\Sigma(\mathscr{P},\mathscr{Q})$ indeed form a convex set. Though an object defined in R^p, its dimensionality is only $p-q$; it is contained in the fiber growing out of the point $\mathbf{r}_0 \in \mathscr{Q}$:

$$\Sigma(\mathscr{P},\mathscr{Q}) \subseteq \gamma^{-1}(\mathbf{r}_0) \cap \mathscr{P},$$

where the point $\mathbf{r}_0$ is the barycenter of $\mathscr{Q}$:

$$\mathbf{r}_0 = \frac{1}{\mathrm{vol}(\mathscr{Q})}\int_{\mathscr{Q}} \mathbf{x}\,d\mathbf{x}.$$

More materials on fiber polytopes can be found in Billera and Sturmfels (1992).

A special kind of fiber polytope is the so-called monotone path polytope, constructed through an interesting projection map. Here the projection γ is given as follows: choose a fixed but otherwise arbitrary p-dimensional row vector $\mathbf{c}$, and form a dot product with any column vector $\mathbf{x} \in \mathscr{P} \subset R^p$. The set of points

$$\mathscr{Q} = \{\mathbf{c}\cdot\mathbf{x}\ :\ \mathbf{x}\in\mathscr{P}\}$$

define a one-dimensional polytope $\mathscr{Q} = [c_{\min}, c_{\max}] \subseteq R^1$ such that

$$c_{\min} = \min_{\mathbf{x}\in\mathscr{P}} \mathbf{c}\cdot\mathbf{x}, \quad c_{\max} = \max_{\mathbf{x}\in\mathscr{P}} \mathbf{c}\cdot\mathbf{x}.$$

With such $\mathscr{Q}$, the corresponding fiber polytope $\Sigma(\mathscr{P},\mathscr{Q})$ is the *monotone path polytope*. The coordinates for a particular section $\sigma : \mathscr{Q} \ni x \to \sigma(x) \in \mathscr{P}$ is

$$\frac{1}{c_{\max} - c_{\min}} \int_{c_{\min}}^{c_{\max}} \sigma(x)\,dx.$$

References

Alvo, M., & Cabilio, P. (1993). Rank correlations and the analysis of rank-based experimental designs. In M. Fligner, & J. Verducci (Eds.), *Probability models and statistical analyses for ranking data* (pp. 141–156). New York: Springer.

Alvo, M., Cabilio, P., & Feigin, P. D. (1982). Asymptotic theory for measures of concordance with special reference to average Kendall tau. *Annals of Statistics*, *10*, 1269–1276.

Babington-Smith, B. (1950). Discussion of Professor Ross's paper. *Journal of the Royal Statistical Society, Series B*, *12*, 53–56.

Balinski, M. L., & Russakoff, A. (1974). On the assignment polytope. *SIAM Review*, *16*, 516–525.

Baryshnikov, Y. (1993). Unifying impossibility theorems: A topological approach. *Advances in Applied Mathematics*, *14*, 404–415.

Baryshnikov, Y. (2000). On isotopic dictators and homological manipulators. *Journal of Mathematical Economics*, *33*, 123–134.

Bayer, M. M., & Lee, C. W. (1993). Combinatorial aspects of convex polytopes. In P. M. Gruber, & J. M. Wills (Eds.), *Handbook of convex geometry*. Amsterdam: Elsevier.

Billera, L. J., & Sturmfels, B. (1992). Fiber polytopes. *Annals of Mathematics*, *135*, 527–549.

Birkhoff, G. (1946). Tres observaciones sobre el algebra lineal, Revista Facultad de Ciencias Exactas. *Puras y Applicadas Universidad Nacional de Tucuman, Serie A (Matematicas y Fisica Teoretica)*, *5*, 147–151.

Block, H. D., & Marschak, J. (1960). Random ordering and probabilistic theories of responses. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hostelling*. Stanford, CA: Stanford University Press.

Bolotashvili, G., Kovalev, M., & Girlich, E. (1999). New facets of the linear ordering polytope. *SIAM Journal of Discrete Mathematics*, *12*, 326–336.

Bowman, V. J. (1972). Permutation polyhedra. *SIAM Journal of Applied Mathematics*, *22*, 580–589.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs. I. *Biometrika*, *39*, 324–345.

Brams, S. J., & Fishburn, P. C. (1978). Approval voting. *American Political Science Review*, *72*, 831–847.

Brams, S. J., & Fishburn, P. C. (1983). *Approval voting*. Boston, MA: Birkhäuser.

Brualdi, R. A., & Gibson, P. M. (1976). Convex polyhedra of doubly stochastic matrices. IV. *Linear Algebra and Its Applications*, *15*, 153–172.

Brualdi, R. A., & Gibson, P. M. (1977a). Convex polyhedra of doubly stochastic matrices. I. Applications of the permanent function. *Journal of Combinatorial Theory (A)*, *22*, 194–230.

Brualdi, R. A., & Gibson, P. M. (1977b). Convex polyhedra of doubly stochastic matrices. II. Graph of Σ_n. *Journal of Combinatorial Theory (B)*, *22*, 175–198.

Brualdi, R. A., & Gibson, P. M. (1977c). Convex polyhedra of doubly stochastic matrices. III. Affine and combinatorial properties of Ω_n. *Journal of Combinatorial Theory (A)*, *22*, 338–351.

Chichilnisky, G. (1980). Social choice and the topology of spaces of preferences. *Advances in Mathematics*, *37*, 165–176.

Chichilnisky, G., & Heal, G. (1983). Necessary and sufficient conditions for a resolution of the social choice paradox. *Journal of Economic Theory*, *31*, 68–87.

Cohen, M., & Falmagne, J. C. (1990). Random utility representation of choice probability: a new class of necessary conditions. *Journal of Mathematical Psychology*, *34*, 88–94.

Cook, W. D., & Kress, M. (1992). *Ordinal information and preference structures: Decision models and applications*. Englewood Cliffs, NJ: Prentice Hall.

Cook, W. D., & Seiford, L. M. (1982). On the Borda-Kendall consensus method for priority ranking problems. *Management Science*, *28*, 621–637.

Cook, W. D., & Seiford, L. M. (1990). A data envelopment model for aggregating preference ranking. *Management Science*, *36*, 1302–1310.

Copeland, A.H. (1951). A "reasonable" social welfare function. Mimeograph, University of Michigan.

Critchlow, D. E., Fligner, M. A., & Verducci, J. S. (1991). Probability models on ranking. *Journal of Mathematical Psychology*, *35*, 294–318.

Doignon, J. P., & Fiorini, S. (to appear). The facets and the symmetries of the approval-voting polytope, *Journal of Combinatorial Theory (Series B)*, to appear.

Doignon, J. P., & Regenwetter, M. (1997). An approval-voting polytope for linear orders. *Journal of Mathematical Psychology*, *41*, 171–188.

Doignon, J. P., & Regenwetter, M. (2002). On the combinatorial structure of AV polytope. *Journal of Mathematical Psychology*, *46*, 554–563.

Falmagne, J.-Cl., & Regenwetter, M. (1996). A random utility model for approval voting. *Journal of Mathematical Psychology*, *40*, 152–159.

Feigin, P. D., & Alvo, M. (1986). Intergroup diversity and concordance for ranking data: An approach via metrics for permutations. *Annals of Statistics*, *14*, 691–707.

Feigin, P. D., & Cohen, A. (1978). On a model for concordance between judges. *Journal of the Royal Statistical Society, Series B*, *40*, 203–213.

Fishburn, P. C. (1992). Induced binary probabilities and the linear ordering polytope—A status-report. *Mathematical Social Sciences*, *12*, 67–80.

Fligner, M. A., & Verducci, J. S. (1986). Distance based ranking models. *Journal of the Royal Statistical Society, Series B*, *48*, 359–369.

Gaiha, P., & Gupta, S. K. (1977). Adjacent vertices on a permutahedron. *SIAM Journal of Applied Mathematics*, *32*, 323–327.

Hardy, G., Littlewood, J. E., & Pólya, G. (1934). *Inequality*. Cambridge, UK: Cambridge University Press.

Heal, G. (1997). *Topological social choice*. New York: Springer.

Jones, M., Zhang, J., & Simpson, G. (2003). Aggregation of utility and social choice: A topological characterization. *Journal of Mathematical Psychology*, *47*, 545–556.

Kendall, M. G. (1962). *Rank correlation methods* (3rd ed.). London: Griffin.

Koppen, M. (1991). Random utility representations of binary choice probabilities. In J. P. Doignon, & J.-C. Falmagne, (Eds.), *Mathematical psychology: Current developments* (pp. 181–201). New York: Springer.

Koppen, M. (1995). Random utility representation of binary choice probabilities—Critical graphs yielding critical necessary conditions. *Journal of Mathematical Psychology*, *39*, 21–39.

Lauwers, L. (2000). Topological social choice. *Mathematical Social Sciences*, *40*, 1–39.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter, (Eds). *Handbook of mathematical psychology*, Vol. III (pp. 249–410). New York: Wiley.

Mallows, C. L. (1957). Non-null ranking models. I. *Biometrika*, *44*, 114–130.

Marley, A. A. J. (1992). A selective review of recent characterizations of stochastic choice models using distribution and functional equation techniques. *Mathematical Social Sciences*, *23*, 5–29.

McCullagh, P. (1993). Models on spheres and models for permutations. In M. Fligner, & J. Verducci (Eds.), *Probability models and statistical analyses for ranking data* (pp. 278–283). New York: Springer.

McMullen, P. (1971). On zonotopes. *Transactions of the American Mathematical Society*, *159*, 91–109.

Merlin, V. R., & Saari, D. G. (1997). The Copeland method II: manipulation, monotonicity, and paradoxes. *Journal of Economic Theory*, *72*, 148–172.

Nitzan, S., & Rubinstein, A. (1981). A further characterization of Borda ranking method. *Public Choice*, *36*, 153–158.

Ovchinnikov, S. (1996). Means on ordered sets. *Mathematical Social Sciences*, *32*, 39–56.

Pekeč, A. (2000). Computation complexity issues in random utility modeling. Paper presented at the Random Utility 2000 Conference, Fuqua School of Business, Duke University.

Rado, R. (1952). An inequality. *Journal of the London Mathematical Society*, *27*, 1–6.

Regenwetter, M., & Grofman, B. (1998a). Choosing subsets: A size-independent probabilistic model and the quest for a social welfare ordering. *Social Choice Welfare*, *15*, 423–443.

Regenwetter, M., & Grofman, B. (1998b). Approval voting, Borda winners, and Condorcet winners: Evidence from seven elections. *Management Science*, *44*, 520–533.

Regenwetter, M., Marley, A. A. J., & Joe, H. (1998). Random utility threshold models of subset choice. *Australian Journal of Psychology*, *50*, 165–174.

Regenwetter, M., & Marley, A. A. J. (2001). Random relations, random utilities and random functions. *Journal of Mathematical Psychology*, *45*, 864–912.

Regenwetter, M., Marley, A. A. J., & Grofman, B. (2002). A general concept of majority rule. *Mathematical Social Sciences*, *43*, 405–428.

Saari, D. G. (1990). The Borda dictionary. *Social Choice and Welfare*, *7*, 279–317.

Saari, D. G. (1992). Millions of election outcomes from a single profile. *Social Choice and Welfare*, *9*, 277–306.

Saari, D. G. (1993). *Geometry of voting*. New York: Springer.

Saari, D. G., & Merlin, V. R. (1996). The Copeland method I: Relationships and the dictionary. *Economic Theory*, *8*, 51–76.

Schoute, P.H. (1911). Analytic treatment of the polytopes regularly derived from the regular polytopes. Cited in Ziegler, G. M. (1995). Lectures on Polytope. New York: Springer (corrected second printing, 1998).

Suck, R. (1992). Geometric and combinatorial properties of the polytope of binary choice-probabilities. *Mathematical Social Sciences*, *23*, 81–102.

Suck, R. (1997). Polytopes in measurement and data analysis. Review of Ziegler's (1995) Lectures on Polytope. *Journal of Mathematical Psychology*, *41*, 299–303.

von Neumann, J. (1953). A certain zero-sum two-person game equivalent to the optimal assignment problem. In: H. W. Kuhn, & A. W. Tucker (Eds.), *Contributions to the theory of games*, Vol. II (pp. 5–12). Princeton, NJ: Princeton University Press. *Annals of Mathematical Studies, no. 28*.

Yemelichev, V. A., Kovalev, M. M., & Kravtsov, M. K. (1984). *Polytopes, graphs and optimization*. (translated by G.H. Lawden). Cambridge: Cambridge University Press.

Young, H. P. (1974). An axiomatization of Borda's rule. *Journal of Economic Theory*, *9*, 43–52.

Young, H. P. (1975). Social choice scoring functions. *SIAM Journal of the Applied Mathematics*, *28*, 824–838.

Ziegler, G. M. (1995). *Lectures on polytope*. New York: Springer (corrected second printing, 1998).

Ziegler, G.M. (1999). Lectures on 0/1-polytopes. In: G. Kalai, & G. M. Ziegler (Eds.), *Polytopes—combinatorics and computation*, DMV-Seminars Series. Basel: Birkhäuser-Verlag. Available: http://www.math.tu-berlin.de/~ziegler/.