Approval Voting Approach to Subset Selection

Peter C. Fishburn
Retired from AT&T Laboratories, Florham Park, NJ 07932

Saša Pekeč
Fuqua School of Business, Duke University, Durham, NC 27708, pekec@duke.edu

We propose a new class of subset-preference aggregation methods, based on appropriately modifying the approval voting procedure. Each voter selects a set of alternatives, indicating their approval of any subset that contains sufficiently many alternatives from the selected set. The subset approved by most voters is selected. We show that this approach both alleviates computational burden of reporting preferences over exponentially large number of subsets, while still accounting for possible synergetic subset values. Further, we also provide a characterization of this novel class of subset preference aggregation procedures.

1. Introduction

Selecting a collection of alternatives rather than choosing a single one is commonplace in many settings. Traditionally, this problem has been studied in the social choice literature in the context of selecting committees and subcommittees of legislative bodies (e.g., Francis, 1982; McElroy, 2006), but the (s)election of the "best" subset (committee, group, team) has many other natural applications, such as a company shareholders' electing and approving its corporate directors (e.g., Easterbrook and Fischel, 1983; Sjostrim and Kim, 2007; Cai et al., 2009; Yermack, 2010), a consulting firm's assigning consultants to a project team, or a sports team coach's deciding on the players for the starting line-up (e.g., Fitzpatrick and Askin, 2005; Feng et al., 2010).

In some cases, the selection is made by a single decision-maker who might consider multiple criteria. In many cases, however, the selection procedure has to aggregate preferences (opinions, votes) from a large number of stakeholders. In this paper, we focus on this aggregation problem, with the objective of maximizing the number of stakeholders who are satisfied by the chosen subset of alternatives. For example, the selected committee should be the one that is "approved" by the maximum number of voters. The voter approval of a committee could be expressed in various forms, e.g., from ensuring that a majority of committee members are approved by the voter to ensuring minimal representation by having at least one committee member approved by the voter.

We propose a new class of subset-preference aggregation methods, based on appropriately modifying the approval voting procedure (Brams and Fishburn, 2007, designed it for selecting a single alternative). We discuss aggregation properties of this class and contrast them to those of ad-hoc approaches which simply propose selecting top k performing alternatives according to some standard voting procedure. These properties indicate shortcomings of the latter approaches, as voters' preferences over subsets are ignored by design. Finally, we show that our class is unique under some conditions, i.e., we provide an axiomatization that characterizes this class of subset preference aggregation methods.

A large social choice theory literature is devoted exclusively to methods of aggregating preferences into a consensus choice. Its central results are impossibility theorems showing that no aggregation procedure can simultaneously satisfy some rather reasonable properties, in particular Arrow's Theorem (Arrow, 1951, 2012) and Gibbard-Satterhwaite Theorem (Gibbard, 1973; Satterthwaite, 1975). Some of the research in this field considers restrictions to avoid these impossibility results (e.g., restrictions on preference structures, Sen, 1984, 2017). In our setting, we restrict voters' subset preference structures by requiring consistency in the sense that a voter's dis/approval of a subset (committee) is derived from their dis/approval of individual alternatives in it (committee members).

The social choice approach to preference aggregation is useful in operations research beyond aforementioned settings of team selection or election of corporate directors, both of which are naturally modeled by and analyzed via voting or elections procedures. For example, product assortment decisions of a retailer can be cast as a problem of selecting a subset of alternatives (products) among available ones. There are many aspects that influence the retailer's product assortment decision, from possible physical constraints on the supply side to substitutability and complementarity effects on demand side, to corresponding pricing decisions (e.g., Borin et al., 1994; Cachon et al., 2005; Gaur and Honhon, 2006; Kök et al., 2008; Mantrala et al., 2009). In certain settings, however, social choice methods could provide guidance for a product assortment decision. Consider a subscription-price digital media streaming service firm (e.g., Netflix). Given the subscription pricing, it is important to maximize the number of subscribers who find many of available products appealing and thus value the service. As content licensing fees tend to be highest for the most popular products, such firms leverage heterogeneity of consumer preferences and are aiming at "long tail" product assortment, i.e., they offer a large number of low-fee low-demand products, thereby ensuring that every subscriber can find a sufficient number of (often "quirky") products that appeal to them (e.g., Brynjolfsson et al., 2003, 2006; Goel et al., 2010). This "long tail" strategy is considered as one of the main competitive advantages of online firms (e.g., Elberse, 2008; Brynjolfsson et al., 2011). In order to benefit from the "long-tail" strategy, firms invest in the recommender system that guides subscribers through a large product assortment to discover products they might value, but the firm first and foremost needs to carefully build its product assortment (e.g., Gomez-Uribe and Hunt, 2015; Tan et al., 2017). Since it is impossible to asses subscribers' values for each individual product they have not experienced yet (e.g., movie which they have not seen) and since licensing fees are often bundled by a content provider (e.g., movie studio), the firm can gather broad subscriber preferences over genres and select its assortment to reflect these. Subscriber preferences over genres can be assessed indirectly through ratings or by direct elicitation. (For example, in addition to collecting movie ratings regardless of the movie being watched on Netflix or not, Netflix asks new subscribers to select all movie genres they enjoy watching and stores this information in their profile.) Hence, information from every subscriber can be represented as a vote for genres the subscriber "approves" of. Similarly, if one further assumes that subscriber satisfaction (and consequently revenues) corresponds to the existence of a sufficient number of available products that the subscriber approves of, the firm could aggregate subscriber preferences over genres to guide them in building the most cost-effective product assortment. In doing so, the firm has to take into account which product bundles can be contracted with content providers, so there could be restrictions on which product assortments are even possible. For example, if the firm acquires a small number of expensive blockbuster movies, there might be less money available to build "long enough tail", thereby possibly losing many subscribers who would not be able to find sufficiently many products appealing, and consequently would churn.

The methodology we present in this paper can readily be applied to such large consumer preference data and yield a selection of genres (product types) that maximizes the number of customers who "approve" it. Hence, our subset selection procedure and, more generally, social choice approach provide a way to aggregate large consumer preference profiles and guide decision-makers in selecting optimal product assortment. (Given complexities associated with processing large and noisy data sets and with optimizing under numerous assortment constraints that often yield intractability, we believe there might be a value in the simple-to-implement procedure that proposes a reasonable, if not optimal, solution.)

In our presentation, for concreteness, we describe our subset-preference aggregation procedure in a classical voting setting: voters need to elect a committee consisting of candidates that are running for election. The admissible committees are known before the votes are cast, i.e., the size and/or structure of committees eligible to be elected could be predetermined. For example, a corporate board of directors has a predetermined number of positions known before the election; a parliamentary committee has a predetermined number of seats for the majority party and for opposition representatives; there are five starting players representing their conference in the NBA all-star game (until recently the starting line-up consisted of the top center, top two forwards, and top two guards, according to fan votes cast for individual players). Even when the set of admissible committees is restricted and the number of candidates under consideration is small, it can contain a prohibitively large and unmanageable number of eligible committees. (E.g., there are $\binom{30}{7} >$ 2M different seven-member committees that could be constructed from a pool of 30 candidates. Identifying several most preferred candidates from a pool of 30 is feasible, but even considering a small portion of more than two million possible seven-member committees might be too much to ask from a voter.) Thus, it is necessary for practical applicability to avoid any intractability issues due to the large number of eligible committees, such as placing any communicational burden on voters who might have limited bandwidth. Hence, it is a prerequisite that representation and communication of voters' preferences is succinct.

Our procedure aggregates voters' preferences which are expressed in terms of approvals for individual candidates, just as in approval voting. One interpretation of such approval in the context of committee selection might be that a voter approves of a particular candidate representing them in the committee. (A similar approach towards expressing voter preferences have been proposed in Brams et al., 2007, but the vote aggregation procedure proposed there is different than ours.) The key feature of our procedure is that the aggregation rule is defined to capture voters' preferences over admissible committees, instead of constructing a winning committee based on voters' preferences over individual candidates. Informally, the procedure assumes that the voter approves of every committee that contains a sufficiently large number of candidates they approve of. (As will be discussed, the definition of "sufficiently large" intersection could depend on the size of the committee; all this is predefined and publicly known prior to the selection of the set of candidates.) While a voter can approve more than one committee in this way, our procedure by construction does not allow for different intensities of approvals. The selected (winning) committee is the committee approved of by the most voters. The procedure is formally introduced in Section 2.

In Section 3, we investigate principal differences between the composition of the winning committee provided by our procedure and that of the committee constructed by combining candidates the maximize the sum of votes they got in a single-candidate election. The latter approaches utilize and emphasize preferences over individual candidates, potentially ignoring important information. In contrast, we establish that our procedure picks up on team-specific preferences and allows both for the winning committee size not to be maximal (i.e., enlarging the committee with any candidate could be decremental for its value), and for the top individual candidate not to be in the winning committee (thereby potentially picking up inherent synergetic value of the team which could dominate individual candidate value; akin to the saying *There's no "I" in "team"*).

We also introduce and discuss two sets of basic properties of generic subset choice functions. We show that our choice functions satisfy nullity, anonymity, as well as certain voter profile partition consistency and inclusivity properties. Further, in the restricted case of just a single voter casting a vote, we establish a form of monotonicity, completeness and consistency for our procedure. All these are formally introduced and discussed in Section 4. We then use these properties to characterize our class of subset choice functions.

In addition to designing an aggregation procedure that is (in our view) applicable for addressing problems of interest to operations research community, our work also contributes to the social choice literature addressing the problem of committee (s)election. Until recently, most of the work in this area focused on either treating subsets as individual alternatives and analyzing properties and performance of classical single-winner voting procedures on committee selection, or on constructing the winning committee based on the vote count of properly designed single candidate elections, with voters and/or candidates partitioned to elect a designated committee member (e.g., Fishburn, 1981a,b; Barberà et al., 1991, 2005; Ratliff, 2006; Ratliff and Saari, 2014). Recent works have also adopted modifications of approval representation of subset preferences introduced in Fishburn and Pekeč (2004), albeit with different approaches towards aggregation of preferences over subsets (Brams et al., 2007; Kilgour, 2010; Kilgour and Marshall, 2012; Duddy, 2014; Kilgour, 2016; Subiza and Peris, 2017). In recent years, motivated by many problems that require aggregating and ranking vast amounts of user preference data in variety of digital business settings, several works in the emerging theoretical computer science area of computational social choice also considered the problem of committee selection. Specifically, Meir et al. (2008) establish that strategizing and manipulating multi-winner elections is hard, while Alon et al. (2016) discuss complexity of several classes of aggregation procedures of preferences that have binary representation (our procedure belongs to one of classes they study). Elkind et al. (2017) and Aziz et al. (2017) relate computational issues central to computational social choice to properties of different committee voting procedures. These works combining the classical social choice theory with computational issues relevant to possible implementations in the data-rich environments are central for understanding and design of such aggregation methods for practical applications.

In the next section we introduce our subset-preference aggregation procedure. In Section 3, we discuss the size and structure of winning committees proposed by our procedure. We provide a characterization of our class of choice functions in Section 4 and conclude the paper with brief concluding remarks.

2. Threshold Choice Model

We assume throughout that there are n voters ($i = 1, 2, \ldots, n$) and m candidates ($j = 1, 2, \ldots, m$). The set of all subsets of $[m] = \{1, 2, \ldots, m\}$ is denoted by $2^{[m]}$, and the set of all potential committees with one or more candidates is $2^{[m]} \setminus \{\emptyset\}$.

We denote by ξ the set of *admissible committees* for a particular application and assume that ξ is a nonempty subset of $2^{[m]} \setminus \{\emptyset\}$. The set of all k-member committees is denoted by ξ_k. Thus, when a committee of k candidates is to be elected and all k-member committees are admissible, we set $\xi = \xi_k$. When all nonempty committees are admissible, $\xi = 2^{[m]} \setminus \{\emptyset\}$.

Let V_i be the vote of voter i. We allow V_i to be any subset of $[m]$, including the empty set. A *voting ballot profile*, or *profile* for short, is an n-tuple

$$V = (V_1, V_2, \ldots, V_n)$$

We often write V_i in simplified notation, denoting, e.g., $V_i = \{3,5,6\}$ by 356.

The set of all possible profiles is denoted by $\mathcal{V} = \left(2^{[m]}\right)^n$. Let V_iS denote the number of candidates in committee S that are also in the vote of voter i, i.e.,

$$V_iS = |V_i \cap S|.$$

Thus, the n-tuple $VS = (V_1S, V_2S, \ldots, V_nS)$ tells us how many candidates in S each voter selected.

Example 1 A committee of three is to be chosen from eight candidates (1,2,...,8). There are nine voters (1,2,...,9) with the following voting ballot profile:

Voter	1	2	3	4	5	6	7	8	9
AV Ballot	2	12	12	13	37	45	46	47	48

We have $n = 9$, $m = 8$, $\xi = \xi_3$, and the profile is $V = (2, 12, 12, 13, 37, 45, 46, 47, 48)$. For committee $S = \{1, 3, 4\}$, $VS = (0, 1, 1, 2, 1, 1, 1, 1, 1)$. $\square$

Given ξ, let C be a map from $\mathcal{V}$ to the set of nonempty subsets of ξ. We refer to C as a *choice function*, and to $S \in \xi \cap C(V)$ as a *choice for profile* V. Without loss of generality, we restrict our attention to ξ and C such that the following hold:

R1 (Choice discrimination) $\forall S \in \xi$, $\exists V \in \mathcal{V}$, $T \in \xi$, such that $S \in C(V)$ and $T \notin C(S)$;

R2 (Vote discrimination) $\forall S \in \xi$, $\exists V \in \mathcal{V}$ such that $S \notin C(V)$.

R1 guarantees that $S \in \xi$ is a choice for at least one voter profile V; if this were not the case, there would be no need for S to be an admissible committee, and we could have simply considered $\xi \setminus \{S\}$ instead of ξ, without loss of generality. Furthermore, R1 excludes a possibility that whenever S is a choice for some profile $V \in \mathcal{V}$, then all committees are choices for V, i.e., $C(V) = \xi$; if this were the case, we could have again considered $\xi \setminus \{S\}$ instead of ξ, without loss of generality.

R2 excludes a possibility for S to be a choice regardless of V, i.e., it disallows for $S \in C(V)$ for all $V \in \mathcal{V}$; if this were the case, we could have once again considered $\xi \setminus \{S\}$ instead of ξ, without loss of generality.

A *threshold function*, *TF* for short, is a map t from ξ into non-negative real numbers; let Υ denote the set of threshold functions. A threshold function t provides a natural way to define approval of committee S by voter i:

$$V_i S \geq t(S) \iff i \text{ approves } S. \tag{1}$$

So, voter i approves of committee S if and only if $V_i S$, the number of candidates in S that voter i approves of, is at lest as large as $t(S)$, the threshold for committee S. Further, given ξ and $t \in \Upsilon$, the *choice function* C_t *for threshold function* t is defined by

$$C_t(V) = \{S \in \xi : |\{i : V_i S \geq t(S)\}| \geq |\{i : V_i T \geq t(T)\}|,\ \forall T \in \xi\}. \tag{2}$$

Thus, $S \in \xi$ is a choice for profile V and threshold function t, if as many voters approve committee S (according to t) as any other admissible committee. We refer to C_t for $t \in \Upsilon$ as a *threshold choice function*, or simply *TCF* for short.

Remark 1 Note that a TCF does not account for the intensity of allignment between voter's vote and an admissible committee. It simply records whether $V_i S$ is above the threshold $t(S)$, and not the actual value of $V_i S$.

Consider an alternative method based on approval voting ballots, in which voter i assigns $V_i S$ votes to a committee S. This method is equivalent to choosing a committee of candidates that maximizes the sum of individual approval voting scores of candidates from a committee:

$$\sum_{i=1}^{n} V_i S = \sum_{i=1}^{n} \sum_{j \in S} V_i\{j\} = \sum_{j \in S} \sum_{i=1}^{n} V_i\{j\} = \sum_{j \in S} |\{i : j \in V_i\}|.$$

In other words, this alternative method of accounting for actual values of $V_i S$ (i.e., the number of members of committee S approved by voter i) coincides with the simplistic approach of using approval voting to rank individual candidates and then construct a committee consisting of the candidates receiving most votes. In the next section we discuss and contrast properties of TCFs to those of procedures that leverage votes tallied for individual candidates to construct the winning committee. $\square$

Example 2 *(Example 1 revisited)* Before illustrating threshold functions on this example, note that under ordinary approval voting for individual candidates (where voter i approves of candidate j if and only if $j \in V_i$, and the ranking of candidates is determined by the number of voters who approve of each of them), the three candidates with the most approval votes are candidates 1,2, and 4. Thus, if one were to use the approval voting results for individual candidates to construct a three-member committee, the elected committee would be {1,2,4}. As mentioned in Remark 1, this outcome could also be reached by assigning $V_1S + \ldots + V_9S$ votes to committee S. Committee $\{1,2,4\}$ gets 10 votes, while any other three member committee gets at most nine votes.

Next consider two threshold rules.

$\mathbf{t \equiv 1.}$ Voter i "approves" of a committee S if and only if at least one of their "approved" candidates is among the three members of S, i.e., if and only if $V_iS \geq 1$. The only three-member committee approved of by all voters is {2,3,4}. Thus $C_t(V) = \{\{2,3,4\}\}$.

$\mathbf{t \equiv 2.}$ Voter i "approves" of a three-member committee S if and only if the majority (i.e., at least two) of the committee members are approved by her, i.e., if and only if $V_iS \geq 2$. Under this rule, voters 2, 3 and 4 approve of committee {1,2,3}, and no other committee of three has more than two approving voters. Thus $C_t(V) = \{\{1,2,3\}\}$. □

Finally, we note that use of TFCs alleviates the burden from voters who do not need to report preferences over the set ξ of possibly as many as $2^m - 1$ admissible committees, and only report preferences over m candidates. This also holds true for any approach that would use tallies of *votes cast for individual candidates* to construct a winning committee (just like approval voting approach described in Remark 1). However, unlike those approaches, TFCs take into account voters' preferences over admissible committees by first interpreting voter-provided information (m bits of data) using (1), to infer voter preferences over committees and create approval votes for all committees in ξ (potentially as many as $2^m - 1$ bits of data), and only aggregating such *votes cast for committees* to determine a winning committee. Note that this can readily be achieved in time polynomial in $(nm + |\xi|)$ (the first term corresponds to the number of bits needed to represent the approval voting ballot profile, and the second is the number of admissible committees) by simply calculating $|\{i : V_iS \geq t(S)\}|$. Hence, we have

Proposition 1 *TCF can be computed in time polynomial in* $(nm + |\xi|)$.

Any voting procedure at the minimum would have to compute votes cast for each of potential election winners, so if the set of admissible committees ξ is large, any procedure counting *votes for committees* will naturally need at least $O(|\xi|)$ time. Hence, if ξ is exponentially large in terms of m but not counted towards the problem's input size (as it could be predifined or described by, e.g, $\log m$ bits needed to encode $k \leq m$ if admissibale committees need to be of the fixed size, i.e., ξ_k), the computational complexity could readily escalate. In fact, Alon et al. (2016) establish complexity of several classes of binary aggregation procedures and computing TCFs belongs to one of the classes they study. Here we just demonstrate that TCF computation encodes a classical NP-hard problem, provided predetrmined succinct representation of input data.

Proposition 2 *Let* $\xi = \xi_k$ *and let* $|V_i| = 2$ *for all voters* i. *Computing TCF for* $t \equiv 1$ *is NP-complete.*

Proof. There is a one-to-one correspondence between all such voter profiles and graphs with vertex set $[m]$ and edges $V_1, V_2, \ldots, V_n$. Thus, determining whether there exists $S \in \xi$ approved by all voters, i.e., such that $V_i S > 0$ for all i, is equivalent to determining whether S is a vertex cover for the corresponding graph. Determining whether a graph has a vertex cover of size k is one of the fundamental NP-complete problems (Garey and Johnson, 1979). Therefore, determining whether there exists $S \in \xi_k$ approved by all voters is also NP-complete. □

Note that complexity stems from reduction of the size of the input, and not from the fact that TCF computation requires comparing any admissible committee S with every voter profile V_i. Even selecting a committee that maximizes the sum of scores of individuals in the committee (as is the case with any method that first aggregates votes for individuals and then, based on these votes chooses an admissible committee) is computationally hard: given the list of individual votes $v(j)$, $j = 1, ..., m$, finding $S \in \xi$ that maximizes $\sum_{j \in S} v(j)$ is also a cannonical NP-complete problem (Garey and Johnson, 1979).

3. TCFs and Top Candidates

We now consider some properties of TCFs. Our focus is on highlighting differences between the size and structure of winning committees using TCFs and those composed from top candidates of an election of individual candidates.

Note that when a winning committee is constructed by maximizing the sum of vote counts for each individual candidate, (a) the winning committee S has to be maximal in ξ, i.e., if $S \subset T$ then $T \notin \xi$, and (b) if ξ is closed under single candidate exchanges (i.e., $S \in \xi$ implies $S \setminus \{j\} \cup \{l\} \in \xi$, for every $j \in S, l \notin S$; e.g., ξ_k), then the candidate with most votes in the individual election must be a member of the winning committee.

We show that neither of these properties needs to hold for TCFs. In what follows, we will demonstrate this even within the limited class of *cardinal* TF, i.e., $t \in \Upsilon$ such that for every $S \in \xi$, $t(S)$ depends only on $|S|$. We also call t a *constant* TF if $t(S) = t(T)$ for all $S, T \in \xi$. A constant TF is clearly a cardinal TF, and if $\xi \subseteq \xi_k$ for some k, then a cardinal TF is also a constant TF (cf., Example 2).

3.1. Winning committee need not be maximal

We first establish that (a) above need not hold for TCFs.

For larger admissible sets, for example $\xi = 2^{[m]} \setminus \{\emptyset\}$, a cardinal TF is *nondecreasing* if, for all $S, T \in \xi$, $|S| < |T|$ implies $t(S) \le t(T)$. Similarly, a cardinal TF is *nonincreasing* if, for all $S, T \in \xi$, $|S| < |T|$ implies $t(S) \ge t(T)$. (Apart from constant TF's, we will not pay much attention to nonincreasing TF's because it seems odd to have $t(S) > t(T)$ when $|S| < |T|$.) The following lemma provides a connection to choice functions.

Lemma 1 *If t is a nonincreasing cardinal TF then, for all $S,T \in \xi$ for which $S \subset T, S \in C_t(V) \Rightarrow T \in C_t(V)$.*

Proof. When t is nonincreasing and $S \subset T$, we have $V_iS \leq V_iT$ and $t(S) \geq t(T)$, so if a voter approves of S, i.e., if $V_iS \geq t(S)$, then that voter also approves of T, since $V_iT \geq V_iS \geq t(S) \geq t(T)$. It follows that if $S \in C_t(V)$ then $T \in C_t(V)$. □

Because Lemma 1 applies to constant TF's, such functions favor larger committees as choices when ξ includes committees of various sizes. For example, if $t(S) = 1$ for all $S \in \xi$, and $\xi = 2^{[m]} \setminus \{\emptyset\}$, then $[m]$, the committee of the whole, is a choice for every $V \in \mathcal{V}$. We therefore consider constant functions primarily in restricted settings such as $\xi = \xi_k$.

Note that cardinal TFs satisfy *neutrality*: for every permutation π of [m] and every $S = \{j_1, \ldots, j_k\}$, $t(S) = t(\{\pi(j_1), \ldots, \pi(j_k)\})$, i.e., they treat candidates equally.

Remark 2 Neutrality may make cardinal TFs inappropriate in some settings. For example, suppose a committee of five is to be chosen from 18 candidates who divide naturally into three distinct groups. If it is desirable but not mandatory to have at least one candidate from each group on the chosen committee, then t could be biased in favor of representativeness. With denoting ξ_5^r the set of committees in ξ_5 that represent precisely r of the three groups, one might take $t(S) = 2$ for $S \in \xi_5^3$, $t(S) = 3$ for $S \in \xi_5^2$, and t(S)=5 for $S \in \xi_5^1$. However, if only the committees in S_5^3 are admissible, we could let $\xi = \xi_5^3$ and take t constant there.

Other restrictions on t could be considered for larger admissible sets such as, for all $S,T \in \xi$, (i) $1 \leq t(S) \leq |S|$, (ii) $S \subset T \Rightarrow t(S) + 1 \leq t(T)$, and/or (iii) $S \cap T = \emptyset \Rightarrow t(S \cup T) \leq t(S) + t(T)$. For example, when $\xi = 2^{[m]} \setminus \{\emptyset\}$, (i) and (ii) imply that $t(S) = |S|$ for all $S \in \xi$. □

The following lemma for this case provides a counterpart to Lemma 1.

Lemma 2 *If $\xi = 2^{[m]} \setminus \{\emptyset\}$ and $t(S) = |S|$ for all $S \in \xi$ then, for all $S,T \in \xi$, such that $S \subset T$, $T \in C_t(V) \Rightarrow S \in C_t(V)$.*

Proof. Under the lemma's hypotheses, including $S \subset T$, every voter who approves of T also approves of S, and it follows that if $T \in C_t(V)$ then $S \in C_t(V)$. □

Lemma 2 shows that some TF's in the general case could favor smaller committees as choices from ξ. If necessary, this could be counteracted by adopting a secondary choice rule that only larger members of $C_t(V)$ are "acceptable". It could also be counteracted by adopting an entirely different threshold function that is biased towards committees of a certain size.

A related idea that might tend to elect moderate-sized committees concerns majority threshold functions. We define the *majority* TF by $t(S) = |S|/2$ for all $S \in \xi$, and the *strict majority* TF by

$t(S) = (|S|+1)/2$ for all $S \in \xi$. The definitions apply to all admissible sets, including ξ_k, where the majority and strict majority TF's are equivalent for C when k is odd. In Example 2 we have already seen the (strict) majority threshold function ($t \equiv 2$) when $\xi = \xi_3$. The following example illustrates the strict majority TF with $\xi = 2^{[m]} \setminus \{\emptyset\}$.

Example 3 Suppose $n = 12$, $m = 8$, and

$$V = (123, 15, 16, 278, 23, 24, 2578, 34, 347, 46, 567, 568).$$

Under the strict majority TF, the maximum approvals for a 1-member committee is 4, for a 2-member committee it is 2 (34,23,56,57,58,78), for a 3-member committee it is 5 (only for 234), for a 4-member committee it is 3 (5678), and for a 5-member committee it is 4 (15678). Hence {2,3,4} is the only member of $C_t(V)$. □

Hence, Lemma 2 and Examples 2 and 3 yield the following result.

Proposition 3 *The winning committee chosen by TCF need not be maximal in* ξ.

In the context of the example discussed in the introduction, product assortment decision for subscription-based media streaming service firm, Proposition 3 suggests that enlarging the assortment is not necessarily optimal. Given that each individual subscriber is interested in and can consume a limited number of product types, there might be no need to enhance heterogeneity of product types beyond some optimal subset of types. In contrast, simplistic approaches based on estimating popularity of each product type would suggest continual enlargement of the categories, provided such enlargements are not cost prohibitive. Such strategy does not provide guidance on optimal subset of product types and could potentially backfire since too many choices could confuse users and increase their search costs (e.g., Schwartz, 2004; Brynjolfsson et al., 2011; Tan et al., 2017).

3.2. Winning committee and top candidates

We next demonstrate that (b) above need not hold for TCFs. Example 2 already established this with approval voting used to determine vote counts for individual candidates. The following proposition formalizes this.

Proposition 4 *When* $\xi = 2^{[m]} \setminus \{\emptyset\}$ *and* t *is the majority or the strict majority TF, there exist* n, m, *and* $V \in \mathcal{V}$ *such that candidate 1 is in a majority of the* V_i *and in more* V_i*'s than any of other candidates, but candidate 1 is in no* $S \in C_t(V)$.

Proof. First consider the majority TF. Suppose $n=6$, $m=5$, and $V=(123,124,135,145,25,34)$. Note that $C_t(V)=\{2345\}$ since every voter approves of the 4-member committee 2345, and no other committee has unanimous approval. However, candidate 1 is in the most V_i's (four), while every other candidate is in three V_i's.

Next consider the strict majority TF. Suppose $n=10$, $m=6$, and

$$V=(1234,1236,1245,1256,1345,1346,1456,235,246,356).$$

Because every voter approves of three of the five candidates in {2,3,4,5,6}, this 5-member committee has unanimous approval. No other committee has unanimous approval (the closest being 12356 with 9 approvals), so $C_t(V)=\{23456\}$. But candidate 1 is in seven V_i's, while the others are in six V_i's. □

The examples from the preceding proof can be modified by adding a few more voters and many more candidates without changing the conclusion of Proposition 4. For example, if we add three voters to the latter example and take $V_{11}=V_{12}=V_{13}=\{7,8,\ldots 100\}$ then $C_t(V)$ is still {2,3,4,5,6} and candidate 1 is in more V_i's (a majority) than any other candidate.

The result of Proposition 4 does not just hold for majority TFs, but for a large class of TFs. function t:

Proposition 5 *Let t be a threshold function such that $t(S)\geq 2$ for every $S\in\xi$ for which $1\in S$. Suppose that there exist $T\in\xi$ such that $1\notin T$, $t(T)\leq|T|$ and such that $AT<t(A)$ for every $A\neq T$. Then, for every $n\geq 3$, there exists a $V\in\nu$ such that candidate 1 is in a majority of the V_i's and in more V_i's than any of other candidates, but candidate 1 is in no $S\in C_t(V)$.*

Proof. Take $V_n=T$ and $V_1=V_2=\ldots=V_{n-1}=\{1\}$. Then, $C_t(V)=\{T\}$ since T is the only committee approved by voter n, and none of the other voters approves of any admissible committee. On the other hand, 1 is approved by $n-1$ voters, while any other candidate is approved by at most one voter. □

The preceding proof assumes that there are voters who choose not to approve of any committee by approving candidate 1 only. This is not the case in the next proposition but the same result still holds.

Proposition 6 *Let $S',S",T\in\xi$, $\{1\}=S'\bigcap S''$, $(S'\cup S'')\cap T=\emptyset$. Let $t(S')\leq|S'|, t(S'')\leq|S''|, t(T)\leq|T|$. Suppose that for every A such that $1\in A$, (i) $AT<t(A)$ and (ii) either $S'A<t(A)$ or $S''A<t(A)$. Then, for every $n\geq 9$, there exists a $V\in\nu$ such that candidate 1 is in a majority of the V_i and in more V_i's than any of other candidates, but candidate 1 is in no $S\in C_t(V)$.*

Proof. First observe that for every $n \geq 9$ (and for $n = 7$), there exists positive integers p_1, p_2, p_3 such that $p_1 + p_2 + p_3 = n$, $p_1 + p_2 > p_3$ and $p_3 > max\{p_1, p_2\}$. Let

$$V_i = \begin{cases} S' & \text{if } 1 \leq i \leq p_1 \\ S\text{''} & \text{if } p_1 < i \leq p_1 + p_2 \\ T & \text{if } p_1 + p_2 < i \leq n \end{cases}$$

Our assumptions say that in approval voting for individual candidates, 1 is approved by $p_1 + p_2$ voters while no other candidate is approved by more than $p_3 < p_1 + p_2$ voters. Note that any set A, such that $1 \in A$ is approved by at most $\max\{p_1, p_2\}$ candidates and thus no such set is in $C_t(V)$ because T is approved by $p_3 > \max\{p_1, p_2\}$ voters. □

There can be even more candidates that are top individual choices and that do not belong to any winning committee. For example, when $t \equiv 1$, $k = 3$ and (n,m)=(3,4), profile $V = (123, 123, 4)$ has $C_1(V) = \{124, 134, 234\}$. In this case no $S \in C_1(V)$ contains all three of the most popular candidates. The following theorem notes a stronger result.

Theorem 1 *Suppose $1 \leq t \equiv \alpha < k$. Then there are n, m and a profile V with $C_\alpha(V) = \{S\}$ such that at least k candidates not in S each appears in more V_i's than any candidates in S.*

Proof. Suppose $t \equiv 1 < k$. When $k = 2$, profile

$$V = \{1, 2, 134, 134, 234, 234\}$$

has $C_1(V) = \{12\}$, and each of 3 and 4 is in more V_i's than either of 1 or 2. When $k \geq 3$, let $n = m = 2k$ with

$$V = (1, 2, \ldots, k, \{1, k+1, k+2, \ldots 2k\}, \ldots, \{k, k+1, k+2, \ldots, 2k\}).$$

Then $C_1(V) = \{1, 2, \ldots, k\}$, each $j \leq k$ is in two V_i's and each $j > k$ is in $k > 2$ V_i's.

Suppose $2 \leq \alpha < k$. Let L be a list of the $\binom{k}{\alpha}$ subsets of $\{1, 2, \ldots, k\}$ that have α members. Let $m = 2k$, and let L' be a list of the $\binom{k}{\alpha}$ subsets of $\{1, 2, \ldots, k\}$ with α members in union with $\{k + 1, k+2, \ldots, 2k\}$. For example, when $\alpha = 2$ and $k = 3$, $L = (12, 13, 23)$ and $L' = (12456, 13456, 23456)$. Let r be a positive integer greater than $\alpha/(k - \alpha)$, let $n = (r+1)\binom{k}{\alpha}$, and let

$$V = (L, rL'),$$

where rL' denotes r repetitions of L'. Then $\{1, 2, \ldots, k\}$ has unanimous approval under $t \equiv \alpha$ and is the only member of ξ_k with this property, for any other member of ξ_k lacks the approval of at least one of the first $\binom{k}{\alpha}$ voters. Hence $C_\alpha(V) = \{1, 2, \ldots, k\}$. The number of voters with candidate $j > k$ in their approval sets is $\binom{k}{\alpha}$, and the number with candidate $j \leq k$ in their approval sets is

$$(r+1)\left[\binom{k}{\alpha} - \binom{k-1}{\alpha}\right] = (r+1)\binom{k}{\alpha}\frac{\alpha}{k}.$$

Because our choice of r ensures that $r > (r+1)\frac{\alpha}{k}$, each candidate in $\{k+1, \ldots, 2k\}$ is in more approval sets than each candidate in $\{1, 2, \ldots, k\}$. $\square$

Remark 3 An interesting combinatorial adjunct of Theorem 1 is to determine the minimum n for each $k \geq 2$ which admits a V that satisfies the conclusion of the theorem. The minimum n for $k = 2$ is easily seen to be $n = 4$ with $V = (1, 2, 134, 234)$. Then $C_1(V) = \{\{1, 2\}\}$ and $C_2(V) = \{\{3, 4\}\}$. The minimum n for $k = 3$ is $n = 6$ with $V = (1, 3, 256, 346, 2789, 145789)$, in which case $C_1(V) = \{123\}$, $C_2(V) = \{456\}$ and $C_3(V) = \{789\}$. We are not certain of the minimum n for $k = 4$ but know that it is no greater than 10 because the 10-voter profile

$$V = (1, 2, 3, 4, \{1, 5, 8, 9, 11, 12\}, \{2, 6, 7, 10, 11, 12\}, \{3, 5, 7, 9, 10, 11\},$$

$$\{4, 6, 8, 9, 10, 12\}, \{1, 5, 6, 13, 14, 15, 16\}, \{2, 7, 8, 13, 14, 15, 16\})$$

has $C_1(V) = \{1234\}$, $C_2(V) = \{5678\}$, $C_3(V) = \{\{9, 10, 11, 12\}\}$ and $C_4(V) = \{\{13, 14, 15, 16\}\}$. $\square$

It should be noted that the behavior exhibited by the results presented in this subsection shows only what is possible and not necessarily what is probable. We would ordinarily expect that the most popular candidates will be contained in some of the committees chosen by the majority and and strict majority TFs. However, one can envision situations in which top individual performers are not considered suitable to be on a committee.

The following proposition shows that, when all committees are acceptable, there is only one cardinal TF which ensures that every candidate in any selected committee (i.e., every member of C_t) is approved by at least one voter.

Proposition 7 *Suppose t is a cardinal TF with $1 \leq t(S) \leq |S|$ for every $S \in \xi = 2^{[m]} \setminus \{\emptyset\}$, where $m \geq 2$. Then $t(S) = |S|$ for all $S \in \xi$ if and only if, for every $V \in \nu$ that has $|V_i| > 0$ for at least one i, every $S \in C_t(V)$ contains only candidates with one or more votes.*

Proof. If $t(S) = |S|$ for all $S \in \xi$ and V is not $(\emptyset, \ldots \emptyset)$, then every $S \in C_t(V)$ is a subset of some V_i and therefore has positive support for every candidate therein.

Suppose t is not equivalent to the preceding TF. Then there is a k between 2 and m and an integer α such that $1 \leq t(S) = \alpha \leq k - 1$ for all $S \in \xi_k$. With $K = \{1, 2, \ldots, \alpha\}$, the constant profile $V = (K, K, \ldots, K)$ has unanimous approval for every $S \in \xi_k$ that includes K and $k - \alpha$ other candidates from $[m] \setminus K$ with no votes. $\square$

Thus, the only way to ensure that the committee selected using a cardinal TF contains individual candidates approved by at least one voter is to assume that a voter only approve of a committee when they approve of all of the committee members (i.e., $t(S) = |S|$ for all S). This is a quite restrictive condition and interpretation of voter preferences, once more suggesting the importance of aggregating preferences over committees, rather than over individual candidates.

Referring once again to the product assortment decision of the streaming media retailer from the Introduction, the results of this subsection point to potential shortcomings of focusing on most popular product types (movie genres) when constructing the assortment. In other words, such myopic approach to building product assortment could result in the limited number of product types (as most popular types are likely to have highest licensing fees) which could cripple the long tail and consequently limit the firm's opportunity to exploit heterogeneity of subscriber preferences. Furthermore, a collection of most popular product types need not be optimal, as results from this subsection suggest. Hence, it is important to aggregate subscriber preferences over *collections* of product types, rather than focusing on popularity of *individual* product types.

We close this section by demonstrating that the choice of an appropriate threshold has important and striking consequences on the committee that will be chosen. More precisely, Given ξ_k with $k \geq 2$, a voter's approval set may depend on the constant TF used to determine choice, i.e., on the value of α. However, if V is fixed and we vary α, $C_\alpha(V)$ might vary widely for the different α's. An extreme possibility is noted in the following theorem.

Theorem 2 *Given $k \geq 2$, there are n, m and a corresponding V such that $C_\alpha(V) = \{S_\alpha\}$ for $\alpha = 1, 2, \ldots k$ with the S_α mutually disjoint.*

Proof. Given $k \geq 2$, let S_α be a k-element set for each α in $\{1, 2, \ldots, k\}$ with the S_α mutually disjoint. Let $\bigcup S_\alpha$ be the candidate set, so $m = k^2$. With $\lambda_1, \lambda_2, \ldots, \lambda_k$ as-yet-unspecified positive integers, set

$$n = \sum_{\alpha=1}^{k} \lambda_\alpha \binom{k}{\alpha}$$

and for each α let $V \in \mathcal{V}$ have λ_α copies of each α-member subset of S_α. We will choose the λ_α so that $C_\alpha(V) = \{S_\alpha\}$.

Set $\lambda_k = 1$. Then $C_k(V) = \{S_k\}$ because S_k is the only V_i that contains at least k candidates.

Set $\lambda_{k-1} = 2$. Note that the only V_i's with $|V_i| \geq k-1$ are S_k and the two copies of each $(k-1)$-member subset of S_{k-1}. This ensures that $C_{k-1} = \{S_{k-1}\}$. (For example, when $k = 2$, $V = (1, 1, 2, 2, 34)$ with $S_1 = \{1, 2\}$ and $S_2 = \{3, 4\}$.)

When $k \geq 3$, let $\lambda_{k-2}, \ldots, \lambda_1$ be increasingly larger integers so that $C_\alpha(V) = \{S_\alpha\}$ for $\alpha = k-2, \ldots, 1$. For any such α, the number of V_i which are α-member subsets of S_α equals $\lambda_\alpha \binom{k}{\alpha}$, and a suitably large λ_α will make this greater than the number of V_i that contain α or more of any given k-member subset of $S_{\alpha+1} \cup \ldots \cup S_k$. $\square$

4. Characterization of TCFs

In this section we provide an axiomatization of TCFs. We introduce and discuss two sets of basic properties that TCFs have: full profile symmetry and decomposition properties, and single voter profile properties. We then use these properties to characterize the class of TCFs.

4.1. Full profile properties

Let $t \in \Upsilon$, and denote the associated TCF, C_t.

We say that voter profile U is a *permutation* of a voter profile V if there exists a permutation σ on the set of voters $[n]$ such that $U_i = V_{\sigma(i)}, i = 1, \dots, n$.

We also introduce notation for the two-part profile partitions:

$$V_p^- = (V_1, \dots, V_p, \emptyset, \dots, \emptyset), \qquad V_p^+ = (\emptyset, \dots, \emptyset, V_{p+1}, \dots, V_n);$$
$$V_i^* = (V_1, \dots, V_{i-1}, \emptyset, V_{i+1}, \dots, V_n), \qquad V_i^o = (\emptyset, \dots, \emptyset, V_i, \emptyset, \dots, \emptyset);$$

Consider the following properties of a (generic) choice function C

P1 (Nullity). $C(\emptyset, \dots, \emptyset) = \xi$

P2 (Anonymity). If U is a permutation of V, then $C(U) = C(V)$.

P3 (Partition Consistency). If $C(V_p^-) \cap C(V_p^+) \neq \emptyset$ then $C(V) = C(V_p^-) \cap C(V_p^+)$.

P4 (Partition Inclusivity). If $C(V_i^*) \cap C(V_i^o) = \emptyset$ for $i = 1, \dots, n$, then $C(V) = \bigcup_{i=1}^n C(V_i^*)$.

Property P3 says that if a committee would be chosen in both parts of a two-part partition of the electorate, then all such committees - and only those - will be in the choice set for the whole electorate. Property P4 says that if P3 does not apply for each of the noted n two-part partitions of 1 and $n-1$ voters (i.e. $C(V_i^*) \cap C(V_i^o)$ is empty), then a committee is in the choice set of the whole electorate if and only if it is in the choice set of at least one $(n-1)$-member subelectorate.

Proposition 8 *Any TCF, $C = C_t$, $t \in \Upsilon$, satisfies* P1, P2, P3, *and* P4.

Proof. It is straightforward to check that C_t satisfies P1 and P2.

To show that C_t satisfies P3 and P4, let $\alpha_t(S,V) = |\{i : V_i S \geq t(S)\}|$, denote the number of voters who meet or exceed the threshold $t(S)$ for S. Hence

$$S \in C_t(V) \text{ iff } \alpha_t(S,V) \geq \alpha_t(T,V) \quad \forall T \in \xi.$$

Further, denote

$$\alpha_t^*(V) = \max_{s \in \xi} \alpha_t(S,V).$$

Suppose $S \in C(V_p^-) \cap C(V_p^+)$. Then $\alpha_t(S, V_p^-) = \alpha_t^*(V_p^-)$ and $\alpha_t(S, V_p^+) = \alpha_t^*(V_p^+)$, so $\alpha_t(S, V) = \alpha_t^*(V)$ and $S \in C_t(V)$. Given $T \notin C_t(V_p^-) \cap C_t(V_p^+)$, we have $\alpha_t(T, V_p^-) \leq \alpha_t^*(V_p^-)$ and $\alpha_t(T, V_p^+) \leq \alpha_t^*(V_p^+)$, with $<$ in one or both cases, so $\alpha_t(T, V) < \alpha_t^*(V)$ and $T \notin C_t(V)$. Hence, P3 holds.

Suppose $S \in C_t(V_i^*)$, so $\alpha_t^*(V_i^*) = \alpha_t(S, V_i^*) \geq \alpha_t(T, V_i^*)$ for all $T \in \xi$. Given P4's hypotheses, we have $\alpha_t(S, V_i^o) = 0$, $\alpha_t(R, V_i^o) = 1$ for some $R \in \xi$ and for all such R, $\alpha_t(R, V_i^*) < \alpha_t(S, V_i^*)$. Then $\alpha_t(S, V) \geq \alpha_t(T, V)$ for all $T \in \xi$, so $S \in C_t(V)$. It follows that $C_t(V_i^*) \subseteq C_t(V)$, hence that $\bigcup_i C_t(V_i^*) \subseteq C_t(V)$.

Finally, suppose $S \notin C_t(V_i^*)$ for $i = 1, \ldots, n$, i.e., $S \notin \bigcup_i C_t(V_i^*)$, but contrary to P4, $S \in C_t(V)$. Then $\forall T_i \in C_t(V_i^*)$, we have

$$\alpha_t(S, V_i^*) = \alpha_t(T_i, V_i^*) - 1 = \alpha_t^*(V_i^*) - 1 \leq n - 2;$$
$$\alpha_t(S, V_i^o) = 1, \alpha_t(T_i, V_i^o) = 0.$$

But then $V_i S \geq t(S)$ for all i, so $\alpha_t(S, V) = n$, a contradiction to $\alpha_t(S, V_i^*) \leq n-2$ and $\alpha_t(S, V_i^o) = 1$. Therefore, P4 holds. $\square$

4.2. Single voter profile properties

We next describe properties of a (generic) choice function properties in the case of a single voter, i.e., $n = 1$. Thus, aggregation across multiple profiles is set aside and the focus is on the mapping from a single voter profile to a committee choice.

Throughout this subsection we will work with

$$C^*(A) = C(A) \text{ for all } A \in \nu_1,$$

The C^* notation indicates that the choice function C maps single voter profile, i.e., $n = 1$.

The following partition of ν_1 into two parts, η and ζ, will be used throughout this subsection:

$$\eta = \{A \in \nu_1 : C^*(A) = \xi \}$$
$$\zeta = \{A \in \nu_1 : S \notin C^*(A) \text{for some } S \in \xi\}.$$

A 0/1 matrix for C^* with its rows labeled by all possible single voter profiles (ν_1) and columns labeled by all admissible committees (ξ), and such that 0 denotes $S \notin C^*(A)$ and 1 denotes $S \in C^*(A)$ has the following form:

<table>
<tr><td></td><td></td><td colspan="3">ξ</td></tr>
<tr><td></td><td>η</td><td colspan="3">1</td></tr>
<tr><td rowspan="3">ν_1</td><td rowspan="3">ζ</td><td></td><td>0</td><td></td></tr>
<tr><td></td><td></td><td>1 0</td></tr>
<tr><td></td><td>1</td><td></td></tr>
</table>

(Note that η might be non-empty: for example, P1 imposes $\emptyset \in \eta$. Further note that every row in ζ has a 0 (by definition of ζ) and a 1 (when $C^*(A) \neq \emptyset$). Also, every column in ζ has a 0 and a 1 by S1 and S2: all 0's column would violate R1 and all 1's column would violate R2).

We are now ready to introduce three properties for (generic) choice functions on single voter-profiles. The first is a monotonicity axiom that is easily seen to be necessary for a TCF on a single voter profile $C^* = C_t$ (on $\mathcal{V}_1$) for some t.

Q1 (Monotonicity). For all $A, B \in \mathcal{V}_1$ and all $S \in \xi$: if $A \in \zeta$, $S \in C^*(A)$ and $BS \geq AS$ then $S \in C^*(B)$.

The monotonicity property has the following interpretation: the choice function interprets a single voter profile as a collection of candidates that the voter supports, and the support for any particular candidate is independent of the voter (not) supporting other candidates. Q1 has the following implications:

Lemma 3 *Let Q1 hold. For all $A \in \zeta$ and all $S \in C^*(A)$, $A \cap S \neq \emptyset$, i.e., $AS > 0$.*

Lemma 4 *Let Q1 hold. $C^*([m]) = \xi$.*

Proof. If $AS = 0$ with $A \in \zeta$ and $S \in C^*(A)$, then Q1 implies an all 1's column for S in the preceding matrix, and this violates R2. For Lemma 4, every column S in the matrix has a 1 in the ζ part, so, because $[m]S \geq AS$, Q1 gives $S \in C^*([m])$ for every $S \in \xi$. $\square$

In order to describe the final two properties for C^* we define an integral function τ on ξ:

$$\tau(S) = \min\{BS : B \in \zeta, S \in C^*(B)\}$$

Our structural restrictions assume that τ is unambiguously defined for every $S \in \xi$, Lemma 3 implies that $\tau(S) \geq 1$ for every S and Q1 says that $S \in C^*(A)$ whenever $AS \geq \tau(S)$. (It may be anticipated that this construction, which is well-defined for any choice function, has a property that for a threshold choice function $C^* = C_t$, $\tau = t$.)

The first of our two τ-based conditions is

Q2 (Completeness). $\forall S \in \xi$, $\exists A \in \zeta$ such that $AS = \tau(S) - 1$.

Proposition 9 *TCF on a single voter profile, $C^* = C_t$, satisfies Q2.*

Proof. We presume that $C^* = C_t$ for a positive integral t on ξ, and refer to this as "the model".

Given $S \in \xi$, let $B \in \zeta$ satisfy $BS = \tau(S)$ with $S \in C^*(B)$. Also, let

$$\mathfrak{F} = \{F \in \mathcal{V}_1 : FS = \tau(S) - 1\}.$$

Clearly, $\mathfrak{F} \neq \emptyset$. We suppose to the contrary of Q2 that $\mathfrak{F} \subseteq \eta$ and will derive a contradiction. By the model, $t(S) \leq \tau(S)$.

Suppose $t(S) = \tau(S)$. By Q1 and ξ's structure, some $G \in \xi$ has $GS \leq \tau(S) - 2$. Let $F \in \mathfrak{F}$ be such that $G \subset F$. Because $FS < t(S)$ and $F \in \eta$ by supposition, the model implies $FT < t(T)$ for all $T \in \xi$. Then $GT \leq FT < t(T)$ for all $T \in \xi$, so we obtain the contradiction that, according to the model, $G \in \eta$.

Therefore $t(S) < \tau(S)$. Take $F \subset B$ with $F \in \mathfrak{F}$. Then $FS \geq t(S)$ so, because $F \in \eta$, the model requires $FT \geq t(T)$ for all $T \in \xi$. But then $BT \geq FT \geq t(T)$ for all $T \in \xi$, so we obtain the contradiction that $B \in \eta$.

We conclude, for each $S \in \xi$, that some member of $\mathfrak{F}$ for S is in fact in ζ. □

Property Q2 forbids a cardinality gap in ζ under S (see $\mathfrak{F} \subseteq \eta$ in the preceding proof) between the AS with $S \in C^*(A)$ and those with $S \notin C^*(A)$.

The final property describes properties of single voter profiles $A \in \eta$, i.e., all A such that $C(A) = \xi$. For a threshold choice function, $A \in \eta$ if and only if $f(S) := AS - \tau(S)$ is either nonnegative or strictly negative function on ξ. (Either AS is above the threshold for every S or AS is below the threshold for every S) This is described in

Q3 (Consistency). $\forall A \in \eta$, $\forall B_1, B_2 \in \zeta$ and $\forall S, T \in \xi$: if $B_1 S = \tau(S)$, $B_2 T = \tau(T)$ and $AS \geq B_1 S$, then $AT \geq B_2 T$.

Lemma 5 *Suppose Q1-Q3 hold. For all $A \in \nu_1$, if there are $S, T \in \xi$ such that $AS \geq \tau(S)$ and $AT < \tau(T)$, then $A \in \zeta$.*

Proof. Suppose the lemma's hypothesis hold but $A \in \eta$. There exists $B_1, B_2 \in \zeta$ with $B_1 S = \tau(S)$ and $B_2 T = \tau(T)$. We then have $AS \geq B_1 S$ so, by Q3, $AT \geq B_2 T$, a contradiction to $AT < \tau(T)$. □

We conclude this subsection with characterization of threshold choice functions for single-voter profiles.

Theorem 3 *Suppose C^* satisfies Q1-Q3. Then C^* is a threshold choice function on single voter profiles. Moreover, τ is the unique integral $t \in \Upsilon$ for which $C^* = C_t$.*

Proof. For all $A \in \zeta$ and $S \in \xi$, the definition of τ and Q1 imply $S \in C^*(A)$ iff $AS \geq \tau(S)$. By Q2, τ is the unique integral t for which this is true, and we have $C^* = C_\tau$ on ζ.

When $A \in \eta$, Lemma 5 implies that either $AS \geq \tau(S)$ for all $S \in \xi$, or $AS < \tau(S)$ for all $S \in \xi$. It follows that $C^* = C_\tau$ on η. □

4.3. Representation Theorem

We are now ready to provide a characterization of TCFs on general voter profiles. The first step towards the characterization is to use Properties P2 and P3, introduced in Subsection 4.1, to extend the choice function from a single-voter profile to the set of consensus voter profiles, i.e., $V \in \{\emptyset, A\}^n$

Lemma 6 *Suppose $n \geq 2$ and let C be a choice function satisfying P2 and P3. Then for every A with $\emptyset \subset A \subseteq [m]$ and every profile $V \in \{\emptyset, A\}^n \setminus \{\emptyset\}^n$, we have*

$$C(V) = C(A, A, \ldots, A)$$

In particular, $C(A, \emptyset, \ldots, \emptyset) = C(A, A, \ldots, A)$.

Proof.
By P2, $C(A, \emptyset, \ldots, \emptyset) = C(\emptyset, A, \emptyset, \ldots, \emptyset) = C(\emptyset, \emptyset, \ldots, \emptyset, A)$.
By P3, $C(A, A, \emptyset, \ldots, \emptyset) = C(A, \emptyset, \ldots, \emptyset) \cap C(\emptyset, A, \emptyset, \ldots, \emptyset) = C(A, \emptyset, \ldots, \emptyset)$.
Further uses of P3 give $C(A, A, A, \emptyset, \ldots, \emptyset) = C(A, A, \emptyset, \ldots, \emptyset) \cap C(\emptyset, \emptyset, A, \ldots, \emptyset) = C(A, \emptyset, \ldots, \emptyset)$, and consequently, $C(A, A, \ldots, A) = C(A, \emptyset, \ldots, \emptyset)$.
P2 completes the proof. □

Note that the lemma's final equation defines C^* on a single voter profile through C operating on consensus profiles: $C^*(A) = C(A, A, \ldots, A) = C(A, \emptyset, \ldots, \emptyset)$. Hence, properties Q1-Q3 that were introduced in the context of a single voter profile are well defined for consensus profiles.

Theorem 4 *The choice function C is a TCF if and only if C satisfies properties P1-P4 and Q1-Q3 (for consensus profiles).*

Proof. We have shown necessity of all properties for threshold function as they were introduced in this section.

For sufficiency, note that Theorem 3 covers $n = 1$, and assume henceforth that $n \geq 2$. The general step in this proof will show how to go from $C = C_\tau$ on profiles with at most $k \geq 1$ non-empty V_i, $k < n$, to profiles with at most $k+1$ non-empty V_i. This can be done under Lemma 6 by fixing the last $(n-(k+1))$ V_i at $\emptyset$ (i.e., $V_i = \emptyset$ for $i = k+2, \ldots, n$) and then using P2 to get $C = C_\tau$ for all V with at most $k+1$ non-empty V_i. No real generality is lost if we take $k = n-1$, so we do this in what follows.

Assume that $C = C_\tau$ on profiles with at least one $V_i = \emptyset$. We extend this to $C = C_\tau$ on $\mathcal{V}$.

Using P3, Suppose (I, J) is a non-trivial two-part partition of $[n]$, also suppose V_I, V_J and V are profiles such that V_I equals V for $i \in I$ and is $\emptyset$ otherwise, and V_J equals V for $i \in J$ and is $\emptyset$ otherwise, and finally suppose that $C(V_I) \cap C(V_J) \neq \emptyset$. By P3, $C(V) = C(V_I) \cap C(V_J)$. By the result for $n-1$, $C(V_I)=C_\tau(V_I)$ and $C(V_J)=C_\tau(V_J)$, and by replicating the proof of P3 in Proposition 8, $C_\tau(V)=C_\tau(V_I)\cap C_\tau(V_J)$. Therefore, $C(V)=C_\tau(V)$.

Using P4 and its notation, suppose $C(V_i^*)\cap C(V_i^o)=\emptyset$ for $i=1,\dots,n$. By P4, $C(V)=\bigcup_i C(V_i^*)$. By the result for $n-1$, $C(V_i^*)=C_\tau(V_i^*)$ for $i=1,\dots,n$, and by replicating the proof of P4 in Proposition 8, $C_\tau(V)=\bigcup_i C_\tau(V_i^*)$. Therefore, $C(V)=C_\tau(V)$.

We conclude that $C(V)=C_\tau(V)$ for all $V\in\mathcal{V}$. □

5. Concluding Remarks

The basic premise of the work presented here is that aggregating preferences over subsets could and should take into account actual preferences over subsets, rather than choosing a subset by combining top singleton choices stemming out of some aggregation preferences over single alternatives. We present our work in the context of voting for committees. We develop a subset choice function which modifies the way approval voting ballots are counted, so that a voter approves of a committee if and only if the committee includes a sufficient number of candidates they voted for. The ballots are counted using threshold functions which have natural interpretations. Constant threshold functions simply allow voters to approve of any committee that has at least some fixed number of members that are approved by the voter. Similarly, cardinal threshold functions (such as majority threshold functions) allow voters to approve of any committee that has, e.g., some percentage (e.g., majority) of members that are approved by the voter. We show that our approach faithfully aggregates preferences over committees in a way that captures potential synergetic values of committees, and that does not myopically add individual candidates (even if they are most preferred) to the committee.

The simplicity of the proposed procedure is important since overly-complicated procedures that demand too much information from voters have slim chances of being implemented. By having to select only the set of candidates they approve of, voters avoid the potentially daunting burden of having to report their preference ranking over all possible committees. Furthermore, requiring voters to select a set of candidates instead of requiring that they rank individual candidates, aligns with the purpose of the aggregation procedure (selecting a committee, i.e., a set of candidates). Thus, approval voting, in contrast to other widespread voting methods for selecting individual candidates, seems to be an obvious starting point for developing an aggregation method for committee selection.

We believe that this work and, more generally, social choice perspective on subset-preference aggregation nicely fits many operations research problems, especially in the context of consumer ratings/rankings driven managerial decisions in the presence of vast amounts of often noisy user data. Computationally light procedures based on the (re)fresh(ed) social choice perspective, have a potential of providing reasonable, if not optimal, solutions to many challenging problems of emerging digital business models.

References

Alon, Noga, Robert Bredereck, Jiehua Chen, Stefan Kratsch, Rolf Niedermeier, Gerhard J Woeginger. 2016. How to put through your agenda in collective binary decisions. *ACM Transactions on Economics and Computation* **4**(1) 1–28.

Arrow, Kenneth J. 1951. *Social choice and individual values*. Wiley.

Arrow, Kenneth J. 2012. *Social choice and individual values*, vol. 12. Yale University Press.

Aziz, Haris, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, Toby Walsh. 2017. Justified representation in approval-based committee voting. *Social Choice and Welfare* **48**(2) 461–485.

Barberà, Salvador, Jordi Massó, Alejandro Neme. 2005. Voting by committees under constraints. *Journal of Economic Theory* **122**(2) 185–205.

Barberà, Salvador, Hugo Sonnenschein, Lin Zhou. 1991. Voting by committees. *Econometrica: Journal of the Econometric Society* 595–609.

Borin, Norm, Paul W Farris, James R Freeland. 1994. A model for determining retail product category assortment and shelf space allocation. *Decision sciences* **25**(3) 359–384.

Brams, Steven, Peter C Fishburn. 2007. *Approval voting*. Springer.

Brams, Steven J, D Marc Kilgour, M Remzi Sanver. 2007. A minimax procedure for electing committees. *Public Choice* **132**(3-4) 401–420.

Brynjolfsson, Erik, Yu Hu, Duncan Simester. 2011. Goodbye Pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science* **57**(8) 1373–1386.

Brynjolfsson, Erik, Yu Hu, Michael D Smith. 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science* **49**(11) 1580–1596.

Brynjolfsson, Erik, Yu Jeffrey Hu, Michael D Smith. 2006. From niches to riches: Anatomy of the long tail. *Sloan Management Review* **47**(4) 67–71.

Cachon, Gérard P, Christian Terwiesch, Yi Xu. 2005. Retail assortment planning in the presence of consumer search. *Manufacturing & Service Operations Management* **7**(4) 330–346.

Cai, Jie, Jacqueline L Garner, Ralph A Walkling. 2009. Electing directors. *The Journal of Finance* **64**(5) 2389–2421.

Duddy, Conal. 2014. Electing a representative committee by approval ballot: An impossibility result. *Economics Letters* **124**(1) 14–16.

Easterbrook, Frank H, Daniel R Fischel. 1983. Voting in corporate law. *The Journal of Law and Economics* **26**(2) 395–427.

Elberse, Anita. 2008. Should you invest in the long tail? *Harvard Business Review* **86**(7/8) 88–96.

Elkind, Edith, Piotr Faliszewski, Piotr Skowron, Arkadii Slinko. 2017. Properties of multiwinner voting rules. *Social Choice and Welfare* **48**(3) 599–632.

Feng, Bo, Zhong-Zhong Jiang, Zhi-Ping Fan, Na Fu. 2010. A method for member selection of cross-functional teams using the individual and collaborative performances. *European Journal of Operational Research* **203**(3) 652–661.

Fishburn, Peter C. 1981a. An analysis of simple voting systems for electing committees. *SIAM Journal on Applied Mathematics* **41**(3) 499–502.

Fishburn, Peter C. 1981b. Majority committees. *Journal of Economic Theory* **25**(2) 255–268.

Fishburn, Peter C, Aleksandar Saša Pekeč. 2004. Approval voting for committees: Threshold approaches. *Working Paper, Duke University* .

Fitzpatrick, Erin L., Ronald G. Askin. 2005. Forming effective worker teams with multi-functional skill requirements. *Computers & Industrial Engineering* **48**(3) 593 – 608.

Francis, Wayne L. 1982. Legislative committee systems, optimal committee size, and the costs of decision making. *The Journal of Politics* **44**(3) 822–837.

Garey, Michael R, David S Johnson. 1979. *Computers and intractability: A guide to the theory of NP-completness*, vol. 29. Freeman.

Gaur, Vishal, Dorothée Honhon. 2006. Assortment planning and inventory decisions under a locational choice model. *Management Science* **52**(10) 1528–1543.

Gibbard, Allan. 1973. Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society* 587–601.

Goel, Sharad, Andrei Broder, Evgeniy Gabrilovich, Bo Pang. 2010. Anatomy of the long tail: ordinary people with extraordinary tastes. *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 201–210.

Gomez-Uribe, Carlos A., Neil Hunt. 2015. The Netflix recommender system: Algorithms, business value, and innovation. *ACM Trans. Manage. Inf. Syst.* **6** 13:1–13:19.

Kilgour, D Marc. 2010. Approval balloting for multi-winner elections. *Handbook on approval voting*. Springer, 105–124.

Kilgour, D Marc. 2016. Approval elections with a variable number of winners. *Theory and Decision* **81**(2) 199–211.

Kilgour, D Marc, Erica Marshall. 2012. Approval balloting for fixed-size committees. *Electoral Systems*. Springer, 305–326.

Kök, A Gürhan, Marshall L Fisher, Ramnath Vaidyanathan. 2008. Assortment planning: Review of literature and industry practice. *Retail supply chain management*. Springer, 99–153.

Mantrala, Murali K, Michael Levy, Barbara E Kahn, Edward J Fox, Peter Gaidarev, Bill Dankworth, Denish Shah. 2009. Why is assortment planning so difficult for retailers? a framework and research agenda. *Journal of Retailing* **85**(1) 71–83.

McElroy, Gail. 2006. Committee representation in the european parliament. *European Union Politics* **7**(1) 5–29.

Meir, Reshef, Ariel D Procaccia, Jeffrey S Rosenschein, Aviv Zohar. 2008. Complexity of strategic behavior in multi-winner elections. *Journal of Artificial Intelligence Research* **33** 149–178.

Ratliff, Thomas C. 2006. Selecting committees. *Public Choice* **126**(3-4) 343–355.

Ratliff, Thomas C, Donald G Saari. 2014. Complexities of electing diverse committees. *Social Choice and Welfare* **43**(1) 55–71.

Satterthwaite, Mark Allen. 1975. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory* **10**(2) 187–217.

Schwartz, Barry. 2004. *The paradox of choice: Why more is less*. HarperCollins New York.

Sen, Amartya K. 1984. *Collective Choice and Social Welfare*. Elsevier.

Sen, Amartya K. 2017. *Collective Choice and Social Welfare: Expanded Edition*. Penguin UK.

Sjostrim, William K Jr, Young Sang Kim. 2007. Majority voting for the election of directors. *Conn. L. Rev.* **40** 459.

Subiza, Begoña, Josep E Peris. 2017. A representative committee by approval balloting. *Group Decision and Negotiation* **26**(5) 1029–1040.

Tan, Tom Fangyun, Serguei Netessine, Lorin Hitt. 2017. Is Tom Cruise threatened? An empirical study of the impact of product variety on demand concentration. *Information Systems Research* **28**(3) 643–660.

Yermack, David. 2010. Shareholder voting and corporate governance. *Annu. Rev. Financ. Econ.* **2**(1) 103–125.