

Dimensional Preferences: Combinatorics and Voting

Clay Thomas
claytont@princeton.edu

Yufei Zheng
yufei@cs.princeton.edu

January 15, 2019

Abstract

Restricted classes of total orders are a common tool used to model the preferences of rational agents. We introduce a new class of ordinal preferences called d -dimensional preferences. In part I of this paper, we prove a sequence of results about d -dimensional preferences, demonstrating how an underlying set of outcomes in $\mathbb{R}_{\geq 0}^d$ can give a natural structure to the set of possible preferences.

We are particularly successful when studying 2-dimensional preferences. In part II of this paper, we apply our theory to the subject of *voting schemes* for 2-dimensional preferences. We show that very good voting schemes exist for these preferences. Specifically, we give an incentive compatible voting scheme satisfying independence of irrelevant alternatives which always selects a Condorcet winner.

1 Introduction

In this project, we study *ordinal preferences*, i.e. preferences which prefer certain outcomes over others, but do not have a quantitative notion of the quality of the outcome. Formally, these are simply total orders. As computer scientists, our first instinct is to make no assumptions, e.g. to assume preferences can be any of the $m!$ distinct total orders on m outcomes. However, this typically does not correspond to reality: the preferences of an agent are typically *caused by something*, i.e. there is an underlying structure explaining the preferences. This is one way to model the idea the preferences are *correlated*.

Consider the following motivating example. Suppose there are m school, and each has two attributes which we assume are objective across all preferences: quality of STEM education and quality of liberal arts education. We can model each school as a point $x \in \mathbb{R}_{\geq 0}^2$, where the first coordinate gives STEM quality and second coordinate gives liberal arts quality. Then it is possible that the school preferences of students actually boil down to a simple preference over STEM and liberal arts, which we can model as a *weighted sum* over the two attributes. For example, a student who prefers 30% STEM and 70% liberal arts would “score” a school $x = (x_1, x_2)$ via the sum $0.3x_1 + 0.7x_2$, and that student prefers school x over school y if $0.3x_1 + 0.7x_2 > 0.3y_1 + 0.7y_2$.

Social choice theory, the study of *voting schemes*, has found much theoretical success in studying restricted classes of preferences. Restricting preferences is a way around various impossibility theorems of social choice theory, which state that certain “good” voting schemes do not exist in general. In this project, we investigate the structure of preferences which arise from “weighted sum” scoring, and apply these results to provide good voting schemes for this restricted class of preferences.

1.1 Outline and highlights of results

We formalize this notion for any number of attributes d , and call the resulting preference model d -dimensional preferences. We setup a framework for studying d -dimensional preferences, identifying the correct geometric objects to use when reasoning about them. These objects give interesting geometric explanations for concepts such as opinion and agreement. Using these tools, we are able to prove a general impossibility result, theorem 8, which shows a qualitative way in which preferences get more complicated as d increases.

When $d = 2$, we achieve a good handle on the combinatorial structure of 2-dimensional preferences, including a tight upper bound on the maximal number of distinct linear preferences corresponding 2-dimensional preferences (theorem 7). We find that $d = 2$ represents a very nice point in the trade-off between expressibility and complexity, with enough preferences to be realistic while still having strong structure.

Next, we turn to study the social choice theory of 2-dimensional preferences¹. We compare and contrast 2-dimensional preferences with *single-peaked preferences*, the classic restricted class used in social choice theory. Each model different phenomenon, but we argue that 2-dimensional preferences are more expressive in some ways, and certainly more natural in some cases. We find that many positive results which holds for single-peaked preferences hold for 2-dimensional preferences. In particular, we consider the voting scheme which takes preference weight vectors a_1, a_2, \dots, a_n as inputs, and outputs the favorite outcome of the median-angled (from the positive x axis) vector u . It turns out that this mechanism satisfies essentially all of the basic desirable properties for a voting scheme.

¹We have tried to add sufficient references so that Part II is readable without studying Part I in detail, and just occasionally glancing back for the lemmas needed. Propositions 4 and 6 are the most important for section 8.

Part I

The Combinatorics of d -Dimensional Preferences

2 Definition of d -dimensional preferences

Let $\{x_1, \dots, x_m\} = X \subseteq \mathbb{R}_{\geq 0}^d$ be any set of m distinct points with nonnegative coordinates. We call X the set of *outcomes*. Given any $a \in \mathbb{R}_{\geq 0}^d$, which we call a *preference weight*, define a order $>_a$ on X as follows: $x_i >_a x_j$ if and only if $\langle a, x_i \rangle > \langle a, x_j \rangle$. Let $R(X)$ denote the set of total orders on X . Define

$$P_d(X) = \{> \in R(X) \mid \exists a \in \mathbb{R}_{\geq 0}^d : x > y \iff x >_a y\}$$

Note that we consider only total orders, in particular, ties are not allowed. Some first observations:

- If $d = 1$, then $|P(X)| = 1$, i.e. preferences are completely determined by the underlying set X .
- If $d = n$, then every linear preference on X can occur in $P(X)$.

Proof. Let $X = \{e_i\}$ simply be the standard basis vectors. To induce an ordering i_1, i_2, \dots, i_n , just create a preference vector a which gives weight $1/k$ to coordinate i_k . \square

The above hints that there should be some sort of continuum between $d = 1$ and $d = n$ of how complex a preference set $P_d(X)$ can be. We will find that $d = 2$ is a particular sweet spot between expressibility and simplicity. A core piece of intuition to keep in mind for this paper is that a collection of d -dimensional preferences is really a $d - 1$ dimensional space: preferences $>_a$ are invariant under $\|a\|$, so without loss of generality we can take all preferences to lie on the unit sphere. In particular, 2-dimensional preferences are a well-behaved one dimensional space.

3 Lemmas

In this section, we set up the tools needed to reason about d -dimensional preferences.

3.1 Lemmas based on the structure of X

These first few lemmas relate geometric properties of X to limitations on the structure of $>_a$ for a specific, fixed a .

The most simple way X gives structure to preferences is if one outcome is better in all attributes.

Definition. Let $x, y \in X \subseteq \mathbb{R}_{\geq 0}^d$. We say x dominates y , denoted $x \gg y$, if $x[k] > y[k]$ for each $k = 1, \dots, d$.

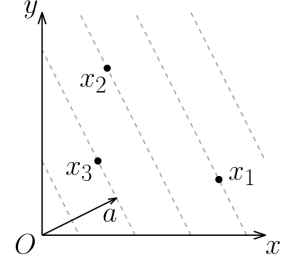


Figure 1: The “level sets” of the preference $>_a$, which ranks outcomes according to distance from the origin along direction a .

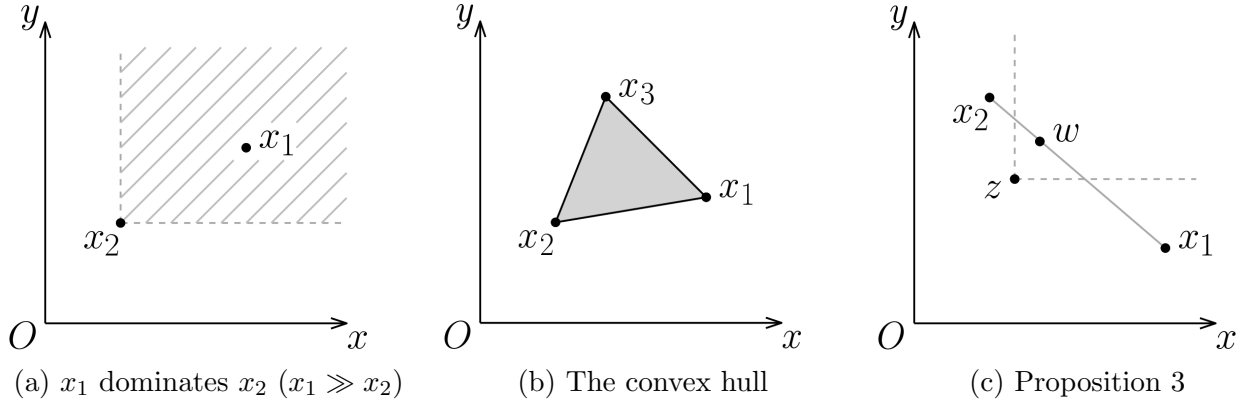


Figure 2

Proposition 1. *If $x \gg y$, then $x >_a y$ for any nonzero $a \in \mathbb{R}_{\geq 0}^d$.*

Proof. Simply observe $a_i x_i \geq a_i y_i$ for each $i = 1, \dots, d$. Because $a \neq 0$, there is also some i where $a_i x_i > a_i y_i$. \square

When comparing different outcomes, a useful tool is the familiar geometric notion of a convex hull. Intuitively, if a preference weight does not like any of a set of options, it will not like any outcome in the hull of those options either. Thus, a point “dominated by the hull” of a set of options (as in figure 9c) cannot be preferred to all those options.

Definition. *For points $x_1, \dots, x_n \in \mathbb{R}^d$, let $\text{hull}(x_1, \dots, x_n) = \{u_1 x_1 + \dots + u_n x_n \mid 0 \leq u_i \leq 1, \sum_{i=1}^n u_i = 1\}$ denote the convex hull of x_1, \dots, x_n .*

Lemma 2. *Let $z, x_1, \dots, x_k \in \mathbb{R}_{\geq 0}^d$ and $a \in \mathbb{R}_{\geq 0}^d \setminus \{0\}$. If $z >_a x_i$ for $i = 1, \dots, k$, then $z >_a w$ for any $w \in \text{hull}(x_1, \dots, x_k)$.*

Proof. We have $\langle a, z \rangle > \langle a, x_i \rangle$ for each $i = 1, \dots, k$. If $w = u_1 x_1 + \dots + u_n x_n$ and $\sum_i u_i = 1$, then $\langle a, w \rangle = u_1 \langle a, x_1 \rangle + \dots + u_n \langle a, x_n \rangle < u_1 \langle a, z \rangle + \dots + u_n \langle a, z \rangle = \langle a, z \rangle$. \square

Proposition 3. *Let $z, x_1, \dots, x_k \in \mathbb{R}_{\geq 0}^d$. Suppose that there exists $w \in \text{hull}(x_1, \dots, x_k)$ such that $w \gg z$. Then no a satisfies $z >_a x_i$ for each $i = 1, \dots, k$.*

Proof. For contradiction, suppose such an a exists. Then $z >_a w$ as well. However, because $w \gg z$, this is a contradiction. \square

Note that lemma 2 and proposition 3 both hold when you reverse all the inequalities in their statements, via the same arguments.

3.2 Lemma relating different preference vectors

When comparing different preference weights, a useful tool is the familiar geometric notion of a convex cone. Intuitively, if a set of weights agree about a certain preference, so does every weight in their cone.

Definition. *For any vectors $a_1, \dots, a_k \in \mathbb{R}_{\geq 0}^d$, define*

$$\text{cone}(a_1, \dots, a_k) = \{u_1 a_1 + \dots + u_k a_k \mid u_i \geq 0 \forall i\}$$

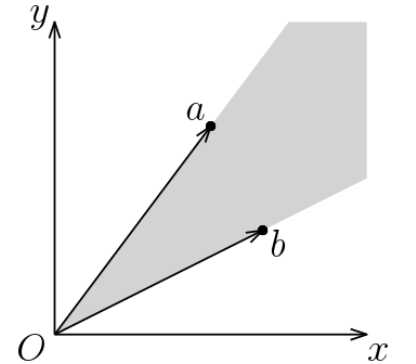


Figure 3: The convex cone

Proposition 4. Suppose that for preference weights a_1, \dots, a_k , we have $x >_{a_1} y, \dots, x >_{a_k} y$. Then for any nonzero $b \in \text{cone}(a_1, \dots, a_k)$, $x >_b y$ as well.

Proof. Let $b = u_1 a_1 + \dots + u_k a_k$. We get $\langle b, x \rangle = u_1 \langle a_1, x \rangle + \dots + u_k \langle a_k, x \rangle > u_1 \langle a_1, y \rangle + \dots + u_k \langle a_k, y \rangle = \langle b, y \rangle$, as desired. \square

3.3 Lemmas for 2 dimensional preferences

This section turns to our main focus: 2 dimensional preferences. Given two points x, y in our set of outcomes X , we give a simple classification of which preference weight prefer x and which prefer y . To do so, we introduce the notion of a weight “indifferent between x and y ”.

Lemma 5. Let $x, y \in \mathbb{R}_{\geq 0}^2$ be distinct points where neither x nor y dominates the other. There exists a unique unit vector $b \in \mathbb{R}_{\geq 0}^2$ such that $\langle b, x \rangle = \langle b, y \rangle$.

Proof. Without loss of generality let $x_1 < y_1, x_2 > y_2$ (if either of these are equalities, the standard basis vectors e_1 or e_2 will do). There exists exactly one unit vector b in $\mathbb{R}_{\geq 0}^2$ such that $b_1/b_2 = (x_2 - y_2)/(y_1 - x_1)$. It’s easy to check that this is a necessary and sufficient condition for having $\langle b, x \rangle = \langle b, y \rangle$. \square

Definition. For $x, y \in \mathbb{R}_{\geq 0}^2$ where neither dominates the other, let the “indifferent vector of x and y ”, denoted $\text{indif}(x, y)$, be the unit vector $b \in \mathbb{R}_{\geq 0}^2$ such that $\langle b, x \rangle = \langle b, y \rangle$.

This next lemma says that the indifference vector of x and y divides the space of preferences into two halves: one that prefers x , one that prefers y . This division is given by another very natural ordering, this one for preference weights. The following definition has an obvious interpretation by considering $\theta(a)$ to be the angle of a from the positive x axis, but we use an equivalent algebraic definition to simplify our proofs. Of course, feel free to ignore this definition, and just use the more natural interpretation of $\theta(\cdot)$ as a function which gives the angle.

Definition. We say a has a larger angle than b , denoted $\theta(a) > \theta(b)$, if $a_2/a_1 > b_2/b_1$.

Proposition 6. Consider any pair of points $x, y \in \mathbb{R}_{\geq 0}^2$, where $x_1 < y_1$ and $x_2 > y_2$, and let $b = \text{indif}(x, y)$. Then for any preference weight vector $a \in \mathbb{R}_{\geq 0}^2$,

1. $\theta(a) > \theta(b)$ if and only if $x >_a y$.
2. $\theta(a) < \theta(b)$ if and only if $y >_a x$.

Proof. Except for the edge case where one vector is a multiple of the other, the two items are exactly the same statement. We prove the first one:

$$\begin{aligned}
 x >_a y &\iff a_1 x_1 + a_2 x_2 > a_1 y_1 + a_2 y_2 \\
 &\iff \frac{a_2}{a_1} > \frac{y_1 - x_1}{x_2 - y_2} = \frac{b_2}{b_1} \\
 &\iff 0 > a_1 b_2 - a_2 b_1 \\
 &\iff \theta(a) > \theta(b)
 \end{aligned}$$

\square

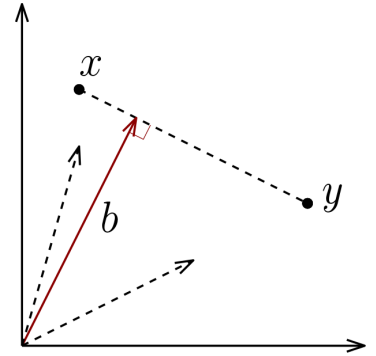


Figure 4: Proposition 6. $b = \text{indif}(x, y)$. Preferences above b (those with higher angle) all prefer x , and those below b (those with lower angle) all prefer y

4 Bounding the number of 2-dimensional preferences

We are able to obtain a tight upper bound on the number of distinct linear preferences which arise from a given two-dimensional outcome set X . In particular, note that $|P_2(X)|$ is much less than $m!$, the total number of linear orders on X .

Theorem 7. *Let $X \subseteq \mathbb{R}_{\geq 0}^2$ with $|X| = m$. We have*

$$|P_2(X)| \leq \binom{m}{2} + 1$$

Proof. Consider the following set of preference weights:

$$B = \{\text{indif}(x, y) \mid x, y \in X \text{ and neither dominates the other}\}$$

Consider the different regions of $\mathbb{R}_{\geq 0}^2$ separated by elements of B . That is, consider the equivalence classes of the relation \sim on $\mathbb{R}_{\geq 0}^2 \setminus B$, where $a \sim a'$ means that $\forall b \in B : \theta(a) < \theta(b) \iff \theta(a') < \theta(b)$. We claim that if $a \sim a'$, then $x >_a y \iff x >_{a'} y$ for all $x, y \in X$. Proof: If $x \gg y$ or $y \gg x$, the order between x and y fixed for any a . Otherwise, $\text{indif}(x, y) \in B$, so by proposition 6, this means $x >_a y \iff x >_{a'} y$.

Thus, the mapping from $\mathbb{R}_{\geq 0}^2 \rightarrow R(X)$ given by $a \mapsto (>_a)$ is constant on equivalence classes of \sim . This map is surjective by definition. Furthermore, a, a' in distinct equivalence classes of \sim give distinct preferences $>_a, >_{a'} \in R(X)$, again by proposition 6, because a and a' are on a different side of some $b \in B$ under the θ order. Thus, there is a bijection between equivalence classes of \sim and preferences in $P_2(X)$.

Note that $|B| \leq \binom{m}{2}$, so \sim divides the $\mathbb{R}_{\geq 0}^2$ into at most $\binom{m}{2} + 1$ regions based on the θ order. Thus $|P_2(X)| \leq \binom{m}{2} + 1$, as desired. \square

Furthermore, this upper bound is tight. Notice that the proof of the upper bound gives a simple condition for the upper bound to be tight: no point of X should dominate another, and every pair of points $x, y \in X$ should give a distinct indifference vector $\text{indif}(x, y)$. To construct such an X with m outcomes, consider $X' = \{(i, m+1-i) \mid i = 1, \dots, m\}$, and randomly perturb each point a bit such that all $\text{indif}(x, y)$ vectors are distinct².

²If you want to get really abstract and fancy, let $\alpha_1, \dots, \alpha_n$ be real numbers which are algebraically independent over \mathbb{Q} , and such that $|\alpha_i| < 1/3$. Then the points $\{(i, m+1-i+\alpha_i)\}_{i=1}^m$ cannot possibly yield a repeated indifference vector, by algebraic considerations.

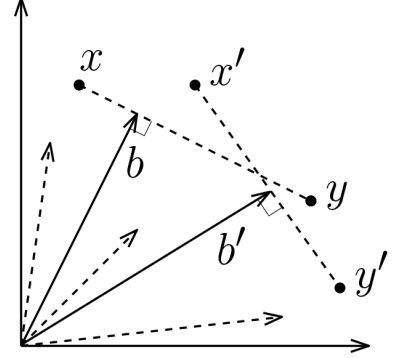


Figure 5: Theorem 7. Indifference vectors b and b' divide preference weight into three regions, which each have the same preferences for x vs. y and x' vs. y' .

5 Impossibility Results

5.1 Large preference cycles are impossible in any dimension

The following preference set may be important in high dimension. We call it k -CYCLE:

$$\begin{array}{cccccccc}
 1 & > & 2 & > & 3 & > & \dots & > & k-2 & > & k-1 & > & k \\
 2 & > & 3 & > & 4 & > & \dots & > & k-1 & > & k & > & 1 \\
 3 & > & 4 & > & 5 & > & \dots & > & k & > & 1 & > & 2 \\
 & & & & & & & & & \vdots & & & & \\
 k-1 & > & k & > & 1 & > & \dots & > & k-4 & > & k-3 & > & k-2 \\
 k & > & 1 & > & 2 & > & \dots & > & k-3 & > & k-2 & > & k-1
 \end{array}$$

Theorem 8. $(d+1)$ -CYCLE is not d -dimensional for any d .

Proof. Suppose for contradiction there existed $a_1, \dots, a_{d+1} \in \mathbb{R}_{\geq 0}^d$ such that $>_{a_i}$ yielded the i th line of $(d+1)$ -CYCLE (in particular, the line whose favorite point is i). The vectors $\{a_i\}$ cannot be linearly independent. Thus, there exists a linear combination of vectors $u_1 a_1 + \dots + u_{d+1} a_{d+1} = 0$. Let $S = \{a_i | u_i > 0\}$ and let $T = \{a_i | u_i < 0\}$. Note that, because every coordinate of each a_i is nonnegative (and no a_i is zero) neither S nor T are empty. We have

$$\sum_{a_i \in S} u_i a_i = \sum_{a_i \in T} -u_i a_i$$

Denote this above vector by b , and note that $b \in \mathbb{R}_{\geq 0}^d \setminus \{0\}$ and $b \in \text{cone}(S) \cap \text{cone}(T)$.

We claim that b satisfies the following:

$$1 >_b 2 >_b 3 >_b \dots >_b d >_b d+1 >_b 1$$

Proof: For each pair $i > i+1$, observe that the inequality is satisfied for all the vectors a_i except for one of them (namely, a_{i+1} which ranks $i+1$ highest). Thus, for either S or T , every vector a_i in the set has the opinion $i >_{a_i} i+1$. Thus, every vector in the cone of that set has the opinion $i > i+1$ as well, by proposition 4. In particular, $i >_b i+1$. Of course, the above argument works for the pair $d+ > 1$ as well.

Thus, transitivity gives us $1 >_b 1$, a contradiction. \square

This is one of our few results which work for any dimension. We'll see that 3-Cycle being impossible for 2 dimensional preferences is a necessary condition for many of the nice properties of 2 dimensional preferences (especially theorem 19). Intuitively, it is possible that "cyclic disagreements" are a core issue for different applications with ordinal preferences (we consider voting here, but other cases include matching or resource allocation). The impossibility of the $d+1$ cycle in d dimensions gives a very nice example of preference sets getting more complicated as dimension increases.

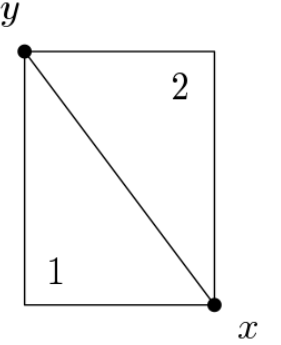
5.2 Combinatorial impossibilities for 2 dimensional preferences

Proposition 9. Let $P = P_2(X)$ be any set of 2 dimensional preferences, where $|X| \geq 3$. It is not possible for every $x \in X$ to be both the favorite and least favorite outcome of some preference in P .

Proof. If any point dominates another, then the dominated point cannot be favorite. Thus we may assume that no point in X is dominated by any other point. Pick two points $x \neq y$ such that x has the largest first coordinate (i.e. x_1) among all of X , and y has the highest second coordinate (i.e. y_2). Consider figure 6. Any remaining point $z \neq x, y$ must lie in region 1 or 2, as all other regions either violate the assumption that $x_1 > z_1$ and $y_2 > z_2$ or cause one point to dominate another. If z lies in region 1, it can never be the favorite, by proposition 3, because it is dominated by some point in $\text{hull}(x, y)$. If z lies in region 2, it can never be the least favorite, this time by the opposite but completely analogous version of proposition 3, because z dominates a point in $\text{hull}(x, y)$. For the corner case where $z \in \text{hull}(x, y)$, observe that z is *always* the middle option among x and y . \square

This proposition, combined with Theorem 8, gives a complete combinatorial classification of two dimensional preferences on exactly three outcomes. Consider all preferences on three outcomes, organized with the two copies of 3-CYCLE in different columns.

$1 > 2 > 3$	$1 > 3 > 2$
$2 > 3 > 1$	$3 > 2 > 1$
$3 > 1 > 2$	$2 > 1 > 3$



In order to make our preference set 2 dimensional, we need to exclude one preference from each cycle. By relabeling, we can without loss of generality assume that the preference with $1 > 2 > 3$ is not in P , and consider removing a preference from the other cycle. If $1 > 3 > 2$ or $2 > 1 > 3$ are removed from P , then P will not violate proposition 9³. If we only remove $3 > 2 > 1$, then every outcome is some preference's favorite and some other preference's least favorite⁴. So this P is not two dimensional.

Figure 6: Proposition 9

³We will see later in section 7 that these preference sets are 2 dimensional. Up to relabeling, they are BADCOMPROMISE and GOODCOMPROMISE, respectively.

⁴This gives SANDWICH from section 7.

Part II

Voting With 2 Dimensional Preferences

6 Definitions

6.1 Voting

Consider a set of preferences P (i.e. linear orders) on a set of outcomes M . We follow the convention that m is the number of outcomes and n is the number of voters. We have the following definitions:

- A *social welfare function on P* is a function $F : P^n \rightarrow P$
- A *social choice function on P* is a function $f : P^n \rightarrow M$. We also call such an f a *voting scheme* or a *voting rule*.
- A welfare function F is *unanimous* if, for any $\succ \in P$, we have $F(\succ, \dots, \succ) = \succ$
- A welfare function F is a *dictatorship* if there exists an i such that $F(\succ_1, \dots, \succ_n) = \succ_i$
- A choice function f is a *dictatorship* if there exists an i such that $f(\succ_1, \dots, \succ_n) = a_i$, where a_i is the favorite outcome of \succ_i
- A welfare function F satisfies *independence of irrelevant alternatives* (IIA) if, whenever $a \succ_i b \iff a \succ'_i b$ and $\succ = F(\succ_1, \dots, \succ_n), \succ' = F(\succ'_1, \dots, \succ'_n)$, we get $a \succ b \iff a \succ' b$. That is, if no voter changes their opinion between a and b , the group preference between a and b does not change.
- A choice function f is *incentive compatible* if, for any $\succ_1, \dots, \succ_n, i, \succ'_i$, we have $f(\succ_1, \dots, \succ_i, \dots, \succ_n) \succeq_i f(\succ_1, \dots, \succ'_i, \dots, \succ_n)$. That is, no voter can get an outcome he strictly prefers by giving a false input to f .
- A collection of preferences \succ_1, \dots, \succ_n , has a Condorcet winner $x \in M$ if, for any other $y \in M$, we have $x \succ_i y$ for more than half of the indices $i = 1, \dots, n$

Let $R(M)$ denote the set of all linear orders on M . Recall that when $P = R(M)$, there are many known impossibility results, the two most famous of which are:

Theorem 10 (Arrow's Impossibility Theorem [Nis07]). *If $|M| \geq 3$, then every unanimous social welfare function on $R(M)$ which satisfies independence of irrelevant alternatives is a dictatorship.*

Theorem 11 (Gibbard Satterthwaite [Nis07]). *If $|M| \geq 3$, then every surjective, incentive compatible social choice function on $R(M)$ is a dictatorship.*

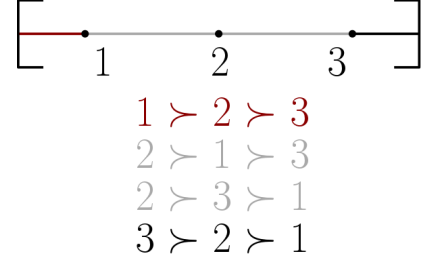
6.2 Single-peaked preferences

Our main point of contrast will be the well-understood class of *single peaked* preferences.

Let $S \subseteq [0, 1]$ be a finite set of m points in the unit interval. We call S the set of *outcomes*. Let $R(S)$ denote the set of linear orders on S . A preference $\succ \in R(S)$ is called *single peaked* if there exists an outcome $p \in S$ (called the “peak” of \succ) such that $x < y < p \implies x \prec y$ and $p < y < x \implies x \prec y$. In other words, the preference has a favorite outcome, and its opinion

strictly decreases as you move farther away from the favorite. Note that no assumption is made about outcomes on different sides of the peak. Define

$$P_{sp}(S) = \{ \succ \in R(S) \mid \succ \text{ is single peaked } \}$$



Single-peaked preferences are studied because the order structure on S gives a natural voting mechanism: poll voters asking for their favorite outcome, i.e. their peak, and output the median peak. It is known that this voting scheme satisfies independence of irrelevant alternatives (so Arrow's theorem does not hold [Nis07]) and that this voting scheme is incentive compatible (so Gibbard Satterthwaite does not hold [SV07]). As we will see, a voting scheme inspired by this will perform well for 2-dimensional preferences, achieving the same positive results as single-peaked preferences.

Figure 7: All single peaked preferences on 3 outcomes

7 Single-peaked versus 2-dimensional preferences

When there are exactly three outcomes, 2-dimensional preferences have strictly more expressive power than single-peaked preferences.

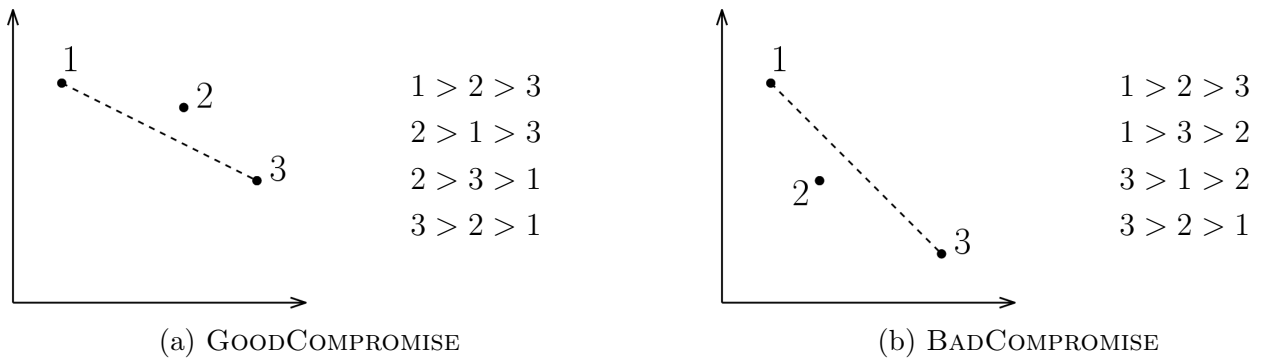
Proposition 12. *Every single-peaked preference set on $m = 3$ outcomes is a 2-dimensional preference set. In particular, up to relabeling it is a subset of $P_2(X)$, for $X = \{(3, 0), (2, 2), (0, 3)\}$.*

Proof. Without loss of generality, relabel the outcomes of the single peaked preferences as $S = \{1, 2, 3\}$. The resulting preferences, as shown in figure 7, are given by GOODCOMPROMISE. \square

Furthermore, BADCOMPROMISE gives an example of a preference set which is 2-dimensional, but not single-peaked:

Proposition 13. *In a single peaked set of preferences, it is not possible for every outcome to be the lowest ranked outcome of some preference.*

Proof. The median outcome (via the standard order on $[0, 1]$) cannot be the lowest ranked. \square



When $m > 3$, neither single-peaked nor 2-dimensional preferences are a subset of the other. One way to see this is to simply count the number of single-peaked preferences, and see that it grows much faster with m than 2-dimensional preferences (recall that we showed in theorem 7 that the maximal size of a 2-dimensional preference set is $O(m^2)$).

Proposition 14. *The number of single-peaked preferences for any set S of m outcomes is at least $2^{\Omega(m)}$*

Proof. Choose the median outcome of S to be the peak. Given a subset $T \subseteq [m - 1]$ of size $|T| = \lfloor m/2 \rfloor$, we can associate a unique single-peaked preference as follows: treat the preference as an array of outcomes, ranked highest to lowest. Put the median of S at index 0. Let the outcomes to the left of the median occupy the indices corresponding to set T , and let those to the right occupy the other indices. There are $\binom{m-1}{\lfloor m/2 \rfloor} \geq 2^{\Omega(m)}$ such subsets T . \square

A cleaner, more constructive way to see this is the following proposition

Proposition 15. *The preference set FLIPFLOP is single-peaked, but not 2-dimensional.*

FLIPFLOP

Proof. To obtain a single-peaked representation, order the outcomes from left to right as follows: 3, 1, 2, 4. All preferences will then be possible with peak 1 or 2.

1 > 2 > 3 > 4
1 > 2 > 4 > 3
2 > 1 > 3 > 4
2 > 1 > 4 > 3

Now we show that FLIPFLOP is not 2-dimensional. Note that neither 1 nor 2 can dominate the other, and the same holds for 3 and 4. So let $a = \text{indif}(1, 2)$ and $b = \text{indif}(3, 4)$. Without knowing the outcome space $X \subseteq \mathbb{R}_{\geq 0}^2$, we can't tell which side of a corresponds to $1 > 2$ versus $2 > 1$ (same with b). However, we do know that at least one of the following implications should hold for any preference $>$ over X :

$$1 > 2 \implies 3 > 4 \quad 1 > 2 \implies 4 > 3 \quad 2 > 1 \implies 3 > 4 \quad 2 > 1 \implies 4 > 3$$

For example, if $\theta(a) < \theta(b)$ and preference weights with smaller angle than a (resp. b) favor $1 > 2$ (resp. $3 > 4$), then $1 > 2 \implies 3 > 4$.

However, none of these implications are satisfied by the entire preference set FLIPFLOP. Thus FLIPFLOP cannot be 2-dimensional. \square

When considering restricted domains of preference, it is perhaps more important to consider which preference sets *do not* fall into that domain. Indeed, if too many preferences are possible, then the impossibility results which hold for arbitrary preferences will apply.

As we've directly shown, 3-CYCLE is not 2 dimensional. This is already a good sign: 3-CYCLE is a classic example of the so-called Condorcet paradox. Despite each preference being a linear order, the majority of voters favor 1 over 2, 2 over 3, and 3 over 1. That is, "collective preference" is cyclic. Indeed, 3-CYCLE is a preference set with no Condorcet winner, and is often used as a basic counterexample for the existence of good voting schemes. The fact that cycle is impossible already seems like good news for voting schemes on two dimensional preferences.

Here's another case:

Proposition 16. *SANDWICH is neither single-peaked nor 2 dimensional*

Proof. SANDWICH is not single-peaked by proposition 13. It's not 2-dimensional by proposition 9. \square

REVERSE	3-CYCLE	SANDWICH		
		$1 > 2 > 3$	2-dimensional	BADCOMPROMISE
$1 > 2 > 3 > 4$	$1 > 2 > 3$	$1 > 3 > 2$	single peaked	FLIPFLOP
$4 > 3 > 2 > 1$	$2 > 3 > 1$	$3 > 2 > 1$	both	REVERSE, GOODCOMPROMISE
	$3 > 1 > 2$	$2 > 3 > 1$	neither	3-CYCLE, SANDWICH

The above discussion gives a fairly clear picture of the relationship between 2-dimensional and single-peaked preferences. While single peaked and 2-dimensional model noticeably different causes of preferences, we argue that 2-dimensional preferences are in some sense more natural. The rest of this section gives some non-formal justification for this highly subjective statement.

In the “school ranking” example given in the introduction, the preference of students is determined by a single underlying attribute of the student: their preference for STEM vs liberal arts. In contrast, single-peaked preferences allow voters to have arbitrary preferences about outcomes on either side of their peak. Two voters with the same peak could have very different preference sets: one could like all the points below the peak before any above the peak, while the other could simply rank options by distance (for example, usual real number distance) from the peak. These two voters seem incompatible within the same underlying situation: one has a sort of “hard budget” at its peak, while the other just wants to get close to his favorite. A common motivation given for single-peaked preferences is that of political policy (with the familiar “left and right” manifest literally). But why would a voter’s favorite policy be the most liberal one which he does not hate? This corresponds to the voter who prefers all outcomes right of his peak to all those left of his peak.

Admittedly, for $m \gg 3$, the size of single-peaked preference sets can be much larger than that of 2-dimensional preferences. However, we believe that single-peaked preferences still miss certain well motivated preference sets, without gaining much essential. Indeed, the ordinal preference set BADCOMPROMISE models the intuitive idea of a bad compromise very well: the “compromise” outcome 2 is no voter’s favorite⁵. However, single-peaked preferences such as FLIPFLOP have no obvious interpretation, except for the voters making arbitrary decisions about outcomes on different sides of their peak.

8 Voting for 2-dimensional preferences

Suppose the set of outcomes X is known. Taking inspiration from single-peaked preferences, we define the following mechanism for voting:

Algorithm 1 Median-angle voting scheme

Poll the preference weight $a_i \in \mathbb{R}_{\geq 0}^d$ from each voter i

Reorder the list a_i by the θ order, so $\theta(a_i) < \theta(a_{i+1})$

Return the favorite outcome of the median-angled preference weight

Intuitively, this seems like a promising voting scheme. Preferences are more similar when angles are closer, and preference angle gives a continuum between one extreme and the other. Thus, the median voter is a natural choice to represent the population. It turns out that the framework

⁵Indeed, our voting scheme given in section 8 will never select outcome 2 in this case!

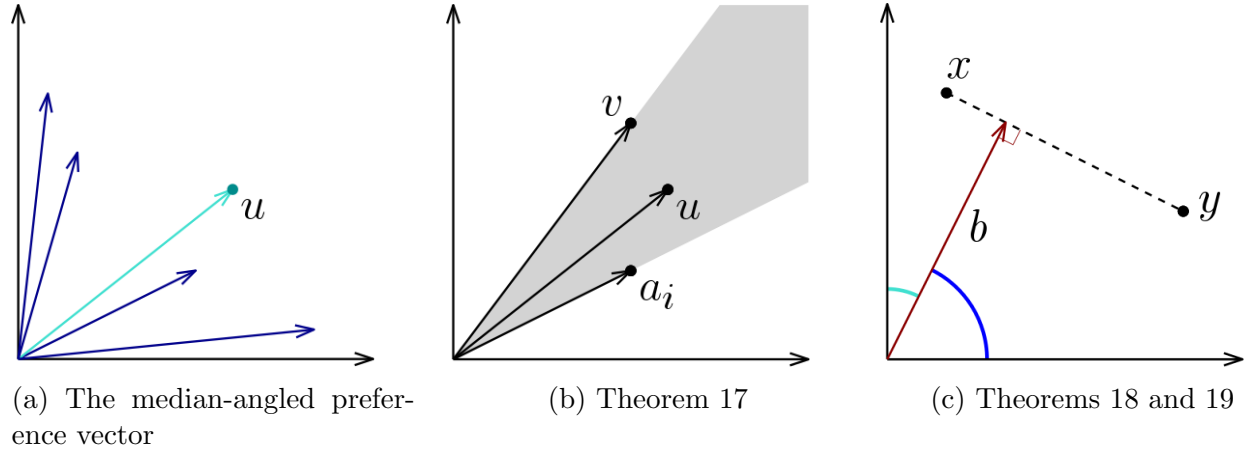


Figure 9

developed in Part I of this paper gives simple proofs of many desirable properties for this voting rule. These properties essentially tell us that 2-dimensional preferences are at least as well-behaved as single-peaked preferences.

Our first result states that a voter cannot strategically manipulate the median-angle voting scheme, i.e. the Gibbard Satterthwaite theorem does not hold for 2-dimensional preferences.

Theorem 17. *Let the number of voters n be odd. The median-angle voting rule for 2-dimensional preferences is incentive compatible.*

Proof. Consider any voter i and suppose the voters true preferences are given by weights a_1, \dots, a_n . Let u denote the median-angled preference vector chosen when all agents report their true preference. Suppose $\theta(a_i) < \theta(u)$ (the other case is symmetric). The only way in which i can change the median-angled preference vector by misreporting his value for a_i is to *increase* its angle, say to some ν . Note that because $\theta(a_i) < \theta(u) < \theta(\nu)$, we have $u \in \text{cone}(a_i, \nu)$.

For the sake of contradiction, assume i could profit from this manipulation, i.e. he could get an outcome he favored chosen by ν . Let y be the outcome selected by u , and let x be selected by ν . We have $x >_\nu y$ and $x >_{a_i} y$. By proposition 4, we get $x >_u y$ as well. However, this contradicts the assumption that u selects y , and completes our proof. \square

We also define the median-angle social welfare function in the obvious way: if the median-angled preference vector is u , return the linear order $(>_u) \in P_2(X)$. Given this definition, we see that median-angled voting satisfies independence of irrelevant alternatives, and thus Arrow's impossibility theorem does not hold for 2-dimensional preferences.

Theorem 18. *The median-angle social welfare function satisfies independence of irrelevant alternatives.*

Proof. Consider any pair of outcomes x, y . If either dominates the other, the IIA axiom becomes easy to verify. So suppose neither dominates the other, without loss of generality take $x_1 < y_1$, $x_2 > y_2$, and let $b = \text{indif}(x, y)$.

Suppose $\{a_i\}$ and $\{a'_i\}$ are collections of preference weights such that $x >_{a_i} y \iff x >_{a'_i} y$ for each i . By proposition 6, this means that $\theta(a_i) > \theta(b) \iff \theta(a'_i) > \theta(b)$. Let u denote the median-angled vector among $\{a_i\}$, and u' for $\{a'_i\}$. Because each pair a_i, a'_i lies on the same side of b , we must have u and u' on the same side of b as well. Thus, $x >_u y \iff x >_{u'} y$, as desired. \square

The next theorem justifies median-angle voting as *the* natural voting scheme for 2-dimensional preferences, using the notion of a Condorcet winner.

Theorem 19. *Among an odd number of 2-dimensional preferences, there is always a Condorcet winner, which is selected by the median angle voting scheme*

Proof. Let u be the median-angled preference vector, and let x be the favorite outcome of u . Given some outcome $y \neq x$, if $y \ll x$, then y is never preferred to x . So suppose neither x nor y dominate each other, and let $b = \text{indif}(x, y)$. Let S be the set of preference weights a such that $x >_a y$. By proposition 6, S either consists of all vectors a with $\theta(a) > \theta(b)$ or all a with $\theta(a) < \theta(b)$. Because $x >_u y$, the median angled preference weight u must lie in S . Thus, more than half of all the input preference weights must lie in S . Thus, more than half of the input preferences agree that $x > y$. Because y was arbitrary, this means x is the Condorcet winner. \square

We note the following important consequence of the previous theorem: if preferences are known to lie in $P_2(X)$ for some X , but no details about X are known, then a voting protocol can simply ask voters for their list of preferences and look for the Condorcet winner (note that there is at most one Condorcet winner). While it is not immediate that this rule is incentive compatible⁶ this still gives a good voting rule for honest agents with two dimensional preferences.

References

- [Nis07] Noam Nisan. Introduction to mechanism design (for computer scientists). In Nisan, Roughgarden, Tardos, and Vazirani, editors, *Algorithmic Game Theory*, chapter 9. Cambridge University Press, Cambridge, 2007.
- [SV07] James Schummer and Rakesh V. Vohra. Mechanism design without money. In Nisan, Roughgarden, Tardos, and Vazirani, editors, *Algorithmic Game Theory*, chapter 10. Cambridge University Press, Cambridge, 2007.

⁶This is because it's possible for a manipulative agent to report a preference list which is not two dimensional with respect to the set X . It's possible for such a manipulator to cause there to be no Condorcet winner (for example, the preference $1 > 2 > 3$ and $2 > 3 > 1$ are certainly two dimensional, so the manipulator could report $3 > 1 > 2$ and create 3-CYCLE). We conjecture that a single manipulator cannot *change* the Condorcet winner, i.e. the manipulated preference set cannot have a Condorcet winner different than the original one, but proving this result requires a bit more thought.