Chapter 12: Social Choice Theory

Felix Munoz-Garcia
School of Economic Sciences
Washington State University[1]

1 Introduction

In this chapter, we consider a society with $I \geq 2$ individuals, each of them endowed with his own preference relation over a set of alternatives, such as the political candidates running for office, or the projects being considered for implementation in a region. We then examine the aggregation of individual preferences into a social preference, that is, a ranking of alternatives for the society. We discuss different methods to aggregate individual preferences and their properties, first in binary sets of alternatives, and then in sets with more than two alternatives (e.g., seventeen candidates competing in the primary of the Republican party to become the party's nominee for U.S. President in the general election). We explore the question, originally posed by Arrow (1953), of whether there exists a procedure to aggregate individual preferences satisfying a minimal set of four desirable properties (spoiler alert: it doesn't exist!).

We then analyze the reactions of the literature to Arrow's inexistence result, which are essentially divided into two groups: those studies that restricted the type of preference relations that individuals can sustain, and those aggregating individual preferences into a social welfare function with a cardinal measure (rather than an ordinal ranking of social preferences). We finish this chapter discussing different voting procedures often used in real life, and their properties.

2 Social welfare functional

Consider a group of $I \geq 2$ individuals and a set of alternatives X. For simplicity, we first consider that set X is binary, thus containing only two elements $X = \{x, y\}$, and later on extend our analysis to sets with more than two alternatives. These two alternatives could represent, for instance, two candidates competing for office, or two policies to be implemented (i.e., the status quo and a new policy). In this binary setting, every individual i's preference over alternatives x and y can be defined as a number:

$$\alpha_i = \{1, 0, -1\}$$

indicating that he strictly prefers x to y when $\alpha_i = 1$, he is indifferent between x and y when $\alpha_i = 0$, or he strictly prefers alternative y to x when $\alpha_i = -1$. Our goal in this chapter is to aggregate

[1] Associate Professor, School of Economic Sciences, Washington State University, 103H Hulbert Hall, Pullman, WA 99164-6210, fmunoz@wsu.edu.

individual preferences with the use of a social welfare functional (also referred to as social welfare aggregator); as defined below.

Social welfare functional. A social welfare functional (swf) is a rule

$$F(\alpha_1, \alpha_2, ..., \alpha_I) \in \{1, 0, -1\}$$

which, for every profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I) \in \{1, 0, -1\}^I$, assigns a social preference $F(\alpha_1, \alpha_2, ..., \alpha_I) \in \{1, 0, -1\}$.

As an example, for individual preferences $(\alpha_1, \alpha_2, \alpha_3) = (1, 0, 1)$ where individuals 1 and 3 strictly prefer x over y while individual 2 is indifferent, the swf could be $F(1, 0, 1) = 1$, thus preferring alternative x over y at the aggregate level. We next describe different properties of swfs, and test whether commonly used swfs satisfy or violate such properties.

2.1 Properties of swf

Paretian. A swf is *Paretian* if it respects unanimity of strict preference; that is, if it strictly prefers alternative x when *all* individuals strictly prefer x, i.e., $F(1, 1, ..., 1) = 1$, but strictly prefers alternative y when *all* individuals strictly prefer y, i.e., $F(-1, -1, ..., -1) = -1$.

This property is not demanding, and is satisfied by many swfs. For instance, Weighted voting, Simple voting, and even Dictatorship are three swfs satisfying it, as we discuss next.

Example 12.1. Weighted voting swf. According to this swf, we first add individual preferences, assigning a weight $\beta_i \geq 0$ to every individual, where $(\beta_1, \beta_2, ..., \beta_I) \neq 0$, as in a weighted sum $\sum_i \beta_i \alpha_i \in \mathbb{R}$. Intuitively, the preference of every individual i is weighted by the importance that such society assigns to his preference, as captured by parameter β_i. After finding the weighted sum $\sum_i \beta_i \alpha_i$, we apply the sign operator, which yields 1 when the weighted sum is positive $\sum_i \beta_i \alpha_i > 0$, 0 when the weighted sum is zero $\sum_i \beta_i \alpha_i = 0$, and -1 when it is negative $\sum_i \beta_i \alpha_i < 0$. Hence, the Weighted voting swf can be summarized as

$$F(\alpha_1, \alpha_2, ..., \alpha_I) = \text{sign} \sum_i \beta_i \alpha_i$$

In order to check if this swf is Paretian, we only need to confirm that, when all individuals strictly prefer alternative x to y, the swf also prefers x to y. Indeed,

$$F(1, 1, ..., 1) = 1, \text{ since } \sum_i \beta_i \alpha_i = \sum_i \beta_i > 0;$$

and similarly, when all individuals strictly prefer alternative y to x, the swf also prefers y to x,

$$F(-1, -1, ..., -1) = -1, \text{ since } \sum_i \beta_i \alpha_i = -\sum_i \beta_i < 0. \quad \square$$

Example 12.2. Simple majority. Simple majority is a special case of weighted majority, whereby the swf assigns the same weights to all individuals, i.e., $\beta_i = 1$ for every individual i. In this setting, if the number of individuals who prefer alternative x to y is larger than the number of individuals preferring y to x, then the swf prefers x over y, i.e., $F(\alpha_1, \alpha_2, ..., \alpha_I) = 1$. The opposite argument applies if the number of individuals who prefer alternative y to x is larger than the number of individuals preferring x to y, where the swf aggregating individual preferences prefers y over x, i.e., $F(\alpha_1, \alpha_2, ..., \alpha_I) = -1$. Finally, note that the Simple majority swf is Paretian since Weighted voting is Paretian. Nonetheless, we next test this property as a practice

$$\begin{aligned} F(1, 1, ..., 1) &= 1, \text{ since } \sum_i \beta_i \alpha_i = N > 0; \text{ and} \\ F(-1, -1, ..., -1) &= -1 \text{ since } \sum_i \beta_i \alpha_i = -N < 0. \quad \square \end{aligned}$$

Example 12.3. Dictatorial swf. We say that a swf is *dictatorial* if there exists an agent d, called the dictator, such that, for any profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I)$:

1. when the dictator strictly prefers alternative x to y, $\alpha_d = 1$, the swf also prefers x to y, $F(\alpha_1, \alpha_2, ..., \alpha_I) = 1$; and

2. when the dictator strictly prefers alternative y to x, $\alpha_d = -1$, the swf also prefers y to x, $F(\alpha_1, \alpha_2, ..., \alpha_I) = -1$.

Intuitively, the strict preference of the dictator α_d prevails as the social preference, regardless of the preference profile of all other individuals $i \neq d$. We can, hence, understand the dictatorial swf as a extreme case of weighted voting where the weight assigned to the individual d's preferences is positive, $\beta_d > 0$, but nil for all other individuals in the society, i.e., $\beta_i = 0$ for all $i \neq d$. Finally, note that since the Weighted voting swf is Paretian, then the Dictatorial swf (as a special case of weighted voting) must also be Paretian. Nonetheless, as an extra confirmation, we show it next:

$$\begin{aligned} F(1, 1, ..., 1) &= 1, \text{ since } \sum_i \beta_i \alpha_i = \beta_d > 0; \text{ and} \\ F(-1, -1, ..., -1) &= -1 \text{ since } \sum_i \beta_i \alpha_i = -\beta_d < 0. \quad \square \end{aligned}$$

Symmetry among agents (or anonymity). The swf $F(\alpha_1, \alpha_2, ..., \alpha_I)$ is *symmetric among agents* (or *anonymous*) if the names of the agents do not matter. That is, if a permutation of preferences across agents does not alter the social preference. More precisely, let

$$\pi : \{1, 2, ..., I\} \to \{1, 2, ..., I\}$$

be an onto function (i.e., a function that, for every individual i, identifies another individual j such that $\pi(j) = i$). Then, for every profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I)$, the swf prefers the same alternative with $(\alpha_1, \alpha_2, ..., \alpha_I)$ and with the "permuted" profile of individual preferences $\left(\alpha_{\pi(1)}, \alpha_{\pi(2)}, ..., \alpha_{\pi(I)}\right)$, that is,

$$F(\alpha_1, \alpha_2, ..., \alpha_I) = F\left(\alpha_{\pi(1)}, \alpha_{\pi(2)}, ..., \alpha_{\pi(I)}\right)$$

Intuitively, the name of the individual preferring x over y, preferring y over x or being indifferent between them, does not affect the socially preferred alternative. As a practice, note that simple majority satisfies anonymity, but weighted voting or dictatorship do not. Indeed, the sum $\sum_i \alpha_i$ coincides when we consider the initial preference profile $(\alpha_1, \alpha_2, ..., \alpha_I)$ and when we consider the "permuted" profile of individual preferences $\left(\alpha_{\pi(1)}, \alpha_{\pi(2)}, ..., \alpha_{\pi(I)}\right)$; thus yielding the same social preference when society uses simple majority to aggregate individual preferences.[2] However, the sum $\sum_i \beta_i \alpha_i$, that we obtain when aggregating individual preferences according to weighted voting, can differ. Consider, for instance, a society of three individuals with a profile of individual preferences given by $(\alpha_1, \alpha_2, \alpha_3) = (1, -1, 1)$. According to weighted voting, $\sum_i \beta_i \alpha_i = \beta_1 - \beta_2 + \beta_3$, which is positive if $\beta_1 > \beta_2 - \beta_3$. However, if we apply the following permutation to individuals' idientities, $\pi(1) = 3$, $\pi(2) = 1$, and $\pi(3) = 2$, weighted voting now yields $\sum_i \beta_i \alpha_i = -\beta_1 + \beta_2 + \beta_3$, which is positive if $\beta_1 < \beta_2 + \beta_3$. It is obvious that weighted voting does not yield a positive outcome (thus indicating a social preference for alternative x over y) before and after the permutation. In particular, it yields the same outcome (thus satisfying anonymity) only if β_1 is intermediate, i.e., $\beta_2 - \beta_3 < \beta_1 < \beta_2 + \beta_3$; but it yields different outcomes otherwise (thus violating anonymity). A similar argument applies to the dictatorial swf.

Neutrality between alternatives. The swf $F(\alpha_1, \alpha_2, ..., \alpha_I)$ is *neutral between alternatives* if, for every profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I)$,

$$F(\alpha_1, \alpha_2, ..., \alpha_I) = -F(-\alpha_1, -\alpha_2, ..., -\alpha_I)$$

That is, if we reverse the preferences of all agents, from $(\alpha_1, \alpha_2, ..., \alpha_I)$ to $(-\alpha_1, -\alpha_2, ..., -\alpha_I)$, then the social preference is reversed as well. For instance, if $(\alpha_1, \alpha_2) = (1, 0)$ and $F(1, 0) = 1$, then when we reverse the profile of individual preferences to $(-1, 0)$, the social preference must become $F(-1, 0) = -1$ for the swf to satisfy neutrality between alternatives. Intuitively, this property is often understood as that the swf doesn't a priori distinguish either of the two alternatives. As a practice, check that simple majority voting satisfies neutrality between alternatives.

Positive responsiveness. Consider a profile of individual preferences $(\alpha'_1, \alpha'_2, ..., \alpha'_I)$ where alternative x is socially preferred or indifferent to y, i.e., $F(\alpha'_1, \alpha'_2, ..., \alpha'_I) \geq 0$. Take now a new

[2] For instance, consider a profile of individual preferences $(\alpha_1, \alpha_2) = (1, 0)$ and its permutation $\left(\alpha_{\pi(1)}, \alpha_{\pi(2)}\right) = (0, 1)$ where $\pi(1) = 2$ and $\pi(2) = 1$, i.e., individual preferences are switched between individuals 1 and 2. Then, the simple majority swf yields x as being socially preferred according to both the initial preference profile $(\alpha_1, \alpha_2) = (1, 0)$, $F(1, 0) = 1$ since $\sum_i \alpha_i = 1 + 0 = 1$, and according to the permuted preference profile $\left(\alpha_{\pi(1)}, \alpha_{\pi(2)}\right) = (0, 1)$, i.e., $F(0, 1) = 1$ since $\sum_i \alpha_i = 0 + 1 = 1$.

profile $(\alpha_1, \alpha_2, ..., \alpha_I)$ in which some agents raise their consideration for x, i.e., $(\alpha_1, \alpha_2, ..., \alpha_I) \geq (\alpha'_1, \alpha'_2, ..., \alpha'_I)$ where $(\alpha_1, \alpha_2, ..., \alpha_I) \neq (\alpha'_1, \alpha'_2, ..., \alpha'_I)$, i.e., $\alpha_i > \alpha'_i$ for at least one individual i. We say that a swf is *positively responsive* if the new profile of individual preferences $(\alpha_1, \alpha_2, ..., \alpha_I)$ makes alternative x socially preferred, i.e., $F(\alpha_1, \alpha_2, ..., \alpha_I) = 1$.

In words, condition $(\alpha_1, \alpha_2, ..., \alpha_I) \geq (\alpha'_1, \alpha'_2, ..., \alpha'_I)$ says that at least one of the components of the new preference profile is larger than in the original profile, thus indicating that the consideration of alternative x increased for at least one individual, i.e., $\alpha_i > \alpha'_i$ for at least one individual i. Therefore, positive responsiveness says that if x was socially preferred to y under the initial profile of individual preferences, then x must still be socially preferred under the new profile of preferences. If alternative x was indifferent to y under the initial profile of preferences, this property says that the swf satisfies positive responsiveness if x becomes socially preferred under the new profile of individual preferences.

Example 12.4. A swf satisfying positive responsiveness. Consider an individual preference profile of $(\alpha'_1, \alpha'_2, \alpha'_3) = (1, 0, -1)$, and assume that the swf in this case yields $F(\alpha'_1, \alpha'_2, \alpha'_3) = 0$, i.e., society is indifferent between alternatives x and y. Now, assume that the preference profile increases the consideration of alternative x, implying that the new preference profile can be any of the following[3]

$$\begin{aligned}(\alpha_1, \alpha_2, \alpha_3) &= (1, 1, -1) \\ &= (1, 0, 0) \\ &= (1, 0, 1) \\ &= (1, 1, 0) \\ &= (1, 1, 1)\end{aligned}$$

Then, if the swf selects $F(\alpha_1, \alpha_2, \alpha_3) = 1$, meaning that alternative x is socially preferred to y, then the swf satisfies positive responsiveness. If, instead, the swf for the new preference profile is $F(\alpha_1, \alpha_2, \alpha_3) = 0$ or $F(\alpha_1, \alpha_2, \alpha_3) = -1$, then such a swf violate this property. $\square$

Example 12.5. Testing positive responsiveness. Simple majority voting satisfies positive responsiveness. Indeed, the sum $\sum_i \alpha_i$ under the new profile of preferences is strictly larger than the sum under the initial profile, $\sum_i \alpha'_i$. Hence, when $\sum_i \alpha'_i = 0$ (entailing that x is socially indifferent to y under the initial profile of individual preferences), the sum $\sum_i \alpha_i$ must be positive and thus select alternative x as socially preferred under the new profile of preferences; while when $\sum_i \alpha'_i > 0$, the sum $\sum_i \alpha_i$ must also be positive, thus implying that x was socially preferred to y

[3] In the first preference profile, individual 2 is the only agent increasing his consideration for alternative x (as α_2 increases from 0 to 1); and in the second perference profile, indidual 3 is the only agent increasing his consideration for x (since α_3 increases from -1 to 0). In the third preference profile, individual 3 is the only agent whose consideration for x increases, but now α_3 increases from -1 to 1; while in the fourth and fifth preference profiles, the consideration of individuals 2 and 3 increase.

under both the initial and new profile of individual preferences. As a practice, note that weighted voting does not necessarily satisfy positive responsiveness, since the sum $\sum_i \beta_i \alpha_i$ under the new preference profile is not necessarily strictly larger than that under the initial profile, $\sum_i \beta_i \alpha_i'$, thus allowing for $\sum_i \beta_i \alpha_i = \sum_i \beta_i \alpha_i' = 0$ which implies $F(\alpha_1', \alpha_2', ..., \alpha_I') = F(\alpha_1, \alpha_2, ..., \alpha_I) = 0$. $\square$

2.2 Arrow's impossibility theorem

Let us now extend our analysis to non-binary sets of alternatives X, e.g., three candidates competing for elected office or, more generally, $X = \{a, b, c, ...\}$. In this context, as shown in the Condorcet's paradox, the aggregation of individual preferences using a majority voting swf, or a weighted voting swf, can be subject to non-transitivities in the resulting social preference. That is, the order in which pairs of alternatives are voted can lead to cyclicalities, as shown in Condorcet's paradox (see Chapter 1, where we aggregated different criteria of a prospective student applying to Ph.D. programs, to obtain his individual preference relation). For illustration purposes, we recall Condorcet paradox below.

Condorcet Paradox. Consider a society with three individuals and three alternatives $X = \{x, y, z\}$, where individual preferences are given by

$$\begin{aligned} x &\succsim^1 y \succsim^1 z && \text{for individual 1,} \\ y &\succsim^2 z \succsim^2 x && \text{for individual 2, and} \\ z &\succsim^3 x \succsim^3 y && \text{for individual 3} \end{aligned}$$

First, note that if individual preferences are aggregated according to majority voting, the resulting social preference is intransitive. Indeed, the social preference is

$$x \succsim^{1,3} y \succsim^{1,2} z \succsim^{2,3} x$$

where the superscripts above each preference symbol represent the individuals who sustain these preferences. For instance, in $x \succsim^{1,3} y$, individuals 1 and 3 prefer alternative x to y (while only individual 2 prefers y to x), and thus alternative x would beat y in a pairwise majority voting. Such intransitive social preference would, hence, lead to cyclicalities.

Second, the above profile of individual preferences is subject to agenda manipulation, i.e., the individual controlling the alternatives that are considered first in a pairwise vote, can strategically alter the agenda on his own benefit. In order to see this result, note that if alternatives x and y are confronted using majority voting, alternative x wins as it receives two votes (from voters 1 and 3) while alternative y only receives voter 2's ballot. The winner of this pairwise majority voting, x, is then paired agains the remaining alternative, z, which yields z as the winner, since z receives two votes (from voters 2 and 3) while alternative x only receives voter 1's ballot. Hence, if the parwise vote is first between alternatives x and y with the winner subsequently confronting alternative z, a sophisticated agenda setter could anticipate that alternative z will be declared the winner. (Individual 3, for instance, would have incentives to set such an agenda for pairwise votes.) If, instead, alternatives x and z are paired first, with the winner confronting alternative y, the outcome changes. Indeed, z wins a pairwise vote against x (as it receives votes from 2 and 3), but the winner, z, then loses against the remaining alternative y (as y is preferred to z by voters 2 and 1), thus declaring y as the winner. This agenda would be beneficial for individual 2. A similar argument applies if alternatives y and z are first matched in a pairwise voting, and the winner of this pair confronting afterwards alternative x. In particular, y would beat z since voters 1 and 2 would vote for y, but the pairwise voting between y and the remaining alternative x ultimately yields x as the winner (since voters 1 and 3 vote prefer x over y). Such voting agenda would be particularly attractive to individual 1.

Given the possibility of intransitive social preferences when we use commonly employed voting methods, such as simple and weighted majority, an interesting question is: Can we design voting systems (i.e., swfs that aggregate individual preferences) that are not prone to the Condorcet's paradox and satisfy a minimal set of "desirable" properties? This was the question Arrow asked himself (for his Ph.D. thesis) obtaining a rather grim result: such a voting procedure does not exist!. This result is commonly known as Arrow's impossibility theorem. Before presenting the theorem, let us first define the four minimal requirements that Arrow considered all swfs should satisfy

1. **Unrestricted domain (U).** The domain of the swf, $\left(\succsim^1, \succsim^2, ..., \succsim^I\right)$, must include all possible combinations of individual preference relations on X.

   In other words, we allow any sort of individual preferences over alternatives.

2. **Weak Pareto Principle (WP).** For any pair of alternatives x and y in X, if $x \succ^i y$ for every individual i, then the social preference is $x \succ y$.

   That is, if every single member of society strictly prefers alternative x to y, society should also strictly prefer x to y. The adjective "weak" in this property is used because WP doesn't require society to prefer x to y if, say, all but one strictly prefer x to y, yet one person is indifferent between these alternatives. (In this case, the social preference doesn't necessarily prefers x to y.)

3. **Non-dictatorship (D).** There is no individual d such that, for all pairs of alternatives $(x, y) \in X$, individual d's strict preference of x over y, $x \succ^d y$, implies a social preference of $x \succ y$ regardless of the preferences of all other individuals $d \neq i$, $\left(\succsim^1, \succsim^2, ..., \succsim^{d-1}, \succsim^{d+1}, ..., \succsim^I\right)$.

   Note that this is a very mild assumption. Indeed, we could consider a "virtual" dictatorship in which an individual d imposes his preference on the rest of individuals for all, but one, pair of alternatives. Such a setting would be considered non-dictatorial, since for a dictatorship to arise we must find that the preferences of individual d dictate the social preference for *all* pairs of alternatives. Consider, for instance, a group of two individuals, with the profile of individual preferences depicted in table 12.1. While social preferences coincide with the preferences of individual 1 in most pairs of alternatives (see first two rows of the table), they do not in one pair (see last row). As a consequence, this swf is not dictatorial.

$\succsim^1$	$\succsim^2$	Soc.Pref. $\succsim$
$x \succsim^1 y$	$y \succsim^2 x$	$x \succsim y$
$y \succsim^1 z$	$z \succsim^2 y$	$y \succsim z$
$z \succsim^1 w$	$w \succsim^2 z$	$w \succsim z$

Table 12.1. A "virtual" dictatorship of individual 1.

4. **Independence of irrelevant alternatives (IIA).** Let $\succsim$ be social preferences arising from the list of individual preferences $\left(\succsim^1, \succsim^2, ... \succsim^I\right)$, and $\succsim'$ that arising when individual preferences are $\left(\succsim'^1, \succsim'^2, ... \succsim'^I\right)$. In addition, let x and y be any two alternatives in X. If each individual ranks x versus y under $\succsim^i$ the same way he does under $\succsim'^i$, then the social ranking of x versus y is the same under $\succsim$ than under $\succsim'$.

   Note that the premise of IIA only requires that, if individual i's preferences are $x \succsim^i y$ between alternatives x and y (e.g., his preference in the morning), then his preferences keep ranking these two alternatives in the same way even if his preferences for a third alternative change, $x \succsim'^i y$ (e.g., his preferences in the afternoon).[4] That is, the premise of IIA says that every individual i must rank x and y in the same way under $\succsim^i$ than under $\succsim'^i$. Then, a swf satisfies IIA if the social ranking of x versus y is the same under $\succsim$ than under $\succsim'$. In other words, even if other alternatives different from x and y change their ranking when we move from $\succsim^i$ to $\succsim'^i$, individual preferences for x and y have not changed and, hence, the social preference for x and y should not change either. The following examples illustrates swfs satisfying or violating IIA.

Example 12.6. Swfs satisfying/violating IIA. Consider the preference profile depicted in table 12.2. Suppose that in the morning (left side) some individuals prefer alternative x to y, $x \succsim^i y$, while others prefer y to x, $y \succsim^i x$. However, they all rank alternative z below both x and y. In addition, suppose that the swf yields a social preference of x over y, i.e., $x \succ y$. During the afternoon (right side), alternative z is ranked above both x and y for all individuals. However, the ranking of alternatives x and y did not change for any individual, i.e., if $x \succsim^i y$ then $x \succsim'^i y$, and if $y \succsim^i x$ then $y \succsim'^i x$; as required by the premise of IIA. Then, IIA says that society should still

[4] The preferences of individual $j \neq i$ could, however, be different from those of individual i, i.e., $y \succsim^j x$ and $y \succsim'^j x$ so individual j ranks alternatives x and y in the same way in the morning and afternoon.

strictly prefer x over y in the afternoon, i.e., $x \succ' y$.

Morning			Afternoon		
$\succsim^1$	$\succsim^2$	Soc.Pref. $\succsim$	$\succsim^1$	$\succsim^2$	Soc.Pref. $\succsim$
.	.	.	z	.	z
.	.	.	.	z	.
x	y	x	x	y	x
y	x	y	y	x	y
.	z	z	.	.	.
z	.	.	.	.	.
$x \succsim^1 y$	$y \succsim^2 x$	$x \succsim y$	$x \succsim^1 y$	$y \succsim^2 x$	$x \succsim y$

Table 12.2. Swf satisfying IIA.

In contrast, table 12.3 illustrates a swf violating IIA. While the preferences over alternatives x and y of individuals 1 and 2 remain constant over time (i.e., the premise of IIA holds), the social preference over these alternatives changes from $x \succsim y$ in the morning to $y \succsim x$ in the afternoon. Hence, IIA does not hold.

Morning			Afternoon		
$\succsim^1$	$\succsim^2$	Soc.Pref. $\succsim$	$\succsim^1$	$\succsim^2$	Soc.Pref. $\succsim$
.	.	.	z	.	z
.	.	.	.	z	.
x	y	x	x	y	y
y	x	y	y	x	x
.	z	z	.	.	.
z	.	.	.	.	.
$x \succsim^1 y$	$y \succsim^2 x$	$x \succsim y$	$x \succsim^1 y$	$y \succsim^2 x$	$y \succsim x$

Table 12.3. Swf violating IIA.

Note that if the preference between alternatives x and y changes for at least one individual from the morning to the afternoon, then the premise of IIA does not hold; as illustrated in table 12.4.

In such a case, we cannot claim that IIA is violated.[5] $\square$

Morning			Afternoon		
$\succsim^1$	$\succsim^2$	Soc.Pref. $\succsim$	$\succsim^1$	$\succsim^2$	Soc.Pref. $\succsim$
.	.	.	z	.	z
.	.	.	.	z	.
x	y	x	x	x	y
y	x	y	y	y	x
.	z	z	.	.	.
z	.	.	.	.	.
$x \succsim^1 y$	$y \succsim^2 x$	$x \succsim y$	$x \succsim^1 y$	$y \succsim^2 x$	$y \succsim x$

Table 12.4. Swf for which the premise of IIA does not hold.

Most of these assumptions are often accepted as the minimal assumptions that we should impose on any swf that aggregates individual preferences into a social preference. Let us now describe Arrow's impossibility theorem, which comes as a surprising, even disturbing, result.

Arrow's impossibility theorem. *If there are at least three elements in the set of alternatives X, then there is no swf that simultaneously satisfies properties U, WP, IIA, and D.*

Proof: We will assume that U, WP and IIA hold, and show that all swfs simultaneously satisfying these three properties must be dictatorial, thus violating one of the four properties. (U is used throughout the proof when we alter the profile of individual preferences, since this property allows for all preference profiles to be admissible.)

Step 1: Consider that an alternative c is placed at the bottom of the ranking of every individual i. Then, by WP, alternative c must be placed at the bottom of the social ranking as well; as depicted in the last column of table 12.5.

$\succsim^1$	$\succsim^2$	...	$\succsim^I$	Soc.Pref. $\succsim$
x	x'	...	x''	x'''
y	y'	...	y''	y'''
.	.			.
.	.			.
.	.			.
c	c	...	c	c

Table 12.5. Alternative c is at the bottoom of everyone's ranking.

[5] For us to claim that the swf violates the IIA, we first need that its premise to be satisfied, and the conclusion is violated, i.e., while individual preferences for x and y do not change throughout the day, the social preference for alternatives x and y changes.

Step 2: Imagine now that we move alternative c from the bottom of individual 1's ranking to the top of his ranking, leaving the position of all other alternatives unaffected. Next, we do the same move for individual 2, then for individual 3, etc.; as illustrated in table 12.6. Let individual n be the first for which raising alternative c to the top of his ranking causes the social ranking of alternative c to increase. The following table places alternative c at the *top* of the social ranking, a result we show next.

$\succsim^1$	$\succsim^2$	...	$\succsim^n$	...	$\succsim^I$	Soc.Pref. $\succsim$
c	c	...	c	...	x''	c
x	x'	...		...	y''	.
y	y'					.
.	.					.
.	.					.
.	.					.
w	w'	...		...	c	w'''

Table 12.6. Alternative c is raised to the top of the ranking for $i = 1, 2, ..., n$.

By contradiction, assume that the social ranking of c increases but *not to the top*, i.e., there is at least one alternative α such that $\alpha \succsim c$ and one alternative β for which $c \succsim \beta$, where $\alpha, \beta \neq c$; as depicted in table 12.7 (left panel). Because alternative c is either at the top of the ranking of individuals $1, 2, ..., n$, or at the bottom of the ranking of individuals $n+1, ..., I$, we can change each individual i's preferences so that $\beta \succ^i \alpha$, while leaving the position of c unchanged for that individual; as depicted in the right panel of table 12.7.

$\succsim^1$	$\succsim^2$	...	$\succsim^n$	...	$\succsim^I$	Soc.Pref. $\succsim$
c	c	...	c	...	x''	α
x	x'	...		...	y''	.
y	y'					c
.	.					β
.	.					.
.	.					.
w	w'	...		...	c	w'''

$\succsim^1$	$\succsim^2$	...	$\succsim^n$	...	$\succsim^I$	Soc.Pref. $\succsim$
c	c	...	c	...	β	α
β	β	...	β	...	z	.
.	α					c
.	.				α	β
α	.					.
.	.		α			.
w	w'	...		...	c	w'''

Table 12.7. Alternative c must be at the top of the social ranking.

We have now changed each individual i's preferences so that $\beta \succ^i \alpha$, while leaving the position of c unchanged for that individual, which produces our desired contradiction:

1. On one hand, $\beta \succ^i \alpha$ for every individual which, byt the WP property, must yield a social preference of $\beta \succ \alpha$; and

2. On the other hand, the ranking of alternative α relative to c, and of β relative to c, have not changed for any individual.[6] By the IIA, this result implies that the social ranking of α relative to c, and of β relative to c, must remain unchanged. Hence, the social ranking still is $\alpha \succsim c$ and $c \succsim \beta$. By transitivity, this yields that $\alpha \succsim \beta$.

However, the result from point 1 ($\beta \succ \alpha$) contradicts that from point 2 ($\alpha \succsim \beta$); yielding the desired contradiction. Hence, alternative c must have moved all the way to the top of the social ranking. (Q.E.D.)

In the next, and final, step of the proof, we show that individual n is a dictator, thus imposing his preferences on the group regardless of the preference profile of all other individuals.

Step 3: Consider now two distinct alternatives a and b, each different from c. In the table on individual and social preferences, let's change the preferences of individual n as follows:

$$a \succ^n c \succ^n b$$

For every other individual $i \neq n$, we rank alternatives a and b in any way but keeping the position of c unchanged (see table 12.8).

$\succsim^1$	$\succsim^2$	...	$\succsim^n$	...	$\succsim^I$	Soc.Pref. $\succsim$
c	c	...	a	...	x''	c
x	x'	...	c	...	y''	.
y	y'		b		.	.
.	.				.	a
.	.				.	b
a	b	...		...	a	.
b	a	...		...	b	
.	.					.
w	w'	...		...	c	w'''

Table 12.8. Individual n must be a dictator.

In the new profile of individual preferences, the ranking of alternatives a and c is the same for every individual as it was just *before* raising alternative c to the top of individual n's ranking in Step 2. Therefore, by IIA, the social ranking of alternatives a and c must be the same as it was at that moment (just *before* raising c to the top of individual n's ranking in Step 2). That is, $a \succ c$, since at that moment alternative c was at the bottom of the social ranking. Similarly, in the new profile of individual preferences, the ranking of alternatives c and b is the same for every individual

[6] That is, if an individual i's preference for alternative α and c was $\alpha \succsim^i c$ ($c \succsim^i \alpha$) before chaning $\beta \succ^i \alpha$ for all individuals, his preference is still $\alpha \succsim^i c$ ($c \succsim^i \alpha$, respectively) after the change. A similar argument applies for individual i's preference between β and c, which remain unaffected after imposing the condition of $\beta \succ^i \alpha$ in the ranking of all individuals.

as it was just after raising c to the top of individual n's ranking in Step 2. Hence, by IIA, the social ranking of alternatives c and b must be the same as it was at that moment. That is, $c \succ b$, since at that moment alternative c had just risen to the top of the social ranking. Summarizing, since $a \succ c$ and $c \succ b$, we have that, by transitivity, $a \succ b$.

Then, no matter how individuals different from individual n rank every pair alternatives a and b, the social ranking agrees with individual n's ranking; thus showing that individual n is a dictator which completes the proof.[7] (Q.E.D.)

In summary, we started with a swf satisfying properties U, WP, and IIA, and showed that the social preference must coincide with that of one individual, thus violating the non-dictatorship property (D). Other proofs of this theorem follow a similar route, by considering that three of the four assumptions hold, and showing that the fourth assumption must be violated.[8]

3 Reactions to Arrow's impossibility theorem

After Arrow's negative result to the search of a swf satisfying his four minimal assumptions, the literature reacted using two main approaches: (1) Eliminating the U assumption, by focusing on specific types of individual preferences, such as the single-peaked preferences that we define below; and (2) Aggregating the intensity of individual preferences (not only the ranking of alternatives for each individual) into a social welfare function that captures the intensity in social preferences. This last approach differs from the swf analyzed in previous sections, as that provided us with a ranking of social preferences (i.e., a cardinal measure), while a social welfare function measures with a real number the welfare that society achieves from each allocation of goods and services (i.e., an ordinal measure). We explore each approach in the next two subsections.

3.1 First reaction - Single-peaked preferences

We informally say that preferences of individual i are single-peaked if we can identify a blissing point (or satisfaction point) at which the individual reaches his maximal utility (i.e., his utility "peak"). Formally, the single peak is defined relative to a linear order in the set of available alternatives X. Hence, we next recall the definition of a linear order (a standard definition in math):

Linear order. A binary relation $\geq$ on the set of alternatives X is a *linear order* on X if it is:

1. reflexive, i.e., $x \geq x$ for every $x \in X$;
2. transitive, i.e., $x \geq y$ and $y \geq z$ implies $y \geq z$; and
3. total, i.e., for any two distinct $x, y \in X$, we have that either $x \geq y$ or $y \geq x$, but not both.

[7] That is, while $a \succ^i b$ for some individuals and $b \succ^j a$ for other individuals, the fact that $a \succ^n b$ for individual n implies that $a \succ b$ for the social ranking, which is true for any two alternatives $a, b \neq c$.

[8] Other approaches use figures to provide a more visual representation of the proof; see, for instance, section 2.4 in Gaertner's (2009) book.

If the set of alternatives is a segment of the real line, i.e., $X \subset \mathbb{R}$, then the linear order $\geq$ can be understood as the "greater than or equal to" operator in the real numbers. We are now ready to define single-peaked preferences.

Single-peaked preferences. The rational preference relation $\succsim$ is *single peaked* with respect to the linear order $\geq$ on X if there is an alternative $x \in X$ with the property that $\succsim$ is increasing with respect to $\geq$ on the set of alternatives below x, $\{y \in X : x \geq y\}$, and decreasing with respect to $\geq$ on the set of alternatives above x, $\{y \in X : y \geq x\}$. That is,

$$\begin{aligned} \text{If } x &\geq z > y \text{ then } z \succ y, \text{ and} \\ \text{If } y &> z \geq x \text{ then } y \succ z, \end{aligned}$$

As suggested above, this definition can be understood as that there is an alternative x that represents a "peak" of satisfaction; and that satisfaction increases as we approach this peak (so there cannot be any other peak of satisfaction).

Example 12.7. Single-peaked preferences. Consider a set of policy alternatives $X = [0, 1]$, e.g., percentage of the federal budget spent in education. Every individual i's utility from alternative $x_k \in [0, 1]$ is

$$u(x_k, \theta_i) = -(x_k - \theta_i)^2$$

where $\theta_i \in [0, 1]$ represents individual i's ideal policy. To understand this utility function, note that it collapses to zero when the policy alternative coincides with the individual's policy ideal, $x_k = \theta_i$; but becomes a negative number both when the policy falls below his policy ideal, $x_k < \theta_i$, and when it exceeds his policy ideal, $x_k > \theta_i$. Graphically, the utility function lies in the negative quadrant for all $x_k \in [0, 1]$, except for $x_k = \theta_i$ where the utility becomes zero. This function is then single-peaked, at $x_k = \theta_i$. $\square$

Example 12.8. Single-peaked preferences and convexity. Consider a set of alternatives $X = [a, b] \subset \mathbb{R}$, i.e., a segment of the real line. Then, a preference relation $\succsim$ on X is single peaked if and only if it is *strictly convex*: That is if, for every alternative $w \in X$, and for any two alternatives y and z weakly preferred to w, i.e., $y \succsim w$ and $z \succsim w$ where $y \neq z$, their linear combination is strictly preferred to w,

$$\alpha y + (1 - \alpha) z \succ w \quad \text{for all } \alpha \in (0, 1)$$

For illustration purposes, Figure 12.1 depicts utility functions satisfying (violating) the single-peaked property in the left-hand panel (right-hand panel, respectively). In particular, note that in both panels $u(y) \geq u(w)$ and $u(z) \geq u(w)$; as required by the premise of convexity. The linear combination of y and z yields a utility $u(\alpha y + (1 - \alpha) z)$ that lies above (below) $u(w)$ when the utility function has a single peak (multiple peaks); as depicted in the left (right) panel.

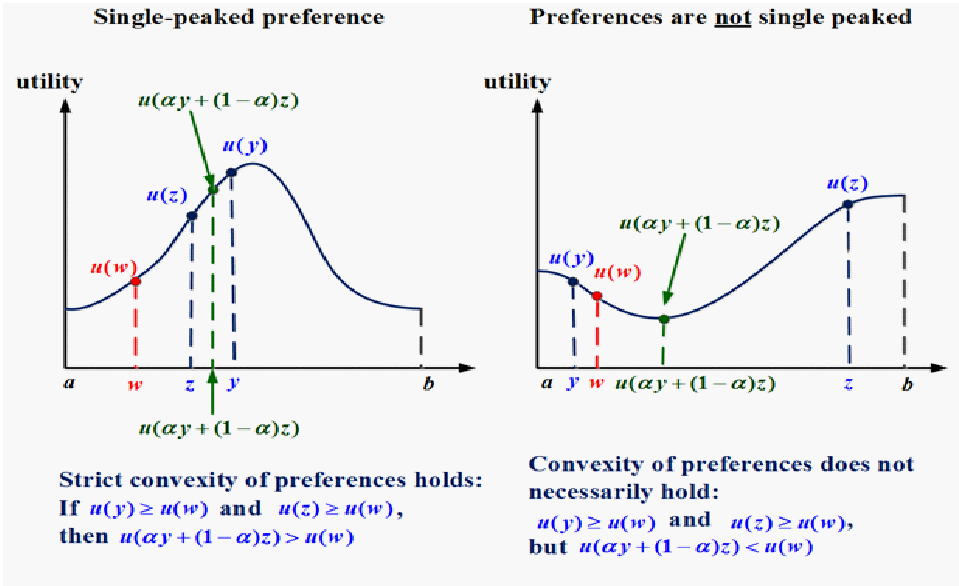

Figure 12.1. Single-peaked preferences and convexity.

Importantly, the single-peaked property is not equivalent to strict concavity in the utility function. Figure 12.2 depicts a utility function that, despite being strictly convex, satisfies the single-peaked property: indeed, $u(y) \geq u(w)$ and $u(z) \geq u(w)$, and the linear combination of y and z yields a utility $u(\alpha y + (1-\alpha)z)$ that lies above $u(w)$. $\square$

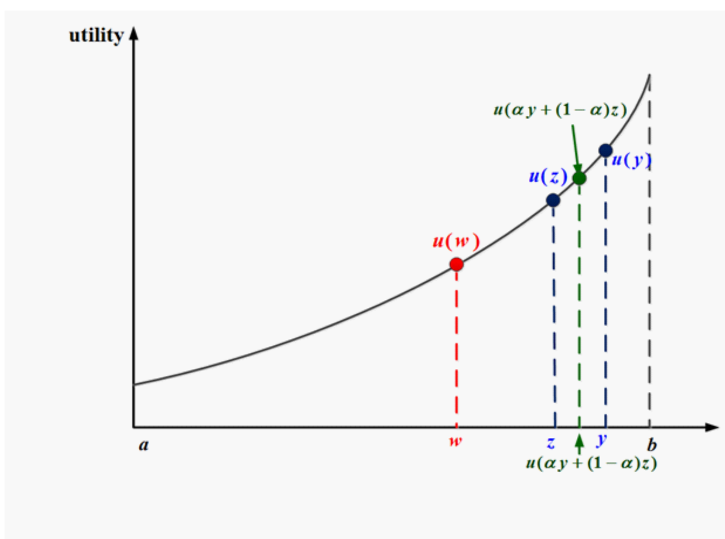

Figure 12.2. Single-peaked preferences and strict concavity.

We will now restrict our attention to settings in which all individuals have single-peaked preferences with respect to the same linear order $\geq$. In this setting, consider pairwise majority voting, which confronts every pair of alternatives x and y against each other, and determines that alternative x is (weakly) socially preferred to y if the number of agents who strictly prefer x to y is larger or equal to the number of agents that strictly prefer y to x. Formally, for any pair of alternatives $\{x, y\} \subset X$, we say that $x\widehat{F}\left(\succsim^1, \succsim^2, ..., \succsim^I\right) y$ to be as "x is weakly socially preferred to y", if

$$\#\left\{i \in I : x \succ^i y\right\} \geq \#\left\{i \in I : y \succ^i x\right\}$$

that is, if the number of votes for alternative x is weakly larger than those to alternative y.

Result. When preferences are single-peaked, the social preferences arising from applying pairwise majority voting have at least a maximal element. That is, there is at least one alternative that cannot be defeated by any other alternatives, i.e, a Condorcet winner exists.

Proof. Before starting our proof, we need a few definitions: First, let x_i denote individual i's peak according to his preference $\succsim^i$. Second, let us define what we mean by a median agent: Agent $m \in I$ is a median agent for the preference profile $\left(\succsim^1, \succsim^2, ..., \succsim^I\right)$ if

$$\#\{i \in I : x_i \geq x_m\} \geq \frac{I}{2} \text{ and } \#\{i \in I : x_m \geq x_i\} \geq \frac{I}{2}$$

That is, the number of individuals whose ideal point is larger than m's ideal point is equal or larger than half of the population. Similarly, the number of individuals whose ideal point is smaller than m's ideal point is equal or larger than half of the population. A natural conclusion of this definition is that, if there are no ties in peaks (i.e., individuals with the same ideal points) and if the number of individuals is odd, then there are exactly $\frac{I-1}{2}$ individuals with ideal points strictly smaller than x_m and, similarly, $\frac{I-1}{2}$ individuals with ideal points strictly larger than x_m; ultimately implying that the median agent is unique.[9]

We are now ready to prove the existence of a Condorcet winner in this setting. Consider a profile of individual preferences $\left(\succsim^1, \succsim^2, ..., \succsim^I\right)$ where, for every individual i, his preference relation $\succsim^i$ is single peaked with respect to the linear order $\geq$. In addition, let $m \in I$ be a median agent with ideal point x_m. Then, the median agent's ideal point x_m is a Condorcet winner, i.e., $x_m\widehat{F}\left(\succsim^1, \succsim^2, ..., \succsim^I\right) y$ for every alternative $y \in X$. (In words, the last sentence can be understood as that the ideal point of the median agent, x_m, cannot be defeated by majority voting by any other alternative y.)

We next prove that x_m is the Condorcet winner. In particular, take any alternative $y \in X$ and suppose that the ideal point of the median agent, x_m, satisfies $x_m > y$ (the argument is analogous if we assume that $y > x_m$). For x_m to be a Condorcet winner, we then need to show that alternative

[9]As an example, consider a setting with five voters with ideal points in $x_i \in [0, 1]$, where $x_1 < x_2 < ... < x_5$. Hence, the median voter is individual 3, with ideal point x_3, leaving the ideal points of voters 1 and 2 to the left of x_3, and the ideal points of voters 4and 5 to the right of x_3.

x_m defeats y, that is,

$$\#\left\{i \in I : x_m \succ^i y\right\} \geq \#\left\{i \in I : y \succ^i x_m\right\}$$

meaning that the number of individuals who strictly prefer x_m to y is larger than the number of individuals who strictly prefer y to x_m. Consider now the set of individuals $S \subset I$ with ideal points to the right-hand side of x_m, that is, $\{i \in I : x_i \geq x_m\}$. Then, their ideal points x_i must satisfy $x_i \geq x_m > y$ for every individual in this set, $i \in S$. Hence, by single-peaked preferences, every individual in this set must weakly prefer to vote for alternative x_m than y, i.e., $x_m \succsim^i y$ for every individual $i \in S$. Finally, because agent m is a median agent, the number of individuals with ideal points to the right-hand side of x_m, i.e., $\#S$, satisfies $\#S \geq \frac{I}{2}$, which entails

$$\#\left\{i \in I : x_m \succ^i y\right\} \geq \#S \geq \frac{I}{2} \geq \#\left(I \backslash S\right) \geq \#\left\{i \in I : y \succ^i x_m\right\}$$

Therefore, focusing on the terms in the extreme left- and right-hand side of the inequality, we obtain that the number of individuals who prefer to vote for x_m is larger than those voting for y. Since alternative y is arbitrary, x_m beats all other alternatives, ultimately making x_m the Condorcet winner; as required. (Q.E.D.)

As a consequence, imposing the assumption of single-peaked preferences guarantees the existence of a Condorcet winner. This is a positive result, as it helps us avoid the cyclicalities described in Condorcet paradox. In other words, the order in which pairs of alternatives are confronted in pairwise majority voting does not affect the final outcome. However, the presence of single-peaked preferences don't necessarily guarantee transitivity; as the next example illustrates.

Example 12.9. Intransitive social preferences. Consider a set of three alternatives $X = \{x, y, z\}$ and $I = 4$ individuals, with the following preference profiles

$$\begin{aligned}
x &\succ {}^1 y \succ^1 z \text{ for individual 1,} \\
z &\succ {}^2 y \succ^2 x \text{ for individual 2,} \\
x &\succ {}^3 z \succ^3 y \text{ for individual 3, and} \\
y &\succ {}^4 x \succ^4 z \text{ for individual 4}
\end{aligned}$$

We thus have that, when we run a pairwise majority voting between alternatives x and y, we obtain

$$\#\left\{i \in I : x \succ^i y\right\} = \#\left\{i \in I : y \succ^i x\right\} = 2$$

that is, the number of individuals preferring x over y (voters 1 and 2) coincides with the number preferring y over x (voters 2 and 4). Similarly, if we confront alternatives y and z in a pairwise majority voting, we find that

$$\#\left\{i \in I : z \succ^i y\right\} = \#\left\{i \in I : z \succ^i y\right\} = 2$$

since individuals 2 and 3 vote for alternative z, while individuals 1 and 4 vote for y. Therefore, we can conclude that alternative x is socially indifferent to y and, similarly, y is socially indifferent to z. More compactly, $z\widehat{F}\left(\succsim^1, \succsim^2, \succsim^3, \succsim^4\right) y$ and $y\widehat{F}\left(\succsim^1, \succsim^2, \succsim^3, \succsim^4\right) x$. For transitivity, we would need $z\widehat{F}\left(\succsim^1, \succsim^2, \succsim^3, \succsim^4\right) x$ to be satisfied. However, this result does not hold. Indeed, when alternatives z and x are presented to voters, the number of individuals preferring x to z (voters 1, 2 and 4) is larger than those preferring z to x (voter 2), that is,

$$\#\left\{i \in I : x \succ^i z\right\} = 3 \text{ and } \#\left\{i \in I : z \succ^i x\right\} = 1$$

thus implying $x\widehat{F}\left(\succsim^1, \succsim^2, \succsim^3, \succsim^4\right) z$, which violates transitivity in the swf.

However, a Condorcet winner exists. To show that, let us run a pairwise majority voting between all pairs of alternatives, in order to test if one alternative beats all others. First, in a pairwise majority voting between x and y, there is a tie since, as described above, two individuals vote for alternative x (voters 1 and 2) and the same number of individuals vote for y (voters 3 and 4). Second, in a pairwise majority voting between y and z, there is a tie since, as discussed above, two individuals vote for alternative z (voters 2 and 3) and the same number of individuals vote for y (voters 1 and 4). Finally, in a pairwise majority voting between z and x, alternative z wins as it receives votes from voters 1, 2 and 4. Hence, alternative z becomes the Condorcet winner.[10] $\square$

Guaranteeing transitivity in the swf. In order to guarantee that the social preference emerging from the swf is transitive, we need to impose two additional conditions: (1) The preference relation of every individual i must be strict (that is, we no longer allow individuals to be indifferent between some alternatives); and (2) the number of individuals I is odd. While the previous example, in which the social preference was intransitive, satisfied condition (1), i.e., individual preferences were strict, it did not satisfy condition (2), as we considered an even number of individuals (four). We next show that these two requirements help us obtain a transitive swf.

Result. *Consider an odd number of individuals I, each of them with strict single-peaked preferences relative to the linear order $\geq$. The social preference must be transitive.*

Proof. Consider a set of alternatives $X = \{x, y, z\}$, where

$$x\widehat{F}\left(\succsim^1, \succsim^2, ..., \succsim^I\right) y \quad \text{and} \quad y\widehat{F}\left(\succsim^1, \succsim^2, ..., \succsim^I\right) z.$$

That is, alternative x defeats y, and y defeats x. Since individual preferences are strict and I is odd, there must be one alternative in X that is not defeated by any other alternative in X. However, such alternative can be neither y (since y is defeated by x) nor z (which is defeated by y). Hence, such alternative has to be x. We can, thus, conclude that $x\widehat{F}\left(\succsim^1, \succsim^2, ..., \succsim^I\right) z$; as required to prove transitivity. (Q.E.D.)

[10] Note that a Condorcet winner, such as alternative z in this example, allows for z to either defeat all other alternatives, or produce a tie when z is confronted to some (but not all) alternatives. In other words, we cannot find another alternative that defeats z in a pairwise majority voting, thus declaring z as the Condorcet winner.

In summary, imposing the assumptions of strict, single-peaked preferences, with an odd number of individuals, helped us guarantee not only acyclic preferences (thus producing a Condorcet winner), but also a transitive social preference relation. While this result is positive, we must recognize that our above discussion only considered that the set of alternatives X was unidimensional, i.e., $X \subset \mathbb{R}$, a segment in the real line. For instance, the alternative being considered for a vote is a number, such as the percentage of government budget that a political candidate plans to spend on military, or on education. In many settings, however, candidates are evaluated on several dimensions, such as their plans for military spending, their experience, an even their looks, thus making their policy proposals multidimensional. A natural question is, then, whether we can still find a Condorcet winner in the social preference when we consider individual preferences that rank policy alternatives according to two or more dimensions. Bad news: a Condorcet winner may not exist in this setting, as End-of-Chapter Exercise #5 illustrates.

3.2 Second reaction - Social welfare function

While the first reaction of the literature to Arrow's impossibility theorem restricted the set of individual preferences being considered, the second reaction allows for the intensity of individual preferences to enter into social preferences. That is, rather than seeking an ordinal measure of social preferences, we obtain a cardinal measure. In particular, this approach uses a social welfare function

$$W\left(u^1(\cdot), u^2(\cdot), ..., u^I(\cdot)\right)$$

with its arguments being the utility levels of all individuals. Since individual preferences can be represented with different utility functions, we next analyze transformations on individual utility functions, and their effects on the social welfare function. In particular, we are interested in guaranteeing that the social ranking of two alternatives x and y, e.g., $W\left(u^1(x), u^2(x), ..., u^I(x)\right) \geq W\left(u^1(y), u^2(y), ..., u^I(y)\right)$, is unaffected if we transform individual utility functions. Note that, otherwise, it would be troublesome if, after applying a monotonic transformation on utility functions (which still represent the same preference relations), the social ranking between alternatives x and y changes. The next subsections investigate conditions on the monotonic transformations on utility functions that guarantee that the social ranking of alternatives is unaffected.

3.3 Utility-level invariance

First, consider a setting in which $u^1(x) > u^1(y)$ for individual 1, and $u^2(x) < u^2(y)$ for individual 2. In addition, assume that $u^1(y) > u^2(x)$, i.e., individual 1 is better off at his least-preferred alternative than individual 2 is. Then,

$$u^1(x) > u^1(y) > u^2(x)$$

where $u^2(y)$ must be larger than $u^2(x)$, but could rank above/below $u^1(x)$ or $u^1(y)$. Figure 12.3 provides an example of this utility ranking. First, note that individual 1 obtains a higher utility from alternative x than from y (see, respectively, points B and A in u^1); while individual 2 enjoys a higher utility from alternative x than from y (as illustrated, respectively, in points C and D in the figure). Second, point A is higher than C, thus implying that $u^1(y) > u^2(x)$, as required.

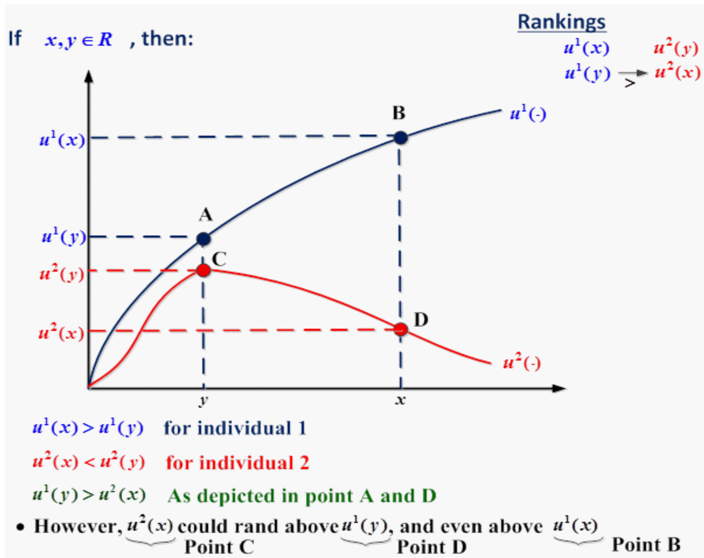

Figure 12.3. Utility-level invariance (motivation).

Assume that, in this context, society deems alternative y as socially preferred to x.[11] Now, consider strictly increasing transformations $\psi^1(\cdot)$ and $\psi^2(\cdot)$ producing the same individual ranking

$$\begin{aligned} v^1(x) &\equiv \psi^1(u^1(x)) > \psi^1(u^1(y)) \equiv v^1(y), \text{ and} \\ v^2(x) &\equiv \psi^2(u^2(x)) > \psi^2(u^2(y)) \equiv v^2(y) \end{aligned}$$

but altering the ranking across individuals, i.e., we started with $u^1(y) > u^2(x)$ but after applying these increasing transformation we obtain $v^1(y) < v^2(x)$. Hence, society would identify alternative x as socially preferred to y. However, this new social ranking is troublesome: indeed, we have not changed the individual ranking over alternatives, yet the social ranking changed! Figure 12.4 superimposes functions $v^1(\cdot)$ and $v^2(\cdot)$ on top of utility functions $u^1(\cdot)$ and $u^2(\cdot)$, showing that the individual ranking of alternatives did not change but the social ranking did.

[11] This social ranking of alternatives could be explained because society seeks to make its least well off individual as well off as possible, thus using the Rawlsian (or maxmin) criterion that we describe below. The assumption that alternative y is socially preferred to x is, however, without loss of generality; so a similar argument applies if x is socially preferred to y.

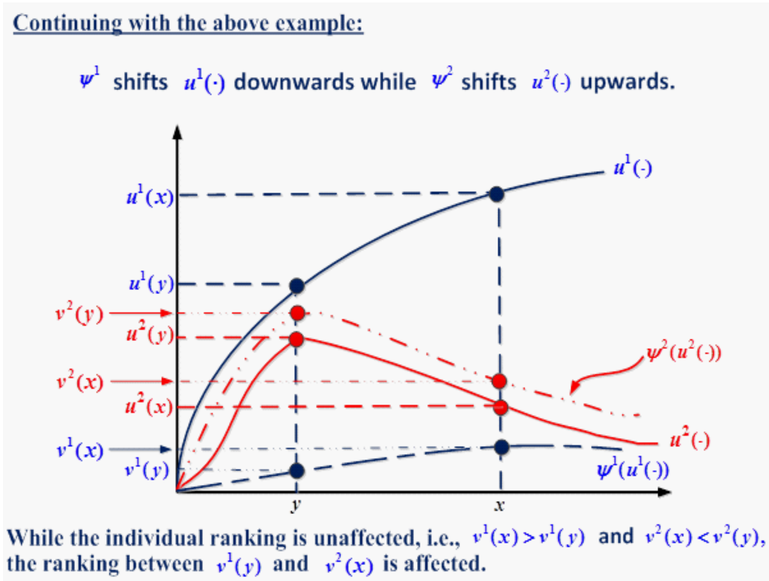

Figure 12.4. Utility-level invariance (transformations).

In order to avoid this possibility, we only need to avoid that the monotonic transformations on individual 1's utility function to be different from that on 2's utility function, i.e., $\psi^1 = \psi^2$. That is exactly what "utility-level invariance" guarantees; as we next define.

Utility-level invariance (Definition): A social welfare function $W(\cdot)$ is *utility-level invariant* if it is invariant to arbitrary, but common, strictly increasing transformations ψ applied to every individual's utility function.

This definition can alternatively be understood as follows. Consider a profile of individual preferences $\mathbf{u} \equiv \left(u^1(\cdot), u^2(\cdot), ..., u^I(\cdot)\right)$, where $\mathbf{u}(x) \equiv \left(u^1(x), u^2(x), ..., u^I(x)\right)$ and $\mathbf{u}(y) \equiv \left(u^1(y), u^2(y), ..., u^I(y)\right)$ denote the profile of individual utility levels from any two alternatives $x \neq y$. Therefore, utility-level invariance (ULI) implies that

$$\text{if } W(\mathbf{u}(x)) > W(\mathbf{u}(y)) \ \text{ then } \ W(\psi\left(\mathbf{u}(x)\right)) > W(\psi\left(\mathbf{u}(y)\right))$$

under a *common* strictly increasing transformation $\psi\left(\cdot\right)$, where

$$\psi\left(\mathbf{u}(x)\right) \equiv \left(\psi\left(u^1(x)\right), \psi\left(u^2(x)\right), ..., \psi\left(u^I(x)\right)\right)$$

and similarly for $\psi\left(\mathbf{u}(y)\right)$. In summary, if we apply the same monotonic transformation $\psi\left(\cdot\right)$ to all individuals' utility function, the social ranking over alternatives remains unaffected.

3.4 Utility-difference invariance

While the utility level that each individual obtains is important in making social choices, and thus the relevance of our above discussion of ULI, another type of information often used in making social choices is related with the utility gain/losses that every individual experiences when he moves from an alternative y to another alternative x. Let us now analyze this utility difference. In particular, consider that individual 1 enjoys a utility gain $u^1(x) - u^1(y)$ when moving from y to x, while individual 2 suffers a utility loss of $u^2(x) - u^2(y)$ from the same change in alternatives, i.e., $u^1(x) - u^1(y) > 0 > u^2(x) - u^2(y)$. A common comparison is, then, whether individual 1's gain of moving from y to x, $u^1(x) - u^1(y)$ is larger (in absolute value) than individual 2's loss, $u^2(y) - u^2(x)$, that is,

$$u^1(x) - u^1(y) > u^2(y) - u^2(x).$$

Figure 12.5a depicts a setting where individuals' utility functions are linear and this ranking holds.[12] For the swf to preserve this information, we need that monotonic transformations on $u^i(x)$ to be linear; as we next define.

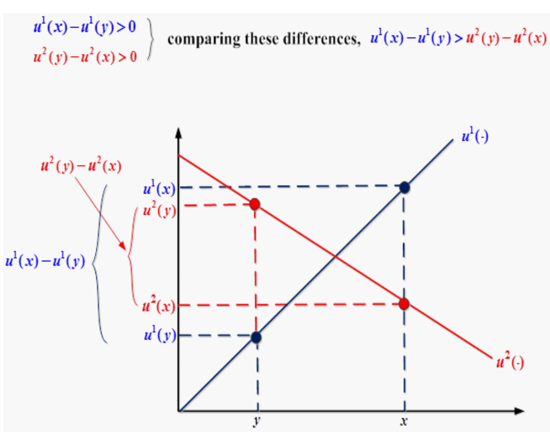

Figure 12.5a. UDI-motivation.

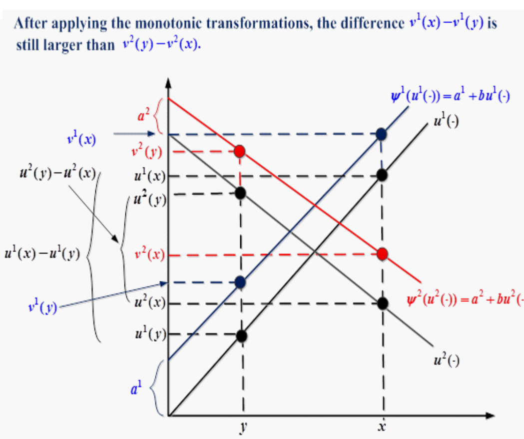

Figure 12.5b. UDI-transformations.

Utility-difference invariance (Definition). A social welfare function $W(\cdot)$ is *utility-difference invariant* if it is invariant to strictly increasing transformations of the following linear form

$$\psi^i\left(u^i(x)\right) = a^i + bu^i(x)$$

where $b > 0$ is common to all individuals.

While the slope of the linear transformation b coincides across all individuals, parameter a^i is

[12] Indeed, individual 1's gain of moving from alternative y to x (see arrow measuring gain $u^1(x) - u^1(y)$ in the vertical axis of figure 12.5a) offsets individual 2's loss (see $u^2(y) - u^2(x)$ in the vertical axis as well).

type-dependent (i.e., the upward or downward shift on the vertical intercept of $u^i(x)$ can vary across individuals), thus implying that the linear transformation can differ for each individual. Figure 12.5b depicts the initial utility function $u^i(x)$ and the monotonic transformation $v^i(x)$ for both individuals. The figure illustrates that, after applying a linear transformation (but not necessarily common) on both individuals' utility function, the initial ranking still holds, i.e., if $u^1(x) - u^1(y) > u^2(y) - u^2(x)$ then $v^1(x) - v^1(y) > v^2(y) - v^2(x)$ is still satisfied.[13]

3.5 Other properties of social welfare functions

We next list two more properties of some swfs.

Anonymity (Definition). Let $\mathbf{u}(x)$ and $\widetilde{\mathbf{u}}(x)$ be two utility vectors of alternative x, where $\widetilde{\mathbf{u}}(x)$ has been obtained from $\mathbf{u}(x)$ after a permutation of its elements. Then,

$$W(\mathbf{u}(x)) = W(\widetilde{\mathbf{u}}(x))$$

In words, the social ranking between alternatives x and y does not depend on the identity of the individuals involved, but only on the levels of utility that each alternative produces. As we describe in future sections, if $W(\mathbf{u}(x)) = \sum_i u^i(x)$, this property holds; but if $W(\mathbf{u}(x)) = \sum_i \beta^i u^i(x)$, where $\beta^i \in [0, 1]$ represents the weight that society assigns to the utility of individual i, anonymity does not necessarily hold.

Hammond Equity (Definition). Let $\mathbf{u}(x)$ and $\mathbf{u}(y)$ be the utility vectors of two distinct alternatives x and y, where $u^k(x) = u^k(y)$ for every individual k except for two individuals, i and j. If

$$u^i(x) < u^i(y) < u^j(y) < u^j(x)$$

then $W(\mathbf{u}(y)) \geq W(\mathbf{u}(x))$.

Intuitively, Hammond Equity (HE) says that society has a preference towards the alternative that produces the smallest variance in utilities across individuals (which corresponds to alternative y in the above definition). Figure 12.6 depicts, for a society with two individuals, the inequality which constitutes the premise for HE. Since this inequality implies that alternative y produces a utility pair that lies closer to the 45-degree line, alternative y is then associated to more equality than x. According to HE, the more equal alternative y generates a larger social welfare than x.

[13] As a remark, note that parameter a^i is allowed to be positive, as in figure 12.5b whereby u^i experiences an upward shift, or negative, where u^i would shift downwards.

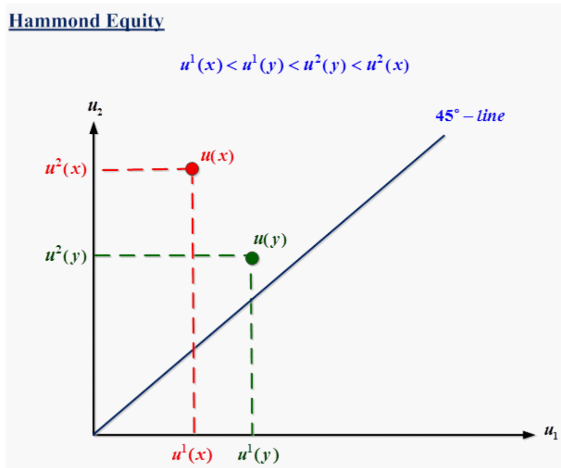

Figure 12.6. Hammond Equity.

While HE seems a reasonable property, it is often criticized because it focuses on equity, but potentially ignores the sum of utility levels that individuals obtain (i.e., the size of the pie). Consider, for instance, the following utility levels, which satisfy the premise in HE

$$u^i(x) = 1 < u^i(y) = 1.1 < u^j(y) = 1.2 < u^j(x) = 100.$$

According to HE, society will prefer alternative y than x, as it entails more equity. However, the sum of utility levels under alternative y is only $1.1 + 1.2 = 2.3$, being much smaller than the sum of utilities with alternative x, $1 + 100 = 101$.

4 Common social welfare functions

We next describe some well-known swf, such as the utilitarian and the Rawlsian. For each of them, we show that they can be characterized by a subset of the properties we just discussed: utility-level invariance (ULI), utility-difference invariance (UDI), anonymity (A), and Hammond Equity (HE).

4.1 Rawlsian social welfare function

This swf considers that the welfare that society obtains from an alternative x coincides with that of the utility level of the worst-off member in the society, that is, $W(x) = \min\left\{u^1(x), ..., u^I(x)\right\}$.

Result: *A strictly increasing and continuous swf W satisfies HE if and only if it can be represented with the Rawlsian form, $W(x) = \min\left\{u^1(x), ..., u^I(x)\right\}$.*

Proof: Note that the "if and only if" clause entails two lines of implication: (1) If W is continuous, strictly increasing, and satisfies HE, then W must be Rawlsian; and the opposite (2) If W is Rawlsian, then W is continuous, strictly increasing, and satisfies HE.

1st line of implication: Suppose that W is continuous, strictly increasing and satisfies HE. We then need to show that W takes the form

$$W(x) = \min\left\{u^1(x), ..., u^I(x)\right\}$$

That is, $W(x) \geq W(y)$ if and only if

$$\min\left\{u^1(x), ..., u^I(x)\right\} \geq \min\left\{u^1(y), ..., u^I(y)\right\}$$

Here is what we are planning to do: Similarly as indifference curves of a Leontieff preference relation in consumer theory, the social indifference curve of a Rawlsian swf must be a right angle (and all kinks are crossed by a ray from the origin).[14] We must then show that, starting from any arbitrary point **a** on the 45-degree line, all points in a horizontal line starting from the 45-degree line, and all points in a vertical line starting from the 45-degree line, yield the same social welfare as point **a**. Consider figure 12.7a, where we choose an arbitrary point **a** on the 45-degree line, and point **u** on the horizontal line extending from **a** to the right. As described above, we seek to show that the welfare in these two points coincide, $W(\mathbf{u}) = W(\mathbf{a})$.

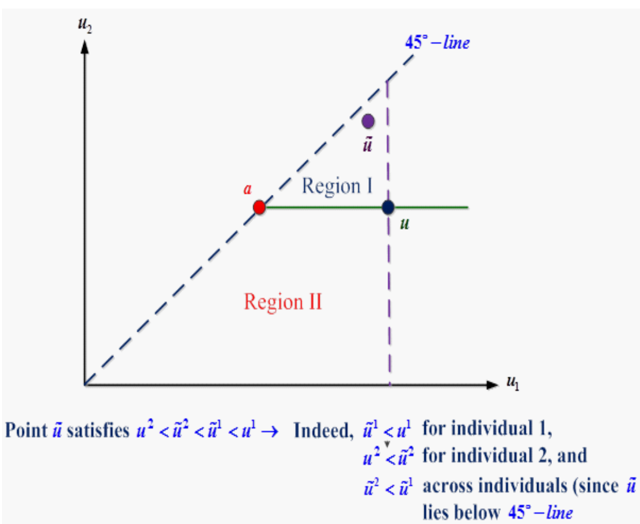

Figure 12.7a. Rawlsian indifference curve-I

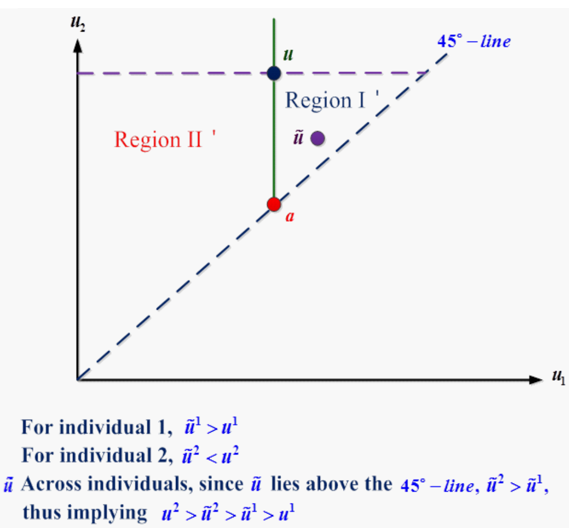

Figure 12.7b. Rawlsian indifference curve-II

Let us define regions I and II; as depicted in the figure. In Region I, consider an arbitrary point

[14] Recall that in consumer theory we represented Leontieff preferences with the utility function $u(x) = \min\{a_1x_1, a_2x_2, ..., a_nx_n\}$ in bundles with n components, i.e., $x = (x_1, x_2, ..., x_n)$, where parameter a_i are often assumed to be weakly positive. We depicted the indifference curves associated to this utility function as right angles, and used them extensively to represent preferences for complementary goods.

$\widetilde{\mathbf{u}}$ (see figure 12.7a). Since $\widetilde{\mathbf{u}}$ lies in Region I, it must satisfy

$$u^2 < \widetilde{u}^2 < \widetilde{u}_1 < u^1$$

Graphically, point $\widetilde{\mathbf{u}}$ is closer to the 45-degree line than $\mathbf{u}$ is, thus reducing the utility dispersion across individuals; as depicted in figure 12.7a. Hence, since HE holds, point $\widetilde{\mathbf{u}}$ is socially preferred to $\mathbf{u}$, i.e., $W(\widetilde{\mathbf{u}}) \geq W(\mathbf{u})$. Importantly, since $\widetilde{\mathbf{u}}$ is any arbitrary point in Region I, this argument can be extended to any point in region I, i.e., $W(I) \geq W(\mathbf{u})$. For the points in region II, since they all lie to the southwest of point $\mathbf{u}$, and W is strictly increasing, then we must have that $W(II) < W(\mathbf{u})$. Hence, $W(I) \geq W(\mathbf{u}) > W(II)$.

What about the points on the frontier between regions I and II, such as point $\mathbf{a}$? By continuity of the swf W, since $W(I) \geq W(\mathbf{u})$ in region I and $W(\mathbf{u}) > W(II)$ in region II, we must have that $W(\mathbf{u}) = W(\mathbf{a})$, as we wished to show.

We can extend the same argument, but now starting from a vertical ray that extends from $\mathbf{a}$ upwards (rather than a horizontal ray); as depicted in figure 12.7b.[15] Because W is strictly increasing, no other points can yield the same social welfare than $\mathbf{a}$ other than the two rays we just examined. That is, the union of the two rays provides us with the social indifference curve for W. In conclusion, W has the same indifference map as function $\min\left\{u^1(x), ..., u^I(x)\right\}$; as figure 12.8 illustrates.

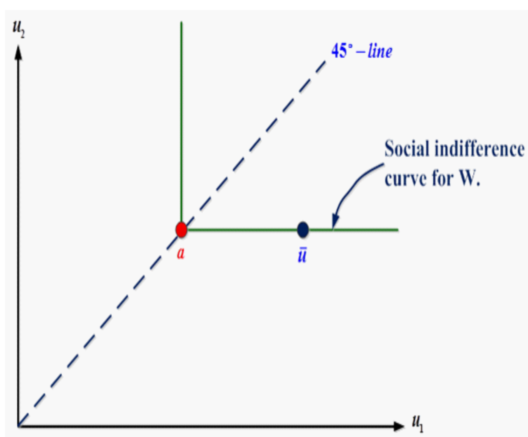

Figure 12.8. Rawlsian indifference curve.

2nd line of implication: In this case, we need to show that if the swf is represented by $W(x) = \min\left\{u^1(x), ..., u^I(x)\right\}$ then HE holds. The premise of HE considers a society where $u^k(x) = u^k(y)$ for every individual k except for two individuals i and j, and assumes that

$$u^i(x) < u^i(y) < u^j(y) < u^j(x)$$

[15] That is, we have just examined the welfare at points below the 45-degree line, but a similar argument applies for points above the 45-degree line.

Then, for HE to hold, we need that society prefers the alternative with the smaller utility dispersion, i.e., $W(\mathbf{u}(y)) \geq W(\mathbf{u}(x))$. Since the premise of HE specifies a complete order for $u^i(x), u^i(y), u^j(y)$ and $u^j(x)$, the only remaining question is the position of $u^k(x) = u^k(y)$ relative to the ranking $u^i(x) < u^i(y) < u^j(y) < u^j(x)$. In particular, $u^k(x) = u^k(y)$ can lie on either of the regions depicted in figure 12.9, and examined in the subsequent discussion.

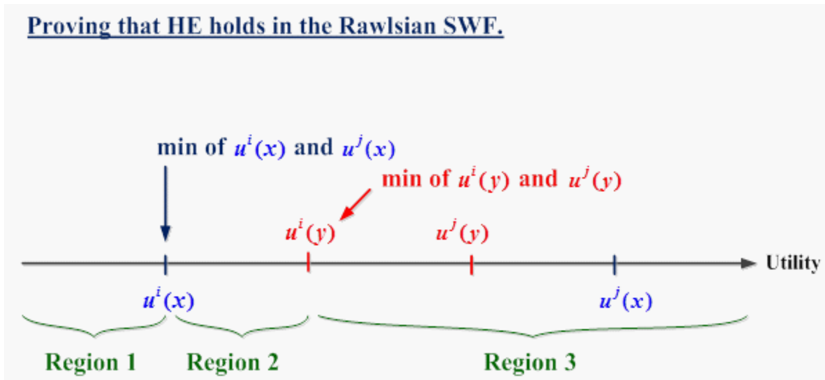

Figure 12.9. HE holds in Rawlsian swfs.

Region 1, where $u^k(x) = u^k(y) < u^i(x)$. In this case, $W(\mathbf{u}(x)) = u^k(x)$ and $W(\mathbf{u}(y)) = u^k(y)$, implying that society is indifferent between alternatives y and x, i.e., $W(\mathbf{u}(y)) = W(\mathbf{u}(x))$, which is allowed according to the HE property (recall that we seek to show $W(\mathbf{u}(y)) \geq W(\mathbf{u}(x))$).

Region 2, where $u^i(x) < u^k(x) = u^k(y) < u^i(y)$. In this case, $W(\mathbf{u}(x)) = u^i(x)$ and $W(\mathbf{u}(y)) = u^k(y)$, which entails that society strictly prefers alternative y to x, i.e., $W(\mathbf{u}(y)) > W(\mathbf{u}(x))$, thus satisfying the HE property. Intuitively, alternative y yields a smaller utility dispersion than x does.

Region 3, where $u^i(y) < u^k(x) = u^k(y)$. In this setting, $W(\mathbf{u}(x)) = u^i(x)$ and $W(\mathbf{u}(y)) = u^i(y)$. As a consequence, society strictly prefers alternative y to x, i.e., $W(\mathbf{u}(y)) > W(\mathbf{u}(x))$, thus satisfying the HE property. Similarly as in Region 2, alternative y yields a smaller utility dispersion than x does.

Hence, we can conclude that, regardless of where $u^k(x) = u^k(y)$ lies, the HE property holds when the swf is represented in the Rawlsian form. (Q.E.D.)

Corollary: *The Rawlsian swf* $W(x) = \min\left\{u^1(x), ..., u^I(x)\right\}$ *satisfies anonymity, and is utility-level invariant.*

Proof. *Anonymity.* Take a utility vector $u^1(x), ..., u^I(x)$, where

$$\min\left\{u^1(x), ..., u^I(x)\right\} = u^k(x)$$

i.e., individual k reaches the lowest utility level under alternative x. Now, perform a permutation on the identities of these individuals, and apply the min operator on their vector of utility levels again. The min is still $u^k(x)$.

ULI. Let us now show that the Rawlsian swf also satisfies utility-level invariance (ULI). For simplicity, let us first define what we need to show. In particular, consider a strictly increasing transformation common to all individuals, $\psi : \mathbb{R} \to \mathbb{R}$. If the social welfare of alternative x is larger than that of y, $W(\mathbf{u}(x)) \geq W(\mathbf{u}(y))$, then the social ranking is preserved after applying the common strictly increasing transformation ψ to all individuals' utility function. That is,

$$W\left(\psi\left(u^1(x)\right), ..., \psi\left(u^I(x)\right)\right) \geq \psi\left(W\left(u^1(x), ..., u^I(x)\right)\right)$$

Let us now show that ULI holds. Specifically, let us apply a strictly increasing transformation common to all individuals $\psi : \mathbb{R} \to \mathbb{R}$,

$$W\left(\psi\left(u^1(x)\right), ..., \psi\left(u^I(x)\right)\right) = \psi\left(W\left(u^1(x), ..., u^I(x)\right)\right)$$

where the transformation ψ is factored out the swf W since it is common to all individuals.[16] Therefore,

$$W\left(\psi\left(u^1(x)\right), ..., \psi\left(u^I(x)\right)\right) \geq W\left(\psi\left(u^1(y)\right), ..., \psi\left(u^1(y)\right)\right)$$

can be rewritten as follows by factoring out the common transformation ψ

$$\psi\left(W\left(u^1(x), ..., u^I(x)\right)\right) \geq \psi\left(W\left(u^1(y), ..., u^I(y)\right)\right)$$

which is equivalent to

$$W\left(u^1(x), ..., u^I(x)\right) \geq W\left(u^1(y), ..., u^I(y)\right);$$

as required by utility-level invariance. (Q.E.D.)

4.2 The Utilitarian SWF

This swf assigns an equal weight to the utility level of each individual, and it is probably the most commonly used swf in economics.

$$W(x) = u^1(x) + u^2(x) + ... + u^I(x) = \sum_{i=1}^{I} u^i(x)$$

Hence, in a society with two individuals, $W = u^1(x) + u^2(x)$, or solving for u^2, $u^2(x) = W - u^1(x)$, thus being represented by a straight line with slope -1 in the $(u_1, u_2)-$quadrant. We next show that the utilitarian swf satisfies two of the properties presented above, A and UDI, if the swf is strictly increasing and continuous in u_i.

Theorem. *The utilitarian swf* $W(x) = \sum_{i=1}^{I} u^i(x)$ *satisfies A and UDI.*

Proof*: Anonymity.*When W takes the utilitarian form, A holds since the utility level of each

[16] If, for instance, the common transformation is linear, i.e., $\psi\left(u^i(x)\right) = a + bu^i(x)$ where $a, b > 0$, then the expression we just found becomes $\psi\left(W\left(u^1(x), ..., u^I(x)\right)\right) = a + b \min\left\{u^1(x), ..., u^I(x)\right\}$.

individual receives the same weight. That is, a permutation on the identities of individuals will not alter the social ranking of alternatives; as required by anonymity.

UDI. In addition, when W takes the utilitarian form, UDI holds as well, since

$$\text{if } \; W(x) = u^1(x) + u^2(x) \geq u^1(y) + u^2(y) = W(y),$$

then a monotonic transformation $a^i + bu^i(x)$ to $u^i(x)$ yields

$$\begin{aligned} & \left(a^1 + bu^1(x)\right) + \left(a^2 + bu^2(x)\right) \\ \geq \; & \left(a^1 + bu^1(y)\right) + \left(a^2 + bu^2(y)\right) \end{aligned}$$

which collapses to

$$b\left[u^1(x) + u^2(x)\right] \geq b\left[u^1(y) + u^2(y)\right]$$

which is satisfied since $u^1(x) + u^2(x) \geq u^1(y) + u^2(y)$, and $b > 0$ by definition. Hence, UDI holds for the utilitarian swf. (Q.E.D.)

We next show the opposite line of implication: that a swf satisfying A and UDI can only be represented with the utilitarian form.

Theorem. *A strictly increasing and continuous swf satisfying A and UDI can only be represented with the utilitarian form.*

Proof*:* For illustration purposes, consider figure 12.10a, which depicts the $(u_1, u_2)-$quadrant. First, take a point $\overline{\mathbf{u}}$ on the 45-degree line, where $\overline{u}_1 = \overline{u}_2$; and let γ represent the sum of the two components of point $\overline{\mathbf{u}}$, i.e., $\overline{u}^1 + \overline{u}^2 \equiv \gamma$.

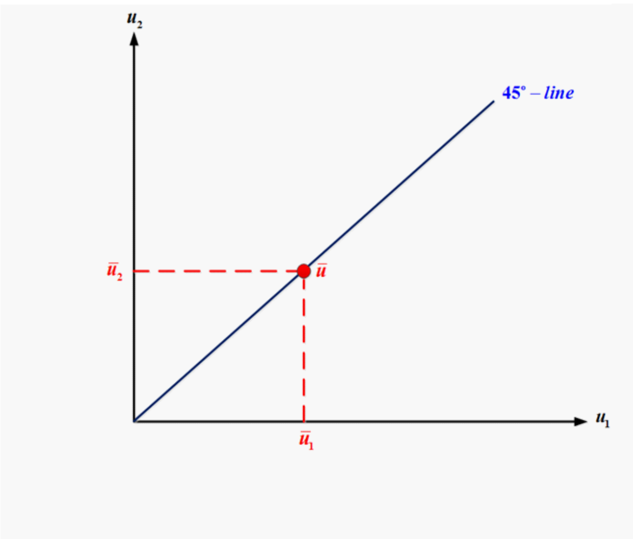

Figure 12.10a. Utilitarian swf-I

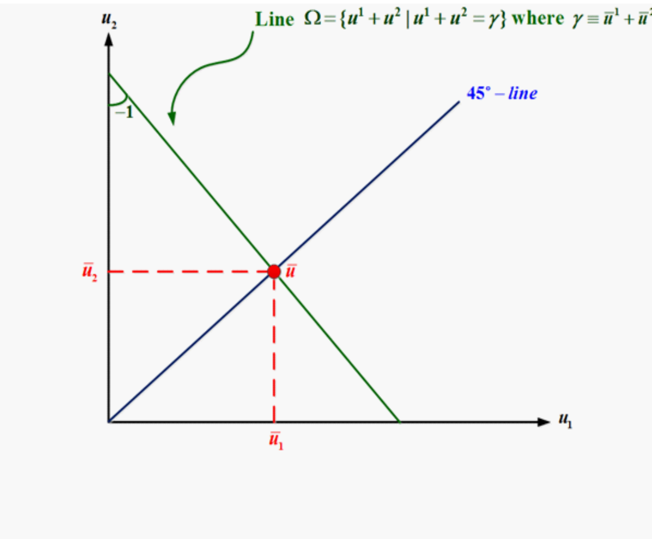

Figure 12.10b. Utilitarian swf-II

Now, let us consider the set of points for which the sum of their two components, $u^1 + u^2$, yields exactly γ. That is,

$$\Omega = \left\{u^1 + u^2 \mid u^1 + u^2 = \gamma\right\}$$

Solving for u_2, yields $u^2 = \gamma - u^2$, implying that set Ω can be depicted by the line originating at γ, decreasing in u^1 with a slope of -1, and that crosses the 45-degree line at exactly point $\overline{\mathbf{u}}$; as depicted in figure 12.10b.

Here is what we are planning to do: Since the social indifference curve of a utilitarian swf must be linear, i.e., $u^2 = W - u^1$, we seek to show that all points in line Ω yield the same social welfare as point $\overline{\mathbf{u}}$ does.

$$W(\Omega) = W(\overline{\mathbf{u}}).$$

Therefore, let us start choosing any point in line Ω, distinct from $\overline{\mathbf{u}}$, such as point $\widetilde{\mathbf{u}}$ in figure 12.11. Without loss of generality, point $\widetilde{\mathbf{u}}$ is located below the 45-degree line, but a similar argument applies if $\widetilde{\mathbf{u}}$ lies above the 45-degree line. Hence, point $\widetilde{\mathbf{u}}^T$ is just a permutation of $\widetilde{\mathbf{u}}$, i.e., if $\widetilde{\mathbf{u}} = (\widetilde{u}^1, \widetilde{u}^2)$ point $\widetilde{\mathbf{u}}^T$ becomes $\widetilde{\mathbf{u}}^T = (\widetilde{u}^2, \widetilde{u}^1)$, thus lying above the 45-degree line.

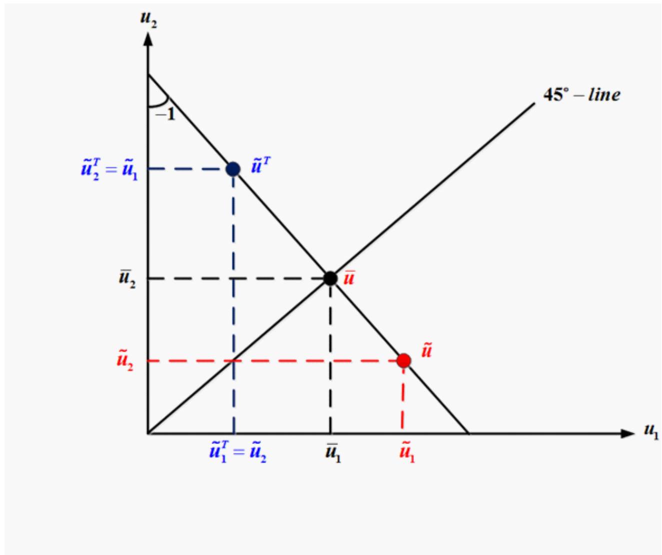

Figure 12.11. Utilitarian swf-III.

By condition A, points $\widetilde{\mathbf{u}}$ and $\widetilde{\mathbf{u}}^T$ must be ranked the same way relative to $\overline{\mathbf{u}}$.[17] That is, if $W(\overline{\mathbf{u}}) \geq W(\widetilde{\mathbf{u}})$, then such social ranking is maintained at point $\widetilde{\mathbf{u}}^T$, i.e., $W(\overline{\mathbf{u}}) \geq W(\widetilde{\mathbf{u}}^T)$. Likewise, if $W(\widetilde{\mathbf{u}}) \geq W(\overline{\mathbf{u}})$, then such social ranking also holds at point $\widetilde{\mathbf{u}}^T$, i.e., $W(\widetilde{\mathbf{u}}^T) \geq W(\overline{\mathbf{u}})$.

Suppose that $W(\overline{\mathbf{u}}) > W(\widetilde{\mathbf{u}})$. Under UDI, this social ranking must be unaffected by linear transformations of the form $\psi^i\left(u^i(\cdot)\right) = a^i + bu^i(\cdot)$. In particular, let $b = 1$ and $a^i = \overline{u}^i - \widetilde{u}^i$, yielding

$$\psi^i\left(u^i(x)\right) = \underbrace{\overline{u}^i(x) - \widetilde{u}^i(x)}_{a^i} + u^i(x)$$

Applying this transformation to $\widetilde{\mathbf{u}}$ yields $\psi^i\left(\widetilde{u}^i(x)\right) = \overline{u}^i(x) - \widetilde{u}^i(x) + \widetilde{u}^i(x) = \overline{u}^i(x)$, that is, $\left(\psi^1\left(\widetilde{u}^1\right), \psi^2\left(\widetilde{u}^2\right)\right) = \overline{\mathbf{u}}$. Graphically, this linear transformation applied to point $\widetilde{\mathbf{u}}$ (which is located away from the 45-degree line) moves us to point $\overline{\mathbf{u}}$, located on the 45-degree line. In addition, if we apply this transformation to point $\overline{\mathbf{u}}$, we obtain

$$\psi^i\left(\overline{u}^i(x)\right) = \overline{u}^i(x) - \widetilde{u}^i(x) + \overline{u}^i(x) = 2\overline{u}^i(x) - \widetilde{u}^i(x)$$

However, since point $\overline{\mathbf{u}}$ lies on the 45-degree line, $2\overline{u}^i(x) = \widetilde{u}^i(x) + \widetilde{u}^j(x)$, which helps us rewrite the result from the transformation on $\overline{\mathbf{u}}$ as

$$\psi^i\left(\overline{u}^i(x)\right) = 2\overline{u}^i(x) - \widetilde{u}^i(x) = \underbrace{\left[\widetilde{u}^i(x) + \widetilde{u}^j(x)\right]}_{2\overline{u}^i(x)} - \widetilde{u}^i(x) = \widetilde{u}^j(x)$$

[17] Indeed, we only changed the identities of the individuals receiving utility levels u^1 and u^2, but did not alter the size of u^1 nor that of u^2. Importantly, note that A does not imply that the society is indifferent between points $\widetilde{\mathbf{u}}$ and $\widetilde{\mathbf{u}}^T$. (We actually plan to show this point for the utilitarian swf, but we have not shown that yet!).

That is, $\left(\psi^1\left(\overline{u}^1\right), \psi^2\left(\overline{u}^2\right)\right) = \widetilde{\mathbf{u}}^T$. Therefore, point $\widetilde{\mathbf{u}}$ is transformed into $\overline{\mathbf{u}}$, and point $\overline{\mathbf{u}}$ is transformed into $\widetilde{\mathbf{u}}^T$.

In summary, we started assuming that $W(\overline{\mathbf{u}}) > W(\widetilde{\mathbf{u}})$ holds, and showed that, by UDI, we must have that $W(\widetilde{\mathbf{u}}^T) > W(\overline{\mathbf{u}})$ also holds. Hence, $W(\widetilde{\mathbf{u}}^T) > W(\overline{\mathbf{u}})$ and $W(\overline{\mathbf{u}}) > W(\widetilde{\mathbf{u}})$, which implies $W(\widetilde{\mathbf{u}}^T) > W(\widetilde{\mathbf{u}})$, thus violating A. Therefore, our initial assumption, i.e., $W(\overline{\mathbf{u}}) > W(\widetilde{\mathbf{u}})$, cannot hold. A similar argument applies if we, instead, start our proof assuming that $W(\overline{\mathbf{u}}) < W(\widetilde{\mathbf{u}})$. We can therefore conclude that $W(\overline{\mathbf{u}}) = W(\widetilde{\mathbf{u}})$ must hold which, together with A, implies that

$$W(\overline{\mathbf{u}}) = W(\widetilde{\mathbf{u}}) = W(\widetilde{\mathbf{u}}^T)$$

Finally, since point $\widetilde{\mathbf{u}}$ was an arbitrary point in the line Ω, we can claim that the social welfare at point $\overline{\mathbf{u}}$ is the same as that of *any* point along the line Ω, i.e., $W(\overline{\mathbf{u}}) = W(\Omega)$. (Q.E.D.)

We can next expand our previous results to the "generalized utilitarian" swf of the form

$$W(x) = \sum_{i=1}^{I} \alpha^i u^i(x)$$

where $\alpha^i > 0$ represents the weight society assigns to individual i. As a practice, show that the generalized utilitarian swf does not satisfy A.

Example 12.10. Generalized utilitarian swf. For the case of two individuals, the generalized utilitarian swf becomes $W = \alpha^1 u^1 + \alpha^2 u^2$ which, solving for u^2, yields a social indifference curve of

$$u^2 = \frac{W}{\alpha^2} - \frac{\alpha^1}{\alpha^2} u^1,$$

thus being still a straight, negatively sloped line, but the slope is now $-\frac{\alpha^1}{\alpha^2}$. Figure 12.12 depicts three social indifference curves, depending on the value of the $\frac{\alpha^1}{\alpha^2}$ ratio. In order to interpret this ratio, consider a society seeking to increase individual 1's utility in one more unit while still maintaining the welfare level unaffected (graphically represented by a right-ward movement along the social indifference curve). When society assigns a larger weight to the utility of individual 1 than 2, $\alpha^1 > \alpha^2$, the ratio becomes larger than 1 in absolute value, i.e., $-\frac{\alpha^1}{\alpha^2} > -1$, implying that the amount of u^2 that society is willing to give up (in order to increase u^1 by one unit) is relatively large. The opposite argument applies when weights satisfy $\alpha^1 < \alpha^2$, as the ratio now satisfies $-\frac{\alpha^1}{\alpha^2} < -1$, entailing that society is willing to give up a small utility u^2 from individual 2 in order to increase u^1 by one unit. Finally, note that the utilitarian swf can be understood as a special case of the generalized utilitarian when weights coincide, i.e., $\alpha^1 = \alpha^2$. $\square$

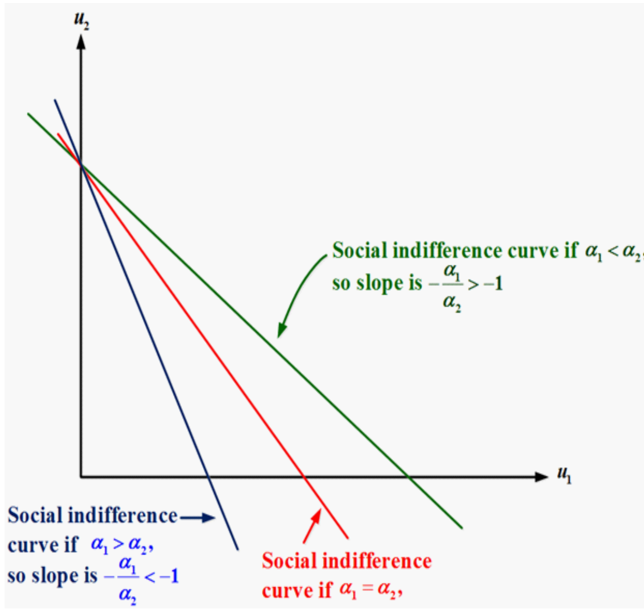

Figure 12.12. Social indifference curves for a generalized utilitarian swf.

4.3 Flexible form SWF

In the analysis of certain policies, e.g., moving from alternative x to y, we might be interested in the percentage change in utility that each individual experiences, $\frac{u^i(x)-u^i(y)}{u^i(x)}$, and whether such a percentage is larger for individual i than for j. That is,

$$\frac{u^i(x) - u^i(y)}{u^i(x)} > \frac{u^j(x) - u^j(y)}{u^j(x)}$$

If we seek to maintain the ranking of percentage changes across individuals invariant to monotonic transformations on the utility functions we need monotonic transformations to be *linear* and *common* among individuals, $\psi(u^i) = bu^i$, where $b > 0$ for all i. Indeed, applying this transformation on the above inequality yields

$$\frac{bu^i(x) - bu^i(y)}{bu^i(x)} > \frac{bu^j(x) - bu^j(y)}{bu^j(x)}$$

which, factoring b out, reduces to

$$\frac{u^i(x) - u^i(y)}{u^i(x)} > \frac{u^j(x) - u^j(y)}{u^j(x)}$$

Utility-percentage invariance (Definition). A swf is utility-percentage invariant if it is invariant to arbitrary, but linear and common, strictly increasing transformations of the form $\psi(u^i) = bu^i$, where $b > 0$ for every individual i.

As a consequence, if a swf satisfies utility-percentage invariant (UPI), it must also satisfy:

- Utility-level invariance (ULI), since for that we need that the strictly increasing transformations are *common* across individuals, i.e., $\psi^i(\cdot) = \psi^j(\cdot)$ for any two individuals $i \neq j$; and

- Utility-difference invariance (UDI), since for that we need that the strictly increasing transformation for each individual to be *linear*, i.e., $\psi^i(u^i) = a^i + bu^i$ where $b > 0$.

Hence, UPI can be understood as a special case of ULI and of UDI. Interestingly, UPI allows for a large class of swfs, whereby the Rawlsian and utilitarian described in previous sections are just special cases. Let us start demonstrating that UPI yields homothetic social indifference curves.

Result. *A strictly increasing swf satisfying UPI must be homothetic.*

Proof: Consider figure 12.13. Let us start by choosing an arbitrary point $\overline{\mathbf{u}}$. Since the swf W is strictly increasing, the social indifference curve must be negatively sloped. Next, depict a ray from the origin that crosses point $\overline{\mathbf{u}}$, as ray OA in the figure, and choose a point on this ray, e.g., $b\overline{\mathbf{u}}$ where $b > 0$.

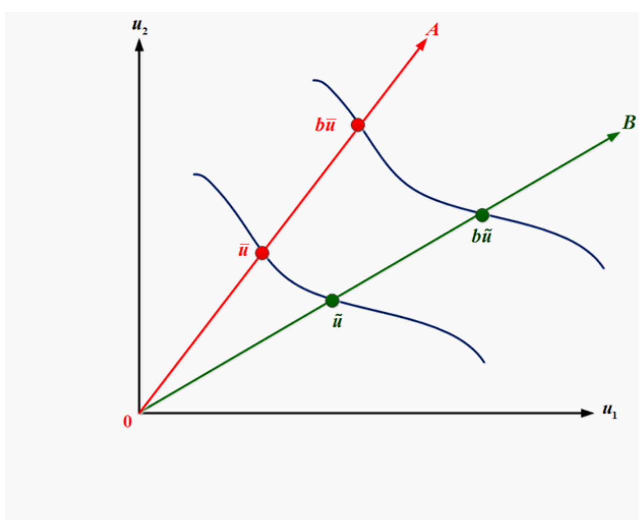

Figure 12.13. Homothetic social preferences-I

Using a similar argument, select another point, $\widetilde{\mathbf{u}}$, lying on the same social indifference curve, i.e., $W(\widetilde{\mathbf{u}}) = W(\overline{\mathbf{u}})$, depict a ray from origin passing through point $\widetilde{\mathbf{u}}$, such as ray OB in figure 12.13, and choose a point through this ray, i.e., $b\widetilde{\mathbf{u}}$, where $b > 0$. We have then multiplied utility vectors $\overline{\mathbf{u}}$ and $\widetilde{\mathbf{u}}$ by the same factor b, i.e., we applied a common transformation $\psi(u^i) = bu^i$ to all individuals' utilities. Hence, by UPI, if $W(\widetilde{\mathbf{u}}) = W(\overline{\mathbf{u}})$ then $W(b\widetilde{\mathbf{u}}) = W(b\overline{\mathbf{u}})$, entailing that, graphically, points $b\overline{\mathbf{u}}$ and $b\widetilde{\mathbf{u}}$ must lie on the same social indifference curve.

Recall that we need to show homotheticity of the social indifference curve, which entails that:

- The slope at point $\overline{\mathbf{u}}$ must coincide with that in point $b\overline{\mathbf{u}}$, and
- The slope at point $\widetilde{\mathbf{u}}$ must coincide with that in point $b\widetilde{\mathbf{u}}$.

Figure 12.14 depicts a chord connecting points $\overline{\mathbf{u}}$ and $\widetilde{\mathbf{u}}$, chord CC. If these points are sufficiently close to each other, the slope of chord CC approximates the slope of the social indifference curve at point $\overline{\mathbf{u}}$. By a similar argument, the chord that connects points $b\overline{\mathbf{u}}$ and $b\widetilde{\mathbf{u}}$ (namely, chord DD in figure 12.14) approximates the slope of the social indifference curve at point $b\overline{\mathbf{u}}$. Furthermore, since points $b\overline{\mathbf{u}}$ and $b\widetilde{\mathbf{u}}$ have both been increased by the same factor b, the slope of chord CC must coincide with that of DD.

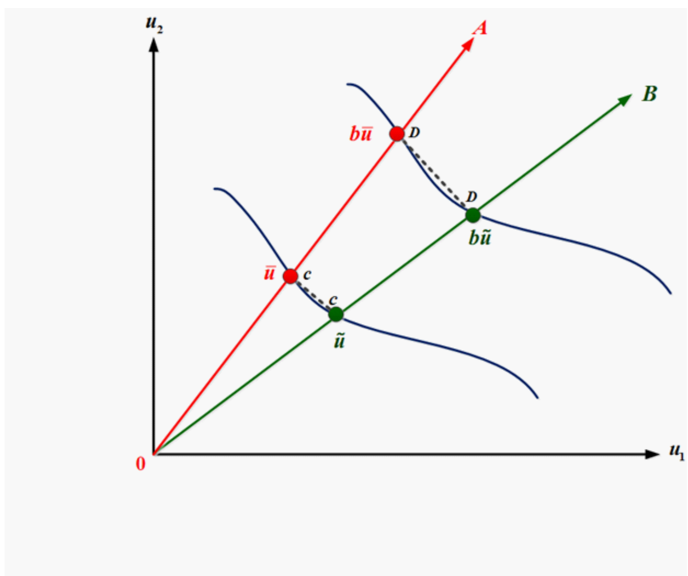

Figure 12.14. Homothetic social preferences-II

In addition, if we choose a point $\widetilde{\mathbf{u}}$ closer and closer to $\overline{\mathbf{u}}$, the slope of chords CC and DD still coincide, but their slopes better approximate those of the social indifference curve through each point. In the limit, the slope of the social indifference curve at point $\overline{\mathbf{u}}$ coincides with that at point $b\overline{\mathbf{u}}$, proving homotheticity. (Q.E.D.)

As a summary, we can encompass all previous functional forms of swfs into the following swf, which exhibits a familiar CES form:

$$W(x) = \sum_{i=1}^{I} \left[\left(u^i(x) \right)^{\rho} \right]^{\frac{1}{\rho}}$$

where $0 \neq \rho < 1$. Hence, the constant elasticity of social substitution between the utility of any two individuals, σ, can be expressed as $\sigma = \frac{1}{1-\rho}$. This swf satisfies three properties mentioned above: A, WP, and quasiconcavity. [Show that as a practice.] In addition, it satisfies a common assumption in consumer theory: strong separability. Formally, the MRS_{u^i,u^j} only depends on the utility of individuals i and j, u^i and u^j, but does not depend on the utility from any other individual $k \neq i, j$. Indeed, the MRS_{u^i,u^j} of this CES swf is

$$MRS_{u^i,u^j} = -\left(\frac{u^i}{u^j}\right)^{\rho-1}$$

which is independent on u^k. For illustration purposes, figure 12.15 depicts three social indifference curves of a CES swf: (1) $\rho \to 1$, corresponding to linear social indifference curves, i.e., utilitarian swf; (2) $-\infty < \rho < 1$, curvy social indifference curves, resembling Cobb-Douglas indifference curves in consumer theory; and (3) $\rho \to -\infty$, corresponding to right-angled social indifference curves, i.e., Rawlsian swf.

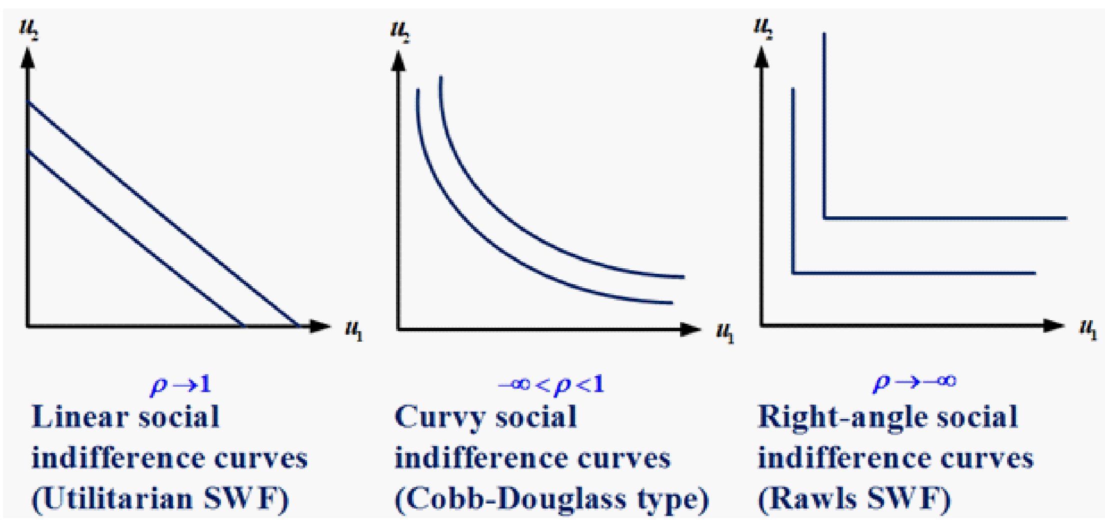

Figure 12.15. CES swf (three cases).

5 Theories of justice

In previous sections, we described the properties of different swfs, but did not discuss how societies choose one swf over another. As we next discuss, the literature has mainly considered two approaches: Harsanyi's and Rawls' approach. Both approaches assume that individuals do not yet know which position they will occupy in society. That is, before being borned, every individual i cannot perfectly anticipate his utility level u^i, thus not knowing whether he will be one of the individuals with the highest or lowest utility level in society. This assumption by both approaches is commonly referred to as individuals' "veil of ignorance," after Rawls (1971). However, these approaches differ in how every individual i assigns probabilities to each of the possible positions he could occupy, as we discuss below. While both approaches seem at first glance incompatible, we show below that they can actually be modeled as special cases of a more general ("unified") approach.

Harsanyi's approach. Harsanyi claims that individuals assign an *equal probability* to the prospect of being in any possible position in society, which is often referred to as the "principle of insufficient reason." Hence, if there are I people in society, there is a probability $\frac{1}{I}$ that individual i will end up in the position of any of these I individuals, yielding a utility $u^j(x)$, thus implying that i's expected utility is

$$\sum_{i=1}^{I} \frac{1}{I} u^i(x)$$

Therefore, when society chooses between two alternatives x and y, alternative x is socially preferred if

$$\sum_{i=1}^{I} \frac{1}{I} u^i(x) > \sum_{i=1}^{I} \frac{1}{I} u^i(y) \Longleftrightarrow \sum_{i=1}^{I} u^i(x) > \sum_{i=1}^{I} u^i(x)$$

which exactly coincides with the condition provided by the utilitarian swf. This explains why Harsanyi's approach is often used to support the use of utilitarian swfs.

Rawls' approach. In contrast, Rawls claims that individuals have no empirical basis for assigning probabilities to each position, whether equal or unequal probabilities to each position.[18] Assuming people are risk averse, he argues that individuals would order alternatives according to which one provides him with the highest utility in case he ends up as society's worst-off member. Thus, alternative x is socially preferred to y if and only if

$$\min\left\{u^1(x), ..., u^I(x)\right\} \geq \min\left\{u^1(y), ..., u^I(y)\right\}$$

which is a purely maximin criterion, i.e., society should choose the alternative that maximizes the utility of the worst-off individual.

Unification of both approaches. Let us now show that Rawls' and Harsanyi's approaches can be modeled as special cases of a more general setting. First, take a utility function $u^i(x)$ for individual i. Since the underlying preferences of this individual can also be represented by monotonic transformations of $u^i(x)$, we can apply the following concave transformation to $u^i(x)$

$$v^i(x) \equiv -u^i(x)^{-a}, \text{ where } a > 0$$

where $v^i(x)$ can be understood as the vNM utility function of this individual, with parameter a

[18] That is, Rawls viewed every individual i's original position in a setting of complete ignorance.

capturing his degree of risk aversion; as Figure 12.16 depicts.[19]

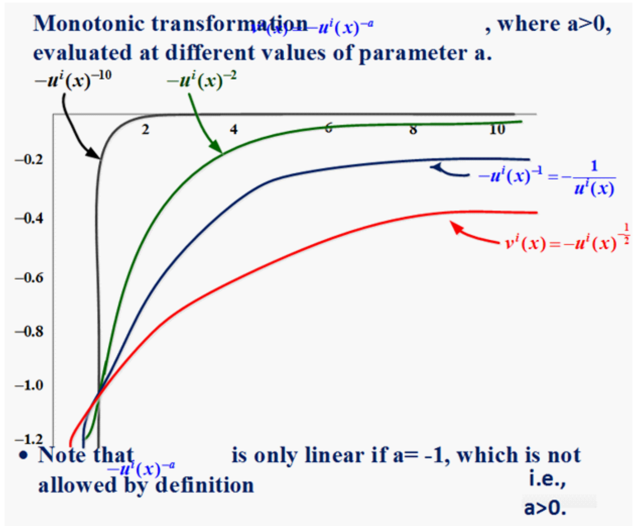

Figure 12.16. Concave transformation of $u^i(x)$.

Using the Harsanyi's approach on this monotonic transformation, yields a social welfare function

$$W = \sum_{i=1}^{I} v^i(x) \equiv -\sum_{i=1}^{I} -u^i(x)^{-a}$$

Importantly, the social ranking of alternatives provided by this swf must coincide with that of its monotonic transformation W^*, where $W^* \equiv (-W)^{-\frac{1}{a}}$, entailing

$$W^* \equiv (-W)^{-\frac{1}{a}} = \left(-\sum_{i=1}^{I} -u^i(x)^{-a}\right)^{-\frac{1}{a}} = \left(\sum_{i=1}^{I} -u^i(x)^{-a}\right)^{-\frac{1}{a}}$$

Hence, if parameter $-a = \rho$, we can express the swf W^* as

$$W^* = \left(\sum_{i=1}^{I} -u^i(x)^{\rho}\right)^{\frac{1}{\rho}}$$

which coincides with the CES swf described in previous sections. Therefore, when $\rho \to -\infty$, the parameter of risk aversion a becomes $a \to \infty$, and the above swf approaches the maximin criterion by Rawls as a limiting case. In addition, the Rawlsian criterion becomes a special case of Harsanyi's approach when individuals become infinitely risk averse. Finally, when $-\infty < \rho < 1$,

[19] Note that function $v^i(x) \equiv -u^i(x)^{-a}$ is only linear if $a = -1$, which is not allowed since by definition parameter a satisfies $a > 0$. In addition, when a decreases, function $v^i(x)$ becomes more concave, thus reflecting a higher degree of risk aversion.

the parameter of risk aversion a satisfies $a \in [0, +\infty)$. (Note that we do not claim $a > -1$ since $a > 0$ by definition.) In that scenario, individuals are risk averse (but not infinitely), and social indifference curves are curvy.

6 Alternatives to majority voting

In previous sections, we criticized both majority voting, as it could lead to ciclycalities and agenda manipulation, and the Condorcet criterion, as it could lead to no candidate being selected as the winner. A natural question is, then, whether other voting procedures, especially those commonly observed in real life elections, produce Condorcet winners. In this section, we take a relatively applied approach, first describing voting procedures and then comparing them in terms of their properties.

6.1 A list of voting procedures

We first describe two familiar voting procedures: majority rule, and the Condorcet winner.

Majority rule. Society chooses the candidate who is ranked first by more than half of the voters.

Condorcet criterion: Society chooses the candidate who defeats all others in pairwise elections using majority rule.

Example 12.11. Applying majority rule and the Condorcet criterion. Consider three candidates, A, B, and C; and three voters who rank these candidates as follows:

Voter 1	Voter 2	Voter 3
A	A	C
B	C	B
C	B	A

Table 12.9. Applying majority rule and the Condorcet criterion.

For instance, voter 2 ranks candidate A at the top of his list, candidate C next, and candidate B last. If majority voting is used, and each voter cast his ballot for his most preferred candidate, candidate A would be receiving two votes (from voters 1 and 2), candidate B would receive no votes, and candidate C would receive one vote (from voter 3). Hence, candidate A would be the winner according to majority rule. If, instead, the Condorcet criterion was used, a pairwise vote between candidates A and B would yield A as the winner (since A is preferred to B by voters 1 and 2, while B is preferred to A by only voter 3). The winner of this pairwise confrontation, candidate A, would then be paired against the remaining candidate, candidate C, still yielding candidate A as the winner (in this case, A is preferred to C by voters 1 and 2, while only voter 3 prefers C to A). As a consequence, A would be the candidate winning the election according to pairwise majority voting, thus becoming the Condorcet winner. Therefore, the winner according to majority rule and the Condorcet criterion coincide. □

Majority rule with runoff election: If one of the m candidates receives more than half of the votes, then he/she is the winner. Otherwise, a second (runoff) election is held between the two candidates receiving the most votes on the first ballot. The candidate receiving the most votes on the second election is declared the winner.

Example 12.12. Applying majority rule with runoff election. Consider four candidates $\{X, Y, Z, W\}$, and five voters, 1-5, with the following preferences

Voter 1	Voter 2	Voter 3	Voter 4	Voter 5
X	X	Y	W	W
Y	Y	Z	Y	Y
Z	Z	W	Z	Z
W	W	X	X	X

Table 12.10. Applying majority rule with runoff election.

If society uses majority rule with runoff election, candidate X would receive two votes (from voters 1 and 2), candidate Y receives only one vote from voter 3, candidate Z receives no votes, and candidate W receives two votes from voters 4 and 5. Hence, no candidate receives more than 50% of the votes (which in this example would require at least three votes going to the same candidate), and a runoff election is held between the two candidates receiving the most votes on the first ballot, i.e., candidates X and W. In this runoff election, voters 1 and 2 cast their ballot for candidate X (as they prefer X to W), while voters 3-5 vote for candidate W, thus making W the winner. As a practice, note that, if this society uses majority rule (without runoff election) candidates X and W would receive two votes each, thus producing a tie. □

Plurality rule: Each voter ranks the m candidates. Then society chooses the candidate who is ranked first by the largest number of voters.

Example 12.13. Applying plurality rule. Consider the preference profile in Example 12.11. In such a setting, voters 1 and 2 rank candidate A at the top of their list, while voter 3 ranks candidate C. (Candidate B is not ranked first by any voter.) Hence, candidate A is ranked first by the largest number of voters, and is declared the winner under the plurality rule. □

The Hare system: First, each voter indicates the candidate he ranks highest of the m candidates. Second, remove from the list the candidate ranked the highest by the fewest number of voters. Third, repeat the procedure for the remaining $m - 1$ candidates. Continue until only one candidate remains in the list, who is declared the winner.

Example 12.14. Applying the Hare system. Consider again the preference profile in Example 12.11. As we discussed in Example 12.13, candidate A is ranked highest by two voters, C is ranked highest by one voter, while B is not ranked highest by any voter. Hence, candidate B is then removed from the list. Once candidate B is removed from the list, every voter is asked to rank the remaining candidates, A and C, which yields the preference profile in the following table.

Voter 1	Voter 2	Voter 3
A	A	C
C	C	A

Table 12.11. Applying the Hare system.

Candidate A is now ranked highest by two voters (voter 1 and 2), C is ranked highest by one voter (voter 3), which implies that candidate C is removed from the list. Therefore, A is the only candidate remaining, and he is declared the winner. As a practice, note that the winner according to the Hare system coincides with that under plurality voting identified in Example 12.13. $\square$

Variations of the Hare system are used in elections in Australia and Ireland. While the Hare system is often proposed as an alternative to the plurality voting system common in most developed countries, it still suffers from two problems: (1) it can fail to select the Condorcet winner (even if it exists); and (2) it violates monotonicity. (We ask you to show these to points in one of the end-of-chapter exercises.)

The Coombs system: This voting procedure can be understood as the opposite of the Hare system. (To emphasize the differences, the next description italicizes the words that changed relative to the Hare system.) First, each voter indicates the candidate he ranks *lowest* of the m candidates. Second, remove from the list the candidate ranked *lowest* by *most* voters. Third, repeat the procedure for the remaining $m - 1$ candidates. Continue until only one candidate remains in the list, who is declared the winner.

Example 12.15. Applying the Coombs system. Consider three candidates, A, B, and C; and three voters who rank these candidates as follows:

Voter 1	Voter 2	Voter 3
A	A	C
B	C	A
C	B	B

Table 12.12. Applying the Coombs system.-I

(This preference profile is similar to that in Example 12.10 but with a twist in voter 3's preferences.) In this context, candidate B is ranked the lowest by two voters (voters 2 and 3), C is ranked the lowest by only voter 1, while A is not ranked the lowest by any voter. We can then proceed to remove candidate B from the list, as illustrated in the table.

Voter 1	Voter 2	Voter 3
A	A	C
C	C	A

Table 12.13. Applying the Coombs system.-II

We can now identify which candidate is ranked the lowest by the largest number of voters. In particular, candidate C is ranked the lowest by voters 1 and 2, while A is ranked the lowest only by voter 3. Hence, candidate C is removed from the list, leaving candidate A as the only surviving candidate, who is declared the winner according to the Coombs system; a result that coincides with the winner under plurality rule and the Hare system. $\square$

The Borda count: First, each voter gives a score $s \in [1, m]$ to each of the candidates, i.e., he gives m points to his most preferred candidate, $m-1$ points to the second most preferred candidate, ..., and one point to his least preferred candidate. The candidate receiving the highest number of points is declared the winner. This approach to aggregate preferences is rarely used in elections, but it is relatively common in college sports, such as ranking NCAA teams in the US (especially famous for college basketball), or identifying the most valuable player (MVP) in sport tournaments.

Example 12.16. Applying Borda count. Consider three candidates A, B and C; and three voters who are asked to score each candidate with a number 1-3. In this setting, a ballot would ask:

> "Please give a score 1-3 to each of the three candidates in the following list, writing 3 next to your most preferred candidate, 2 for your second most preferred candidate, and 1 for your least preferred candidate."

Here are examples of possible ballots marked by voters 1 and 2.

Voter 1		Voter 2		Total votes	
A	3	A	1	A	4
B	2	B	3	B	5
C	1	C	2	C	3

Table 12.14. Applying the Borda count

Intuitively, candidate A is the most preferred by voter 1, followed by candidate B, and candidate C. In contrast, voter 2 prefers candidate B, followed by C, and ultimately by A. In this context, candidate A receives a total of 4 points, B receives 5 points, and C receives only 3 points. Therefore, candidate B is declared the winner under Borda count. $\square$

Approval voting: First, each voter votes for the k candidates he ranks highest of the m candidates, where k can vary from voter to voter and $k \in (1, m)$. The candidate with the most votes is declared the winner.

Example 12.17. Applying approval voting. Consider four candidates A, B, C, and D; and three voters who are asked to vote for one, two or all three candidates, i.e., $k \in (1, 4)$ which implies that the number of votes, k, must be either $k = 2$ or $k = 3$. In such a setting, a ballot's instructions would read:

> "In the next list of three candidates, please mark a cross next to the candidate (or candidates) you want to vote for. You can mark a cross next to two or three candidates."

Examples of ballots marked by voters 1 and 2 could look like the following:

Voter 1		Voter 2		Voter 3		Voter 4	
A	X	A		A		A	X
B	X	B	X	B	X	B	
C		C	X	C	X	C	
D		D		D	X	D	X

Table 12.15. Applying approval voting.

where voter 1 indicates that he deems candidates A and B as acceptable, while C and D is regarded as unacceptable.[20] Approval voting would then sum the number of votes that each candidate receives: candidate A receives two votes (one from voter 1 and another from voter 4), B receives three votes (from voters 1-3), C receives two votes (from voters 2 and 3), and D obtains two votes (from voters 3 and 4). Since B is the candidate receiving the most votes, he is declared the winner according to approval voting. $\square$

6.2 Evaluating voting procedures

Given the different voting procedures suggested above (and other variations we could easily construct), which criteria can we use to compare them? We next briefly describe two common criteria.

Decisiveness. One common (normative) criterion is to check if the voting procedures picks a winner (referred to as decisiveness). As discussed in previous sections, when the number of candidates is only two, $m = 2$, all voting procedures are decisive. However, when $m > 2$, majority voting and the Condorcet criterion are not necessarily decisive, but all other voting procedures are decisive.

Condorcet efficiency. While most voting procedures identify a winner, i.e., they are decisive, such a winner doesn't need to coincide with the Condorcet winner; as the following example

[20] Voter 2 deems candidates B and C as acceptable, but A and D are unacceptable. A similar intuition applies to the votes from voters 3 and 4.

illustrates. Therefore, majority rule selects a Condorcet winner (if one exists), but all other voting procedures may select a winner that is not necessarily the Condorcet winner (even when one exists).

Example 12.18. Winner does not need to be the Condorcet winner. Consider four candidates running for office, $\{X, Y, Z, W\}$, and five voters, 1-5, with the following preference ranking.

Voter 1	Voter 2	Voter 3	Voter 4	Voter 5
X	X	Y	Z	W
Y	Y	Z	Y	Y
Z	Z	W	W	Z
W	W	X	X	X

Table 12.16. Plurality voting does not produce a Condorcet winner.

Using plurality voting, candidate X is ranked the highest by two voters (1 and 2), while candidates Y, Z, and W are ranked highest by only one voter each. Hence, candidate X is ranked highest by the largest number of voters, and becomes the winner according to plurality voting.[21] However, X is not the Condorcet winner. Indeed, if candidates X and Y are confronted in a pairwise vote, Y wins as Y is preferred to X by three out of five voters (i.e., voters 3-5). The winner of this pairwise election, candidate Y, is then confronted against Z in a pairwise vote, where Y wins again since Y is preferred to Z by four voters (1, 2, 3 and 5). Finally, Y faces the remaining candidate W, which yields Y to be the winner again, as Y is preferred to W by four voters (1-4). In summary, candidate Y beats all other candidates in pairwise votes, and thus becomes the Condorcet winner; which does not coincide with the winner according to plurality voting. $\square$

7 End-of-Chapter Exercises

1. [**Majority voting - Some properties**] Consider majority voting between two alternatives x and y, so the preferences of every individual i over these two alternatives can be represented as $\alpha_i = \{1, 0, -1\}$, where $\alpha_i = 1$ indicates that individual i strictly prefers x to y; $\alpha_i = 0$ reflects that he is indifferent between alternatives x and y; and $\alpha_i = -1$ represents that he strictly prefers y to x. Let us check that, in this context, majority voting satisfies the following three properties: (a) symmetry among agents, (b) neutrality between alternatives, and (c) positive responsiveness.

2. **[Three examples of social welfare functionals**] In this exercise, we consider a setting with two alternatives x and y, and discuss three specific social welfare functionals $F(\alpha_1, .., \alpha_I)$ in parts (a)-(c) below. For each functional, determine whether or not it satisfies the three properties of majority voting (symmetry among agents, neutrality between alternatives, and positive responsiveness).

[21] As a practice, find the winner when other voting procedures are used, such as the Borda count.

(a) Let us first consider the lexicographic social welfare functional

$$F(\alpha_1, .., \alpha_I) \begin{cases} \alpha_1 \text{ if } \alpha_1 \neq 0 \\ \alpha_2 \text{ if } \alpha_1 = 0 \text{ and } \alpha_2 \neq 0 \\ \alpha_3 \text{ if } \alpha_1 = \alpha_2 = 0 \text{ and } \alpha_3 \neq 0 \\ ... \end{cases}$$

Intuitively, society selects the alternative that individual 1 strictly prefers. However, if he is indifferent between alternatives x and y, society follows the strict preferences of individual 2 (if he has a strict preference over x or y). If both individuals 1 and 2 are indifferent between x and y, the strict preferences of individual 3 are considered, and so on.

(b) A constant social welfare functional $F(\alpha_1, ..., \alpha_i) = 1$ for all $(\alpha_1, ..., \alpha_I)$, thus representing that society chooses alternative x over y regardless of the profile of individual preferences $(\alpha_1, ..., \alpha_I)$.

(c) A constant social welfare functional $F(\alpha_1, ..., \alpha_I) = 0$ for all $(\alpha_1, ..., \alpha_I)$, thus indicating that society is indifferent between alternatives x and y regardless of the profile of individual preferences $(\alpha_1, ..., \alpha_I)$.

3. [**Winner does not need to be the Condorcet winner-I**] Consider three candidates running for office, $\{X, Y, Z\}$, and five voters, 1-5, with the following preference ranking.

Voter 1	Voter 2	Voter 3	Voter 4	Voter 5
X	X	X	Y	Y
Y	Y	Y	Z	Z
Z	Z	Z	X	X

(a) Find the Condorcet winner.

(b) Find the winner according to the Borda count.

4. [**Winner does not need to be the Condorcet winner-II**] Consider four candidates running for office, $\{X, Y, Z, W\}$, and five voters, 1-5, with the following preference ranking.

Voter 1	Voter 2	Voter 3	Voter 4	Voter 5
Y	W	X	Y	W
X	Z	Z	Z	X
Z	X	W	X	Z
W	Y	Y	W	Y

(a) Find the Condorcet winner.

(b) Find the winner according to the Hare system.

5. [**Cyclicality in bidimensional ranking.**] Consider that the space of alternatives is bidimensional and, in particular, given by the unit square, i.e., $X = [0, 1]^2$. A specific alternative is, hence, represented now by a pair $x = (x_1, x_2)$, rather than a point in the real line. In this setting, consider three individuals with the following utility functions:

$$\begin{aligned} u_1(x_1, x_2) &= -2x_1 - x_2, \\ u_2(x_1, x_2) &= x_1 + 2x_2, \text{ and} \\ u_3(x_1, x_2) &= x_1 - x_2. \end{aligned}$$

   (a) Find the indifference curve of every individual i for a given utility level $\overline{u}$. Are his preferences convex?

   (b) Show that no Condorcet winner exists. That is, demonstrate that, starting from any pair $x = (x_1, x_2)$ you can find another pair $y = (y_1, y_2)$ which is preferred by at least two of the three individuals. Importantly, you must show this result for all possible positions of pair $x = (x_1, x_2)$ on the unit square.

6. [**Social welfare functions**] Consider an economy with two individuals, 1 and 2. Every individual i's utility function is $u^i(x) = \alpha^i x^i$, where $\alpha^i > 0$, and x^i represents individual i's wealth, where $x^1 + x^2 = x$.

   (a) Find the socially optimal wealth distribution, i.e., the pair of wealth levels $\left(x^1, x^2\right)$ that maximizes the social welfare function

$$W(u^1, u^2) = \left(u^1\right)^\theta + \left(u^2\right)^\theta \quad \text{where } \theta \in (0, 1)$$

   (b) *Numerical example.* Use the social welfare function of part (a), but assume that $\alpha^1 = 1$, $\alpha^2 = \alpha$ where $\alpha \in (0, 1)$, and $\theta = 1/3$. Identify the socially optimal wealth levels x^1 and x^2.

7. [**Rawlsian swf satisfying UDI.**] Consider a society evaluating two alternatives x and y according to a Rawlsian swf. In particular, assume that $u^1(x) = 6$ and $u^2(x) = 12$ for alternative x, and $u^1(y) = 4$ and $u^2(y) = 12$ for alternative y.

   (a) Find the alternative that yields the highest social welfare.

   (b) Let us now apply a linear, but potentially asymmetric, strictly increasing transformation $\psi^i\left(u^i(x)\right) = a^i + bu^i(x)$, where $b = 1$. Identify for which values of parameters a^1 and a^2 utility-difference invariance (UDI) holds, and for which values this property does not hold.

8. [**Gibbard-Satterthwaite theorem**] In this chapter, we analyzed the aggregation of individual preferences into a social preference relation satisfying a set of desirable properties.

However, we assumed individual preferences were truthfully reported by each individual. In this exercise, we examine a setting in which individuals do not necessarily truthfully reveal their preferences. In particular, we are interested in social choice functions that are "strategy proof." First, note that a *social choice function* $c\left(\succsim^1, \succsim^2, ..., \succsim^I\right) \in X$ maps the profile of individual preferences $\left(\succsim^1, \succsim^2, ..., \succsim^I\right)$ into an alternative $x \in X$. That is, society uses the social choice function (scf) to "select" an alternative $x \in X$, using the information in the profile of individual preferences $\left(\succsim^1, \succsim^2, ..., \succsim^I\right)$. Hence, we say that a scf $c(\cdot)$ is *strategy-proof* if every individual i prefers the alternative that the scf selects when he reports his true preferences, $c\left(\succsim^i, \succsim^{-i}\right) = x$, than that arising when he misreports his preferences, $c\left(\succsim'^i, \succsim^{-i}\right) = y$, i.e., $x \succsim^i y$, where $\succsim^{-i}$ denotes the profile of individual preferences by all other agents $\left(\succsim^1, ..., \succsim^{i-1}, \succsim^{i+1}, ..., \succsim^I\right)$. In words, if a scf is strategy proof, individuals have no strict incentives to misreport their preferences, regardless of the preferences other individuals report, $\succsim^{-i}$; which holds true even if the other individuals misreport their preferences. We seek to show, in several steps, Gibbard-Satterthwaite's theorem, which says that: If there are three or more alternatives in X, then every strategy-proof scf is dictatorial.[22] In the next questions of this exercise, we will start showing that (1) a strategy-proof scf must exhibit two properties: Pareto efficiency and monotonicity; and (2) every Pareto efficient and monotonic scf must be dictatorial.

We of course need to define what we mean by Pareto efficient scf: A scf is *Pareto efficient* when every individual i's strict preference for x over y, $x \succ^i y$, where $x, y \in X$, yields the scf to select x, i.e., $c\left(\succsim^1, \succsim^2, ..., \succsim^I\right) = x$. We also define what we mean by monotonic scfs: Consider a initial profile of individual preferences, $\left(\succsim^1, \succsim^2, ..., \succsim^I\right)$, yielding that alternative x is chosen by the scf, i.e., $c\left(\succsim^1, \succsim^2, ..., \succsim^I\right) = x$. Assume that the preferences of at least individual i change from $x \succsim^i y$ to $x \succ'^i y$, for every $y \in X$, i.e., alternative x rises to the only spot at the top of his ranking of alternatives, and the preference for x is not lowered for any individual, i.e., $x \not\prec y$. We then say that a scf is *monotonic* if the scf still selects x under the new profile of individual preferences, $c\left(\succ'^1, \succ'^2, ..., \succ'^I\right) = x$. Hence, loosely speaking, a scf is monotonic if it keeps selecting x as socially preferred when x becomes the top alternative for at least one individual.

(a) Show that strategy-proofness implies monotonicity on the scf.

(b) Use monotonicity to show that the scf must be Pareto efficient.

After demonstrating that strategy-proofness implies monotonicity and Pareto efficiency, we are ready to show the main result of Gibbard-Satterthwaite's theorem (namely, that in a context where the set of alternatives has more than three elements, and where the scf satisfies monotonicity and Pareto efficiency, then such scf must be dictatorial). We

[22] The definition of a dictatorial scf is similar to , in the definition in swf. In particular, we say that a scf $c(\cdot)$ is *dictatorial* if there is an individual d (the dictator) such that, if $x \succsim^d y$ for every two alternatives $x, y \in X$, then the scf selects x, i.e., $c\left(\succsim^1, \succsim^2, ..., \succsim^I\right) = x$. That is, a scf is dictatorial if there is an individual d such that $c(\cdot)$ chooses d's top choices, regardless of the preferences of all other individuals.

will demonstrate that using five steps.

(c) *Step 1.* Consider a profile of strict rankings in which alternative x is ranked highest and y lowest for every individual i; as illustrated in the next table. In this setting, Pareto efficiency implies that the scf must select x.

$\succsim^1$	...	$\succsim^{n-1}$	$\succsim^n$	$\succsim^{n+1}$	...	$\succsim^I$	Social choice
x	...	x	x	x	...	x	x
.		.	.	.		.	
.		.	.	.		.	
.		.	.	.		.	
y	...	y	y	y	...	y	

Consider now that we change individual 1's ranking by raising y in it one position at a time. Show that there must exist an individual n for which the social ranking changes when y is raised above x in individual n's ranking.

(d) *Step 2.* Consider now a different profile of individual preferences in which: x is moved to the bottom of individual i's ranking, for all $i < n$, and x is moved to the second last position in individual i's ranking, for all $i > n$. Show that this change in individual preferences does not change the selection of the scf.

(e) *Step 3.* In this step, we use the assumption that the number of elements in the set of alternatives X is equal or larger than 3. For that, we only need to consider an alternative $z \neq x, y$ in our above steps.

(f) *Step 4.* Consider a profile of individual preferences compatible with those in Step 3. Switch the ranking of alternatives x and y for all individuals $i > n$; as depicted in the next table.

$\succsim^1$	...	$\succsim^{n-1}$	$\succsim^n$	$\succsim^{n+1}$	...	$\succsim^I$	Social choice
.		.	x	.	...	.	x
.		.	z	.		.	
.		.	y	.		.	
z	...	z		z	...	z	
y	...	y	.	y	...	y	
x	...	x	.	x	...	x	

Show that alternative x must be socially selected.

(g) *Step 5.* Argue that the scf must be dictatorial.