2 Arrow's impossibility result

2.1. The axiom system and the theorem

When Arrow showed the general impossibility of the existence of a social welfare function in 1951, quite a few welfare economists were confused. Hadn't Bergson, in his seminal paper of 1938, developed the notion of a social welfare function and hadn't Samuelson (1947) successfully employed this concept in various welfare-economic analyses? What went wrong? Was Arrow right and were Bergson and Samuelson wrong or was it just the other way round?

First of all, Arrow's notion of a social welfare function is different from the Bergson–Samuelson concept in so far as Arrow considered an aggregation mechanism that specifies social orderings for any logically possible set of individual preferences (the multiple profile approach). Bergson claimed that for a given set of individual preferences there always exists the real-valued representation of an ordering for the society (single or fixed profile approach). Furthermore, while Bergson emphasized that any set of value propositions may be introduced when the welfare of a community is being analysed (see section 1.2 above), Arrow was very specific on what basic properties a process should fulfil that maps any set of individual orderings into a social preference.

Let us consider a few examples. Imagine that there is a society with n members one of whom is constantly expressing opinions that all the other members of this society view as unacceptable or at least very strange. Therefore, the aggregation scheme could be such that whenever this particular person prefers a to b, society should prefer b to a. Let us assume now that with respect to two particular alternatives c and d, there happens to be complete unanimity, i.e. everybody strictly prefers c to d. Should society now prefer d to c? This outcome would violate one of the basic properties in the sense used above, viz. the weak Pareto principle to be defined below.

Another aggregation rule could declare that whenever a particular option z is among those alternatives about which the members of the society should make up their mind, alternative z should always be preferred to each of the other options. If one requires that this rule be applied to any given set of individual preferences, a clash with the Pareto principle will again occur.

Finally, a third example. Imagine that in a decision between two social alternatives x and y, not only the individuals' preferences between these two alternatives but also the individuals' preferences between x and some other options z and w and also the individuals' preferences between y and the options z and w should be taken into consideration. Actually, there is a class of aggregation rules which does exactly this. Then again, one of Arrow's basic properties would be violated as we shall see in a moment.

We now wish to state and discuss Arrow's general result in greater detail. In order to do this, we will use the notation and definitions introduced in section 1.3.

Let $\mathcal{E}$ denote the set of preference orderings on X and let $\mathcal{E}'$ stand for a subset of orderings that satisfies a particular restriction. $\mathcal{E}'^n$ will denote the cartesian product $\mathcal{E}' \times \cdots \times \mathcal{E}'$, n times. An element of $\mathcal{E}'^n$ is an n-tuple of preference orderings $(R_1, \ldots, R_n)$ or the profile of an n-member society consisting of preference orderings.

A social welfare function in the sense of Arrow is a mapping from $\mathcal{E}'^n$ to $\mathcal{E}$. Arrow's fundamental result says that there does not exist a social welfare function if this mapping which we denote by $f(R_1, \ldots, R_n)$ is to satisfy the following four conditions:

Condition U (Unrestricted domain). The domain of the mapping f includes all logically possible n-tuples of individual orderings on $X (\mathcal{E}' = \mathcal{E})$.

Condition P (Weak Pareto principle). For any x, y in X, if everyone in society strictly prefers x to y, then xPy.

Condition I (Independence of irrelevant alternatives). If for two profiles of individual orderings $(R_1, \ldots, R_n)$ and $(R'_1, \ldots, R'_n)$, every individual in society has exactly the same preference with respect to any two alternatives x and y, then the social preference with respect to x and y must be the same for the two profiles. In other words, if for any pair x, y and for all i, xR_iy iff xR'_iy, and yR_ix iff yR'_ix, then $f(R_1, \ldots, R_n)$ and $f(R'_1, \ldots, R'_n)$ must order x and y in exactly the same way.

Condition D (Non-dictatorship). There is no individual i in society such that for all profiles in the domain of f and for all pairs of alternatives x and y in X, if xP_iy, then xPy.

Condition U requires that no individual preference ordering be excluded a priori. Even the 'most odd' ordering(s) should be taken into consideration. Condition P, the weak Pareto rule, prescribes that if all individuals unanimously strictly prefer x to an alternative y, the same should hold for society's preference. Condition I, perhaps a bit more difficult to understand than the other conditions, demands that the social welfare function be parsimonious in informational requirements. More concretely, if society is to take a decision with respect to some pair of alternatives (x, y), only the individuals' preferences with respect to this pair should be taken into consideration and not more. The individuals' preferences between x and a third alternative z and the preferences between y and z should not count, nor should the individuals' preferences between z and a fourth alternative w play any role in the social decision between x and y. Finally, there should be no individual in society such that whenever this person strictly prefers x over y, let's say, this preference must become society's preference; and this for all pairs of alternatives from X and for all profiles in the domain of f. Such a person who always has his or her way in terms of strict preferences would have dictatorial power in the preference aggregation procedure, and this is to be excluded.

For Arrow, his four conditions on f (or five conditions if the demand that the social preference relation be an ordering is counted as a separate requirement) were necessary requirements in the sense that 'taken together they express the doctrines of citizens' sovereignty and rationality in a very general form, with the citizens being allowed to have a wide range of values' (Arrow, 1951, 1963, p. 31). The aspect of sovereignty shows itself very clearly in conditions U, P, and D.

Theorem 2.1 (Arrow's general possibility theorem (1951, 1963)). For a finite number of individuals and at least three distinct social alternatives, there is no social welfare function f satisfying conditions U, P, I, and D.

2.2. The original proof

On the following pages, we shall prove Arrow's result. Actually, we shall provide three different proofs of his theorem. These proofs highlight different aspects within his impossibility result and we hope that the three ways of proving his theorem provide sufficient insight into why, at the end, there is a *general* impossibility. The first proof follows very closely Arrow's own proof from the 1963 edition of his book as well as Sen's proof in chapter 3* of his book from 1970. Both proofs show in a transparent way that decisiveness over some pair of social alternatives spreads to decisiveness over all pairs of alternatives which belong to a finite set of alternatives. This phenomenon has sometimes been called a contagion property. Sen (1995) speaks of 'field-expansion' in this context. We start with two definitions which will prove to be very helpful in the sequel.

Definition 2.1. A set of individuals V is almost decisive for some x against some y if, whenever xP_iy for every i in V and yP_ix for every i outside of V, x is socially preferred to y (xPy).

Definition 2.2. A set of individuals V is decisive for some x against some y if, whenever xP_iy for every i in V, xPy.

We now concentrate on a particular individual J and denote the 'fact' that person J is almost decisive for x against y by $D(x,y)$ and the 'fact' that J is decisive for x against y by $\bar{D}(x, y)$. It is immediately clear that $\bar{D}(x, y)$ implies $D(x, y)$; so the former is stronger than the latter. If J is decisive no matter how the preferences of all the other individuals look, J is decisive a fortiori if all the other individuals' preferences are strictly opposed to J's. Now comes a very important contagion result which contains the hardest part of the proof.

Lemma 2.1. If there is some individual J who is almost decisive for some ordered pair of alternatives (x, y), an Arrovian social welfare function f satisfying conditions U, P, and I implies that J must have dictatorial power.

Proof. Let us assume that person J is almost decisive for some x against some alternative y, i.e. for some $x, y \in X, D(x, y)$. Let there be a third alternative z and let index i refer to all the other members of the society. According to condition U, we are absolutely free to choose any of the logically possible preference profiles for this society. Let us suppose that the following preferences hold:

$$\begin{array}{ll} xP_Jy, & yP_Jz \quad \text{and} \\ yP_ix, & yP_iz. \end{array}$$

The reader should notice that for all persons other than J the preference relation between x and z remains unspecified. Since $D(x, y)$, we obtain xPy. Then, because yP_Jz and yP_iz for all other persons, the weak Pareto principle yields yPz. But since f *per definitionem* is to generate orderings, we obtain, by transitivity from xPy and yPz, xPz.

The reader will realize that we started off by using condition U. In the next step, we applied condition P. Then, our argumentation used the ordering property of the social preference relation. What about the independence condition? We arrived at xPz without any information about the preferences of individuals other than person J on alternatives x and z. We have, of course, assumed yP_ix and yP_iz, but according to condition I, these preferences have no role to play in the social decision between x and z. Therefore, xPz must be the consequence of xP_Jz alone, regardless of the other orderings (remember that individual preferences are assumed to be transitive). But this means that person J is decisive for x against z and for the first step in our proof, we obtain: $D(x, y) \rightarrow \bar{D}(x, z)$.

Let us consider the second step. Again assume that $D(x, y)$ but the preferences of all members of the society now read

$$\begin{array}{ll} zP_Jx, & xP_Jy \quad \text{and} \\ zP_ix, & yP_ix. \end{array}$$

Notice that this time i's preferences between z and y remain unspecified. We obtain, of course, xPy from $D(x, y)$ and zPx from condition P. The transitivity requirement now yields zPy. An argument analogous to the one in the previous case, using the independence condition, shows that zPy must be the consequence of $zP_J y$ alone. Therefore, in the present situation we obtain: $D(x, y) \to \bar{D}(z, y)$.

In order to demonstrate the contagion phenomenon, we could continue along the lines of the first two steps. This, however, would be a bit boring for the reader. We could also argue via permutations of alternatives. For example, since we have already shown that $\bar{D}(x, z)$ and therefore $D(x, z)$, we could interchange y and z in $[D(x, y) \to \bar{D}(z, y)]$ and show that $D(x, z)$ implies $\bar{D}(y, z)$. Other interchanges would provide further steps in our proof of the lemma.

Given the verbal argumentation in steps 1 and 2, we want to prove the lemma in a rather schematic way. We shall reiterate steps 1 and 2. In the following scheme, $x \longrightarrow y$ stands for 'x is preferred to y' and $x \longleftarrow y$ stands for 'y is preferred to x'. The following six steps can be distinguished:

1. $J : x \longrightarrow y \longrightarrow z$

$$xPy, yPz \to xPz$$

$i : x \longleftarrow y \longrightarrow z$

$$D(x, y) \to \bar{D}(x, z) \to D(x, z)$$

2. $J : z \longrightarrow x \longrightarrow y$

$$zPx, xPy \to zPy$$

$i : z \longrightarrow x \longleftarrow y$

$$D(x, y) \to \bar{D}(z, y) \to D(z, y)$$

3. $J : y \longrightarrow x \longrightarrow z$

$$yPx, xPz \to yPz$$

$i : y \longrightarrow x \longleftarrow z$

$$D(x, z) \to \bar{D}(y, z) \to D(y, z)$$

4. $J : y \longrightarrow z \longrightarrow x$

$$yPz, zPx \to yPx$$

$i : y \longleftarrow z \longrightarrow x$

$$D(y, z) \to \bar{D}(y, x) \to D(y, x)$$

5. $J : z \longrightarrow y \longrightarrow x$

$$zPy, yPx \to zPx$$

$i : z \longrightarrow y \longleftarrow x$

$$D(y,x) \to \bar{D}(z,x) \to D(z,x)$$

6. $J : x \longrightarrow z \longrightarrow y$

$$xPz, zPy \to xPy$$

$i : x \longleftarrow z \longrightarrow y$

$$D(x,z) \to \bar{D}(x,y) \to D(x,y).$$

Our scheme shows that starting from $D(x,y)$, individual J is decisive (and therefore almost decisive) for every ordered pair from the triple of alternatives (x,y,z), given conditions U, P, and I. Therefore, individual J is a dictator for any three alternatives that contain x and y.

Can this contagion property be extended beyond three alternatives? The answer is 'yes'. We do not want to provide the full argument since the reader will easily see how the reasoning works. Let us consider four elements, viz. x, y, u, and v where both u and v are different from x and y. We start with the triple (x, y, u). Due to the result above and due to condition U, we arrive at $\bar{D}(x,u)$ and $D(x,u)$. Next, we take the triple (x,u,v). Since we have $D(x,u)$, the argumentation above shows that $\bar{D}(u,v)$ and $\bar{D}(v,u)$ follow. Therefore, $D(x,y)$ for some x and y implies $\bar{D}(u,v)$ for all possible ordered pairs (u,v). Thus, the contagion result holds for any finite number of alternatives and the lemma is proved.

The remainder of the proof of Arrow's theorem is rather easy. The logical consequence of the lemma above is that we cannot allow an individual to be almost decisive over some ordered pair of alternatives since this would clash with the condition of non-dictatorship. Let us therefore assume that there is no almost decisive individual. As the reader will see shortly, this leads to a contradiction.

Remember that our frame of argumentation is given by conditions U, P, and I together with the ordering property of f.

By condition P, there is at least one decisive set for any ordered pair (x,y), viz. the set of all individuals. Therefore, there also exists at least one almost decisive set. Among all the sets of individuals that are almost decisive for some pair of alternatives, let us choose the smallest one (not necessarily unique). According to the result of the lemma, it must contain at least two individuals, for the case of one almost decisive person would yield dictatorship, and the proof were complete. Let us call this set V and let V be almost decisive for (x,y). We now divide V into two parts: V_1 contains only a single individual,

V_2 contains all the others from V. Let V_3 be the individuals outside of V. Due to condition U, we postulate the following profile:

$$\begin{array}{ll} \text{For } i \text{ in } V_1: & xP_iy \text{ and } yP_iz \\ \text{For all } j \text{ in } V_2: & zP_jx \text{ and } xP_jy \\ \text{For all } k \text{ in } V_3: & yP_kz \text{ and } zP_kx. \end{array}$$

Since V is almost decisive for (x, y), we obtain xPy. Can zPy hold? If this were the case, then V_2 would be almost decisive for (z, y) due to condition I, since zP_jy and all the other individuals (in V_1 and V_3) prefer y to z. However, according to our assumption, V is a smallest decisive set, and V_2 is a strict subset of V. Therefore, zPy is impossible and thus yRz. Transitivity of the social relation now yields xPz. But then the single member of V_1 would be almost decisive, in contradiction to what we have assumed at the outset. The impossibility result now follows from the lemma.

The reader should note two points. The first refers to the profile that we have used above. It has the structure of the so-called paradox of voting to which we shall come back in the next chapter. The second refers to the fact that the preferences of the individuals in V_3 are not needed in our argumentation. In other words, we could have dispensed with a part of the profile of the voting paradox. It would, perhaps, be a good exercise for the reader to check that the last statement actually holds.

2.3. **A second proof**

The second proof can be found in Jehle and Reny (2001) and in Reny (2001). It is largely based on Geanakoplos (1996). While the first proof was particularly good in bringing out the contagion property of decisiveness, the second proof shows very clearly the function of Arrow's independence condition.

The proof starts by postulating a finite set of alternatives X and n individuals who have strict orderings over these alternatives. The social ordering is assumed to be a weak order. We pick any two distinct alternatives a and b from X. In step 1, alternative a is ranked highest and alternative b lowest by every person $i \in \{1, \ldots, n\}$. Condition P then requires that a is strictly at the top of the social ordering. Imagine now that alternative b is raised, step by step or rank by rank, to the top of individual 1's ordering, while the ranking of all other alternatives is left unchanged. Due to the independence condition, a either remains at the top of the social ordering or is replaced by b. If a remains at the top, raise b in individual 2's ranking until it reaches the top, then do the same in the third, fourth, ... individual's ranking. We know from the weak Pareto condition that 'in the end', when we have moved b to the top of every individual's ranking, the social relation will rank b above a. We now focus on individual m where, after b has risen to the top in his or her ordering, b for the first time is socially preferred to a. Figures 2.1 and 2.2 show the situation just before and just after b was raised to the top of individual m's ordering.

R_1	...	R_{m-1}	R_m	R_{m+1}	...	R_n	social order R
b	...	b	a	a	...	a	a
a		a	b	·		·	·
·		·	·	·		·	b
·		·	·	·		·	·
·		·	·	b		b	·

Figure 2.1.

R_1	...	R_{m-1}	R_m	R_{m+1}	...	R_n	social order R
b	...	b	b	a	...	a	b
a		a	a	·		·	a
·		·	·	·		·	·
·		·	·	·		·	·
·		·	·	b		b	·

Figure 2.2.

In step 2, we introduce the following changes into figures 2.1 and 2.2. We move alternative a to the lowest position of individual i's ordering for $i < m$ and move a to the second lowest position in the orderings of $i > m$. With respect to figure 2.2, the reader will realize that moving a downwards does not alter anything in the relationship between b and any of the other alternatives. Therefore, due to condition I, b must remain top-ranked in the social ordering. The only difference between the new constellations, let's call them $1'$ and $2'$, lies in m's ranking of alternatives a and b. Therefore, due to condition I, b must in situation $1'$ remain socially ordered above every alternative but possibly a. But if b were socially ordered at least as high as a in situation $1'$, then, again due to condition I, b would have to be socially ranked at least as high as a in figure 2.1. But this would be in contradiction of what we had obtained in step 1. Therefore, in constellation $1'$, a is top-ranked socially.

In step 3, we focus on any third alternative c which is distinct from a and b. Remember that in situation $1'$, a was ranked lowest for $i < m$ and second lowest for $i > m$. Individual m had a at the top of the ordering. We now construct a profile in figure 2.3 which is such that the ranking of a in relation to any other alternative in any individual's ordering remains the same as in situation $1'$.

A preference profile is picked where every individual has c ordered above b. The main insight within this step is that due to condition I, alternative a must again be top-ranked socially.

R_1	$\dots$	R_{m-1}	R_m	R_{m+1}	$\dots$	R_n	social order R
$\cdot$		$\cdot$	a	$\cdot$		$\cdot$	a
$\cdot$		$\cdot$	c	$\cdot$		$\cdot$	$\cdot$
c		c	b	c		c	$\cdot$
b		b	$\cdot$	a		a	$\cdot$
a		a	$\cdot$	b		b	$\cdot$

Figure 2.3.

In step 4, the preference profile from figure 2.3 is modified in the following way, and this is the only change: for individuals $i > m$, the rankings of alternatives a and b are reversed. What are the consequences of this alteration? Due to condition I, the social ranking of a versus all the other alternatives except for b remains the same. Can b become top-ranked socially? The answer is 'no' since c must be socially preferred to b due to the Pareto condition. Therefore, a is at the top of the social ordering and c is socially ranked above b.

In the final step 5, we construct an arbitrary profile of orderings with a above b in the ordering of person m. For example, the profile could have, as depicted in figure 2.4, alternative c between a and b in m's ordering whereas all the other individuals order c at the top. Condition I disallows the ranking of c to have any effect on the social ranking between a and b. The ranking of a versus c is as in step 4. Due to our inferences in step 4, a must be ranked above c due to condition I, and c is Pareto-preferred to b. Therefore, by transitivity of the social relation, a is preferred to b, and this holds whenever person m orders a above b.

R_1	$\dots$	R_{m-1}	R_m	R_{m+1}	$\dots$	R_n	social order R
c		c	a	c		c	a
$\cdot$		$\cdot$	c	$\cdot$		$\cdot$	$\cdot$
$\cdot$		$\cdot$	b	$\cdot$		$\cdot$	c
b		b	$\cdot$	b		b	$\cdot$
a		a	$\cdot$	a		a	b

Figure 2.4.

If we now permute alternatives b and c in the arguments above, we obtain the same qualitative result. The ranking of a is above alternative c when person

m orders a above that alternative. And this holds for any alternative distinct from a. In other words, individual m has dictatorial power over a versus any other alternative. Since alternative a was chosen arbitrarily in step 1, it is now evident that there is a dictator for every a from X. But can there be different dictators for different alternatives? The reader will easily see that this would lead to contradictions in the construction of a social ranking whenever these 'potential dictators' have individual orderings that are not the same. Therefore, there can only be one dictator for all elements from X.

2.4. A third diagrammatic proof

The third proof provides a diagrammatic representation of Arrow's theorem and was introduced by Blackorby, Donaldson, and Weymark (1984). In order to keep the diagrams two-dimensional, the proof was given for only two individuals (though the authors briefly indicate how their proof can be extended to more than two persons). The reader certainly remembers our remark at the end of the first (original) proof that two individuals would suffice to show the Arrovian impossibility.

The diagrammatic proof unfolds in utility space. Strictly speaking, this would require us to redefine the whole Arrovian set-up in terms of utility functions that are defined in Euclidian space. This would be extremely cumbersome and very tiring for the reader. Therefore, in the process of redefining concepts, we shall try to be as parsimonious as possible.

The first thing for the reader is to remember from a basic course in microeconomics that a preference ordering can be transformed into a utility function if continuity is postulated in addition to the other properties that turn a binary preference relation into an ordering. In other words, the better-than-or-indifferent set and the worse-than-or-indifferent set with reference to any point in Euclidian space are assumed to be closed sets. The second point to remember is that given any preference ordering and its corresponding utility function, any other utility function which is generated by applying a strictly monotone transformation to the original utility function has the same informational content as the original. This property of ordinal utility will prove to be important in the sequel (remember that the Arrovian framework is purely ordinal). Different individuals can pick different strictly monotone transformations without changing or distorting the original information contained in the preference orderings of the n members of the society. This being said, it is very clear that any 'degree' of comparability of utilities across individuals is excluded. When looking back to our reasoning in the first two proofs, the reader will immediately agree that comparability assumptions had nowhere been postulated.

The social welfare function f à la Arrow is now turned into a social evaluation functional F. Its domain are sets of n-tuples of individual utility functions $u_1, u_2, \ldots, u_n$. Each individual $i \in \{1, \ldots, n\}$ evaluates social states $x \in X$ in terms of utility function $u_i(x)$. We postulate that all logically possible n-tuples of utility functions are admissible (unrestricted domain). The functional F then is a mapping from the set of all logically possible n-tuples or profiles of utility functions into the set of all orderings of X, which we denoted by $\mathcal{E}$ earlier on. For $U = (u_1, u_2, \ldots, u_n)$ being a profile, $F(U) = R_U$ is the ordering generated by F, when the utility profile is U.

After unrestricted domain, the second condition on F that we introduce is Arrow's independence of irrelevant alternatives, now defined for n-tuples of individual utility functions. The meaning of condition I is precisely the same as before. If for any two social alternatives $x, y \in X$ and two utility profiles U' and U'', both x and y obtain the same n-tuple of utilities in U' and U'', then $R_{U'}$ and $R_{U''}$ must coincide on $\{x, y\}$. As promised above, we abstain from giving a redefinition of the independence condition (see, however, section 7.3). Nor do we want to redefine the weak Pareto condition which, of course, also has the same meaning as before. However, we now introduce a condition called Pareto indifference, which requires that if all members of the society are indifferent between a pair of alternatives, the same should hold for society's preference over this pair.

Condition *PI* (Pareto indifference). For all $x, y \in X$ and for all U from the (unrestricted) domain, if $U(x) = U(y)$, then $xI_U y$.

$xI_U y$ means that $xR_U y$ and $yR_U x$, and $U(x) = U(y)$ means that $u_i(x) = u_i(y)$ for all $i \in \{1, \ldots, n\}$.

Conditions U, I and PI have very strong implications for F. Sen (1977b) has shown that the three conditions together imposed on F are equivalent to a property called strong neutrality. Strong neutrality requires that the social evaluation functional F ignore all non-utility information with respect to the alternatives, such as names or rights or claims or procedural aspects. The only information that counts is the vector of individual utilities associated with any social alternative. This 'fact' has been termed 'welfarism' in the literature of social choice theory as well as bargaining theory (we briefly discussed this issue in our introduction) and has been sharply criticized from different angles. The total disregard of non-utility information holds not only within a single utility profile but also across profiles. All this is rather debatable and we shall return to the welfarism issue later on in this book.

For the present analysis, however, the welfaristic set-up has a great advantage. Instead of considering the orderings R_U generated by F, we can focus on an ordering R^* of $\mathbb{R}^n$, the space of utility n-tuples, that orders vectors of individual utilities which correspond to the social alternatives from the given set X. The formal result that Blackorby et al. state in this context (it is due to

D'Aspremont and Gevers (1977)) says that if the social evaluation functional F satisfies the three axioms of welfarism (viz. conditions U, I, and PI), there exists an ordering R^* of $\mathbb{R}^n$ such that for all $x, y \in X$ and all logically possible utility profiles $U, xR_U y \leftrightarrow \bar{u}R^*\bar{\bar{u}}$, where $\bar{u} = U(x)$ and $\bar{\bar{u}} = U(y)$.

D'Aspremont and Gevers show that when the functional F fulfils the three axioms of welfarism, the ordering R^* in utility space inherits these properties. Therefore, one can redefine these conditions together with other requirements and impose them directly on R^*. However, we shall refrain from doing this except for one case, because otherwise the reader would, perhaps, feel terribly bored. Let us just look at the formulation of dictatorship within the new framework. The ordering R^* is a dictatorship if and only if there exists an individual $i \in \{1, \ldots, n\}$ such that for all $\bar{u}, \bar{\bar{u}} \in \mathbb{R}^n$, if $\bar{u}_i > \bar{\bar{u}}_i$, then $\bar{u}P^*\bar{\bar{u}}$. Whenever individual i gets more utility under vector $\bar{u}$ than under vector $\bar{\bar{u}}$, $\bar{u}$ is socially ordered above $\bar{\bar{u}}$.

What is now shown diagrammatically is that if the social evaluation functional F satisfies the three axioms of welfarism, the framework of ordinally measurable, and non-comparable utilities together with the weak Pareto rule are necessary and sufficient for the social ordering R^* to be a dictatorship.

The diagram in figure 2.5 will be widely used in the following proof for the two-person case. Let $\bar{u}$ in $\mathbb{R}^2$ be our point of reference. The plane has been divided into four regions. For the moment, we do not consider the boundaries between the regions but only the interior of the four regions. From the weak Pareto principle, it is clear that all utility vectors in region I are socially preferred to the reference point $\bar{u}$, and the latter is preferred to all vectors in region III.

What can be said about the points in region II in comparison with $\bar{u}$ and the points in region IV against $\bar{u}$? In the following, it will be shown in several steps that either all points in II are preferred to $\bar{u}$ and the latter is preferred to all points in IV or all points in IV are preferred to $\bar{u}$ and $\bar{u}$ again is preferred to all points in II.

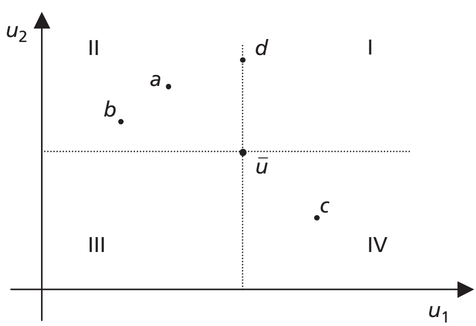

Figure 2.5.

Let us show first that all points in region II (region IV) must be ranked identically against $\bar{u}$. Notice that points in II are such that $u_1 < \bar{u}_1$ and $u_2 > \bar{u}_2$. Consider the points a and b in II and let us assume that $aP^*\bar{u}$. We will now argue that we then obtain $bP^*\bar{u}$ as well. Why? Remember that each of the two persons is totally free to map his or her utility scale into another one by a strictly increasing transformation. It is easy to find a transformation (there are infinitely many) that maps a_1 into b_1 and $\bar{u}_1$ into $\bar{u}_1$. Similarly, one can find another transformation that maps a_2 into b_2 and $\bar{u}_2$ into itself. Figures 2.6(a) and (b) depict two such transformations.

We know that since we are in the framework of ordinal and non-comparable utilities, these transformations do not change the rankings of the two persons. Therefore, if $aP^*\bar{u}$ as assumed, then $bP^*\bar{u}$. Notice that this result holds for any points a, b in the interior of region II. Therefore, all points in the interior of region II are ranked identically with respect to reference point $\bar{u}$ (but not, of course, ranked identically with respect to each other). The reasoning above holds analogously for all points in region IV with respect to $\bar{u}$.

Since R^* is an ordering, three ways of ranking points in region II against $\bar{u}$ are possible: the points in II could be preferred, indifferent, or worse. In our argument above, we had postulated a strict preference against $\bar{u}$. We could also have started by assuming $\bar{u}$ to be preferable to all points in II. The inferences

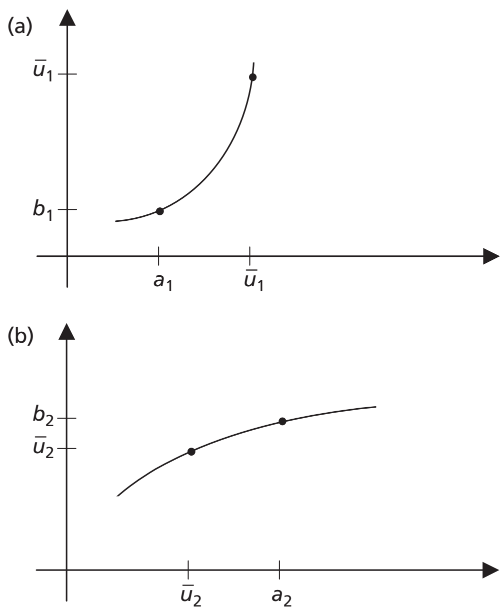

Figure 2.6.

would have been completely analogous. However, indifference between points in II and $\bar{u}$ would lead to a contradiction. We would, for example, have $aI^*\bar{u}$ and $bI^*\bar{u}$. But since R^* is an ordering, we would also obtain aI^*b. Clearly, point a in figure 2.5 must be Pareto-preferred to point b. Therefore, indifference cannot hold.

We now wish to show that the ranking of points in region II against $\bar{u}$ must be opposite to the ranking of points in region IV against $\bar{u}$. Again, we shall use the argument that strictly monotone transformations of individual utility scales do not change the informational content. Let us assume once more that points in region II are preferred to $\bar{u}$, more concretely that $(a_1, a_2)P^*(\bar{u}_1, \bar{u}_2)$. Consider the following transformations for individuals 1 and 2. Change person 1's utility scale such that each point is shifted to the right by $\bar{u}_1 - a_1$, a constant amount, and change person 2's scale such that each point is shifted downwards by $a_2 - \bar{u}_2$, another constant amount. This means that (a_1, a_2) is moved to $(a_1 + (\bar{u}_1 - a_1), a_2 - (a_2 - \bar{u}_2)) = (\bar{u}_1, \bar{u}_2)$ and $(\bar{u}_1, \bar{u}_2)$ is shifted 'south-east' to $(2\bar{u}_1 - a_1, 2\bar{u}_2 - a_2) = (c_1, c_2)$. More briefly, the independent transformations map a into $\bar{u}$ and $\bar{u}$ into c. Since a, by assumption, is preferred to $\bar{u}$, this relationship continues to hold after the transformations, viz. $\bar{u}$ is preferred to c. And from our earlier steps in this proof we infer that $\bar{u}$ is preferred to all points in region IV. Remember that assuming region II to be preferred to $\bar{u}$ was arbitrary. If region II had been assumed to be worse than $\bar{u}$, all points in region IV would turn out to be better than $\bar{u}$.

The proof is almost complete. We still have to deal with points on the boundaries. Consider, for example, point d in figure 2.5. Suppose region II is preferred to $\bar{u}$. For d, there always exists a point in II (such as a) that is Pareto-inferior to d. Therefore, dP^*a and $aP^*\bar{u}$. Transitivity of R^* yields $dP^*\bar{u}$. This result holds for any choice of d. In other words, if two adjacent regions have the same preference relationship to $\bar{u}$, the same ranking holds for any point on their common boundary.

Let us lean back for a moment and see what we have shown. There are two cases possible that are depicted in figures 2.7(a) and (b). If we assume that region II is preferred to $\bar{u}$, then regions I and II and their common boundary are preferred to $\bar{u}$. In this case, the direction of social preference is vertical, and person 2 is a dictator in the sense defined. If region IV is preferred to $\bar{u}$, then regions I and IV and their common boundary are preferred to $\bar{u}$. In this case, the direction of social preference is horizontal, and person 1 is a dictator.

Let us add two more remarks. The first refers to the chosen reference point $\bar{u}$. The position of this point is totally arbitrary for the arguments above. Any other point $\bar{\bar{u}}$ can be reached by transforming the utility scales of person 1 and 2 by adding $\bar{\bar{u}}_1 - \bar{u}_1$ and $\bar{\bar{u}}_2 - \bar{u}_2$ to person 1's scale and person 2's scale, respectively. The proof would then proceed in the same way as before.

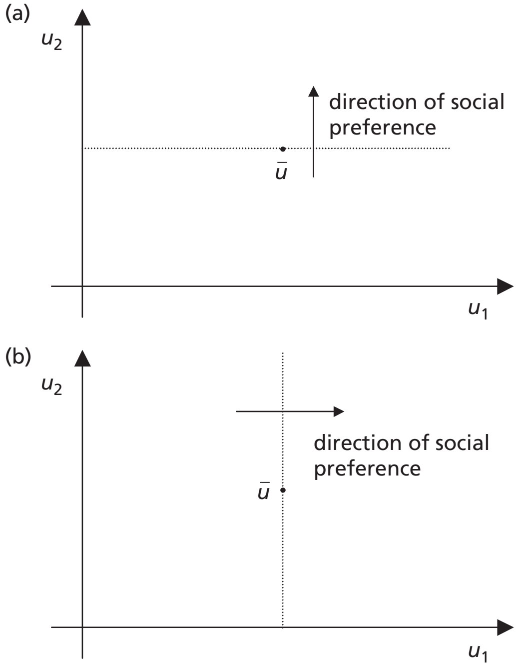

Figure 2.7.

The second remark refers to the fact that we have said nothing about points lying on the dotted lines. As a matter of fact, without introducing a further assumption, nothing precise can be said about the relationship among points on the dotted lines. What we can say is the following. Consider, for example, figure 2.7(a). Because of the informational set-up, any two points on the horizontal dotted line can be ranked either that one of the two points is better than the other, or that it is worse than the other, or that the two points are indifferent. Take a point to the right of $\bar{u}$. If this point is, for example, better than $\bar{u}$, then whenever one moves to the right on any horizontal line, this is an improvement socially. Two things could be done to 'remedy' this situation. One would be to introduce a continuity requirement with respect to R^*. Then all points on the dotted line through $\bar{u}$ would become indifferent to each other. The second thing would be to introduce a strong version of the Pareto principle. A consequence of this assumption would be that whenever the dictator (either person 1 or person 2) is indifferent between two utility allocations, the second person becomes decisive, i.e. determines the social preference. In other words, we obtain a serial or lexicographic dictatorship. Finally, the reader should note how important and far-reaching the assumption of informational invariance with respect to strictly monotone transformations of the individuals' utility scales has been in the proof above.

2.5. A short summary

When Arrow published his by now famous impossibility theorem, the result came as a surprise to various welfare economists. Arrow's negative result was met with disbelief by some. Others, such as Samuelson (1967), claimed that this result might have importance for politics but not so much for economics proper. Others again tried to construct counterexamples to the theorem. Actually, Blau (1957) had a good point which forced Arrow to reformulate the original statement of his theorem to some, though minor degree.

We presented three different proofs to make the logical implications within the Arrovian set-up more transparent and to show the generality of his result. The spread of decisiveness from a single 'cell' (strict preference over one pair) to all other 'cells' in the first proof may be quite stunning for the beginner. The second proof reveals how restrictive (in terms of barring profile information) Arrow's condition of independence of irrelevant alternatives is. The third proof demonstrates the far-reaching consequences of the purely ordinal approach where utilities are determined up to arbitrary strictly monotone transformations, a property that remains largely in the dark in the first two proofs. All the different properties interact, of course, but each proof seems to highlight one of these in particular.

2.6. Some exercises

2.1 Why does the majority rule briefly introduced in Chapter 1 not qualify for an Arrow social welfare function? Please discuss. Show that the weak Pareto principle does not qualify either for an Arrow social welfare function.

2.2 Why is the Borda rule which assigns ranks to positions of alternatives not a counter-example to Arrow's impossibility result? Please discuss.

2.3 Consider the following preference profile for three individuals:

$$xP_1yP_1zP_1w; \quad yP_2zP_2xP_2w; \quad zP_3xP_3yP_3w.$$

According to the simple majority rule we get $yPzPxPw$. Nevertheless, there is something 'going wrong' with this profile. Please discuss.

2.4 Show that if an individual J is decisive over any triple (x, y, z), this individual is also decisive over the quintuple (x, y, z, u, v).

2.5 Show that in the latter part of Arrow's own proof, only the individuals in V_1 and V_2 are needed in order to arrive at a contradiction.

2.6 Write out explicitly so-called constellation $1'$ in the second Arrow proof (section 2.3) and give arguments why b cannot be top-ranked socially though a has lost so many positions in comparison with the situation in figure 2.1.

2.7 Why isn't it possible, in the situation of figure 2.4, for alternative c to be top-ranked socially? This alternative is at the top of $n-1$ orderings and just beaten by a in m's ranking. Please discuss.

2.8 Please construct a positive affine transformation $\varphi(z) = \alpha + \beta z$ such that a_1 is mapped into b_1 and $\bar{u}_1$ is mapped into $\bar{u}_1$ in figure 2.6(a). Do the same in figure 2.6(b) for $\psi(z) = \gamma + \delta z$, i.e. $\bar{u}_2$ is mapped into $\bar{u}_2$ and a_2 is mapped into b_2.

2.9 Show that if the two regions I and II in figure 2.5 are ranked the same in relation to $\bar{u}$, then all points on their common boundary have the same preference relationship to $\bar{u}$.

2.10 Why doesn't the weak Pareto principle help us to determine a preference relationship between points on the dotted line and point $\bar{u}$ in figures 2.7(a) and (b)?

RECOMMENDED READING

Blackorby, Ch., Donaldson, D., and Weymark J. A. (1984). 'Social Choice with Interpersonal Utility Comparisons: A Diagrammatic Introduction'. *International Economic Review*, 25: 327–356.

Reny, Ph. J. (2001). 'Arrow's Theorem and the Gibbard–Satterthwaite Theorem: A Unified Approach'. *Economics Letters*, 70: 99–105.

Sen, A. K. (1970). *Collective Choice and Social Welfare*, chapter 3. San Francisco, Cambridge: Holden-Day.

HISTORICAL SOURCE

Arrow, K. J. (1951, 1963). *Social Choice and Individual Values* (2nd edn.), Chapter 5. New York: John Wiley.

MORE ADVANCED

Sen, A. K. (1995). 'Rationality and Social Choice'. *American Economic Review*, 85: 1–24.

This page intentionally left blank