Decision Theory III - Group Decisions

In this course, decision theory is discussed mostly for the case of a single decision maker, or two
decision makers (bargaining and, later this term, game theory). Now we take a look at some complex
issues related to decision making in case of a group of individuals, all with their own preference order
over a set of rewards—also called ‘outcomes’, ‘options’, or ‘alternatives’.

We first present an important result on combining individual preference rankings into a group pref-
erence. It is important to remark that, when working with preference rankings only, individuals can-
not express the relative strength of their preference (which could, for example, be done using utilities).
Thereafter, we briefly consider how individual utilities could be combined such as to ‘maximise social
well-being’.

1 Social Welfare Functions

We have a collection of rewards \(r_1, r_2, \ldots, r_k\), and a group with \(m\) members. Each member \(i\) has a personal
preference ranking, denoted by \(\succeq_i\). The collection \((\succeq_1, \ldots, \succeq_m)\) is the preference profile for the group.

To simplify notation, in this part of the course we express preference relations between rewards using
the symbols \(>\), \(\sim\) et cetera, without the asterisk, so instead of \(>^*, \sim^*, \ldots\), et cetera, this just to reduce com-
plexity of notation while confusion is unlikely.

Consider the following conditions on preferences

(O1) **Comparability** For each pair of rewards \(r_i, r_j\), we have one and only one of
- \(r_i > r_j\) (\(r_i\) preferred over \(r_j\)),
- \(r_i < r_j\) (\(r_j\) preferred over \(r_i\)), or
- \(r_i \sim r_j\) (no preference between \(r_i\) and \(r_j\)).

(O2) **Transitivity**
- If \(r_i \geq r_j\) and \(r_j \geq r_k\), then \(r_i \geq r_k\).

Note that \(r_i > r_j\) is just a brief notation for \(((r_i \geq r_j) \land (r_i \neq r_j))\). Similarly, \(r_i \sim r_j\) is a brief notation
for \(((r_i \geq r_j) \land (r_i \leq r_j))\). The two conditions above imply that all the usual properties of > and \(\sim\) are
satisfied as well; for instance:
- if \(r_i > r_j\) and \(r_j > r_k\), then \(r_i > r_k\),
- if \(r_i > r_j\) and \(r_j \sim r_k\), then \(r_i > r_k\),
- and so on.

**Definition 1.** A **social welfare function** (SWF) is a function which operates on preference profiles, all
obeying (O1) and (O2), and yields a group (or social) preference \(\succeq_g\) over the rewards which obeys (O1)
and (O2).
2 Four Reasonable (?) Conditions on Social Welfare Functions

We present Arrow’s theory for group decisions. We present his theorem with 4 conditions or axioms, also knowns as Arrow’s axioms, stated below. In the literature these can be found with several variations, sometimes re-formulated to a set of more than 4 axioms and possibly slightly weakened. The formulation chosen here is convenient both with regard to interpretation of the axioms and the proof of the main theorem.

(U) Unrestricted domain $\geq_g$ is defined, and obeys (O1) and (O2), for any preference profiles $(\geq_1, \ldots, \geq_m)$ satisfying (O1) and (O2).

(D) No dictatorship There is no individual $i$ for which $\geq_i$ automatically becomes $\geq_g$.

(P) Pareto’s condition If each individual has $x_i > y_i$, then the group has $x >_g y$.

(I) Independence of irrelevant alternatives If some rewards are removed, and everyone keeps the same preference over other rewards, then so should the group.

3 Arrow’s Impossibility Theorem

Theorem 4. Let there be at least three rewards and at least two individuals. Then there is no SWF that meets all of the four conditions (U), (D), (P) and (I).

This theorem means that for any SWF, preference profiles can be found that contradict at least one of these 4 conditions. There is no “fair” voting system in the sense that you will always have to violate one of the four conditions of Arrow’s theorem. However, there are voting systems that satisfy any three of them (see exercises).

To prove Arrow’s theorem, we introduce a few more definitions.

• An individual $j$ is decisive (DC) for $c$ over $d$ if, whenever $c >_j d$, then $c >_g d$.

• A subgroup $V$ of individuals is almost decisive (ADC) for reward $a$ over $b$ if

$$\left\{(\forall i \in V)(a >_i b) \land (\forall i \notin V)(b >_i a)\right\} \Rightarrow a >_g b$$

• A minimal almost decisive subgroup (MAD) $V$ is ADC for some $a$ over some $b$, and no subgroup of $V$ is ADC for any $c$ over any $d$.

MAD groups must exist. Indeed, by condition (P), the whole group is ADC for any $a$ over $b$, so either the whole group is MAD, or we can eliminate individuals until we reach a MAD group. (This is under the assumption that we do want (P) to be satisfied!)

It is important to notice that condition (I) implies that group preference between $a$ and $b$ only depends on individual preferences between $a$ and $b$.

---

1Kenneth Arrow, born in 1921, was (jointly) awarded the 1972 Nobel Memorial Prize in Economics for a wide variety of contributions including the theory we look at here. He was, and remains, the youngest ever recipient of the prize. In addition to his many direct contributions to economics, many of his PhD students became eminent researchers, with 5 of them also winning the prize in later years, including John Harsanyi who we will encounter soon. Arrow is still active (!) as Emeritus Professor at Stanford.
In his book *Decision Theory: An Introduction to the Mathematics of Rationality*, Simon French calls the above ADC property ‘decisive’, yet we think that our terminology is more appropriate. However, French’s slightly different presentation of Arrow’s theorem and proof are also very interesting.

We only need to use ‘decisive’ for a single individual in the proof below, of course it would be simple to generalize that definition to any subgroup. A dictator would be decisive for all \( a \) over all \( b \).

**Proof of Arrow’s Theorem.** We now have all we need to prove Arrow’s theorem. This will be done in two steps. First, we show that a MAD group has just one person (we discussed above that a MAD group certainly exists). Secondly, we show that this person is a dictator, a contradiction which then proves the theorem. Throughout the proof we make repeated use of the condition (I) by chosing other rewards at convenient places in individuals’ preference rankings.

- **We first prove that the MAD group has 1 person.**

  Suppose \( V \) is a MAD group, which is ADC for \( a \) over \( b \) (two of the possible rewards). Suppose \( V \) has more than 1 individual. Choose an individual \( j \in V \). Let \( W \) be the other individuals in \( V \), and let \( U \) be all individuals not in \( V \) (which could be an empty set). Choose a third reward \( c \). Consider the group preference if individuals’ preferences are (note: we can do this because of the condition (U)):

  1. \( a >_j b >_j c \)
  2. \( c >_i a >_i b \), for all \( i \in W \)
  3. \( b >_i c >_i a \), for all \( i \in U \)

  As \( a >_i b \) for all \( i \in V \), and \( b >_i a \) for all \( i \notin V \), and \( V \) is ADC for \( a \over b \), we have group preference \( a >_b b \).

  As \( c >_i b \) for all \( i \in W \), and \( b >_i c \) for all \( i \notin W \), we cannot have \( c >_i b \) as this would imply that \( W \) is ADC for \( c \) over \( b \), contradicting that \( V \) is MAD (since \( W \subset V \)). Hence, \( b \geq_c c \).

  Since \( a >_b b \) and \( b \geq_c c \), transitivity (part of condition (U)) implies \( a >_c c \). But \( a >_j c \) and \( c >_i a \) for all \( i \neq j \). So, \( j \) is ADC for \( a \) over \( c \). This contradicts the fact that \( V \) is MAD, as \( j \in V \). So, to avoid this contradiction, \( V \) must have only a single member.

- **Next we prove that the single person in the MAD group is a dictator.**

  We show that individual \( j \), from step 1, is DC for every \( c \) over every other \( d \), which means that he is a dictator. From step 1, we know that \( j \) is ADC for some particular reward \( a \) over some particular \( b \). This proof again consists of several steps.

  1. **Individual \( j \) is DC for \( a \) over any \( d \neq b \).**

     (We use that, by (I), we can place \( b \) where we want in individuals preference rankings, while the group preference between \( a \) and \( d \) should be the same independent of \( b \). We can, of course, not assume anything about the ranking of \( a \) and \( d \) for individuals other than \( j \) to prove this property.)

     Suppose \( a >_j b >_j d \) (which implies that \( b \) must be a different reward than \( a \) and \( d \)), and \( b >_j a \) and \( b >_j d \) for all \( i \neq j \). As \( j \) is ADC for \( a \) over \( b \), we have \( a >_b b \). By Pareto (P), we have \( b >_d d \) as all individuals prefer \( b \) over \( d \). Hence, transitivity (part of (U)) implies \( a >_d d \). So, if individual \( j \) holds \( a >_j d \), then \( a >_d d \). So indeed \( j \) is DC for \( a \) over any \( d \neq b \).

  2. **Individual \( j \) is DC for any \( c \neq a \) over any \( d \neq a \).**

     (We compare \( c \) and \( d \), so by (I) we can place \( a \) where we want.)

     Suppose that \( c >_j a >_j d \), and \( c >_i a \) for all \( i \neq j \). As \( j \) is DC for \( a \) over \( d \) (by 1), we have \( a >_d d \). Pareto (P) implies that \( c >_d a \) as all individuals hold this preference. Transitivity implies \( c >_d d \). Hence, if \( c >_j d \) then \( c >_d d \) for any \( c \neq a \) and \( d \neq a \).
3. Individual \( j \) is DC for any \( c \neq a \) over \( a \).

(We now place \( d \) where we want to show that, whenever \( c \succ_j a \), the group holds \( c \succ_a a \) for \( c \neq a \).)

Suppose \( c \succ_j d \succ_j a \), and \( d \succ_i a \) for all \( i \neq j \). As \( j \) is DC for \( c \neq a \) over \( d \neq a \) (by 2), we have \( c \succ_d a \). By (P) we have \( d \succ_a a \), and transitivity gives \( c \succ_a a \).

4. Individual \( j \) is DC for \( a \) over \( b \).

(We know ADC for this combination, but must prove DC. Check that this will complete the proof of step 2, as all pairs of rewards will have been considered. For this step, we conveniently place \( c \).)

Let \( a \succ_j c \succ_j b \) and \( c \succ_j b \) for all \( i \neq j \). By 1 we have \( a \succ_j c \), and (P) gives \( c \succ_j b \), so transitivity indeed gives \( a \succ_j b \).

We have shown that the single person in the MAD group (which exists) is a dictator, which obviously contradicts the condition (D). So, we cannot generally satisfy all 4 conditions in Arrow’s theorem. \( \square \)

### 4 Utilitarianism

We want to achieve the following: Given choice between social options, rank these such as to place at the top the option producing the greatest pleasure for everyone, \( \textit{et cetera} \), and at the bottom the option producing the least pleasure for everyone. The main questions are: (1) How do we measure individual pleasure? (2) How do we combine pleasure of individuals to produce ‘group pleasure’?

John Harsanyi proposed the following theory\(^2\). Each individual expresses utility for each option. Suppose \( m \) citizens face \( r \) social choices \( x_1, x_2, \ldots, x_r \). Each citizen is ‘rational’, so citizen \( i \) has a utility function \( U_i(x) \) over these choices. In addition to these \( r \) options, we assume that one option can be introduced, say \( x_0 \), for which all citizens agree that it would be as bad as what they consider the worst real option. Utility is unique up to a linear transformation, so suppose that each \( U_i \) is scaled to lie in \([0, 1]\), with citizen \( i \)’s utility actually being equal to 0 for his least liked option, as well as for \( x_0 \), and equal to 1 for his most preferred choice.

Introduce the Planner, \( P \), who wants to make good choices for the community and knows each \( U_i \). \( P \) is ‘rational’ and will create utility function \( W \). Suppose further that \( P \) obeys two conditions for ‘social rationality’:

- **Anonymity (A):** It does not matter which utility corresponds to which citizen. \( P \) would create the same \( W \) function under any permutation of \( U_1, \ldots, U_m \).

- **Strong Pareto principle (SP):** If each citizen is indifferent between two options, then so is the planner. And, if no citizen prefers \( x \) to \( y \), and some prefer \( y \) to \( x \), then the planner prefers \( y \) to \( x \).

Let \( U(x_i) = (U_1(x_i), U_2(x_i), \ldots, U_m(x_i)) \) denote the vector of utilities for each choice \( x_i \), then \( W(x_i) \) is a function of \( U(x_i) \), let us write \( W(x_i) = f(U(x_i)) \), for some function \( f \).

**Theorem 6.** If the planner obeys (A) and (SP) then the function \( f \) must be \( f(U_1, U_2, \ldots, U_m) = U_1 + U_2 + \ldots + U_m \), so

\[
W(x) = \sum_{i=1}^{m} U_i(x)
\]

(We can say that the planner must act as sum utilitarian.)

\(^2\)This theory was part of his work for which he was awarded the 1994 Nobel Memorial Prize in Economics, jointly with John Nash (whose theory for bargaining we discussed) and Reinhard Selten. Harsanyi initiated also Bayesian game theory, which we do not address in detail in this module but which is of great interest.
Proof. We prove this theorem for two citizens only, for \( m \) citizens the proof is similar. The proof consists of four steps.

1. Fix the lowest possible value for \( W \) at 0, which value is assumed for the introduced ‘worst possible option for all’ \( x_0 \) with \( U(x_0) = (U_1(x_0), U_2(x_0)) = (0, 0) \). Then \( x_0 \) is rated worst by both citizens, so by (SP) \( x_0 \) is rated worst by \( P \), so \( W(x_0) = f(0, 0) = 0 \).

2. Consider any \( x \) with \( U(x) = (U_1, U_2) \). Consider the gamble
   \[
   G = px \oplus (1-p)x_0
   \]
   (Remember: this is the gamble where with probability \( p \) the outcome is option \( x \), and with probability \( 1-p \) the outcome is option \( x_0 \).) Individual 1’s utility for this gamble is \( U_1(G) = pU_1(x) + (1-p)U_1(x_0) = pU_1 \), and similarly individual 2’s utility for this gamble is \( U_2(G) = pU_2 \). Hence, \( U(G) = (pU_1, pU_2) \), so
   \[
   W(G) = f(pU_1, pU_2) \tag{i}
   \]
   As \( W \) is itself a utility function, we also have that \( W(G) = pW(x) + (1-p)W(x_0) \), so
   \[
   W(G) = pf(U_1, U_2) \tag{ii}
   \]
   Clearly, (i) and (ii) give
   \[
   f(pU_1, pU_2) = pf(U_1, U_2) \tag{*}
   \]

3. Fix \( f(1, 0) = 1 \) (which can be done as utility function \( W \) must be specified up to a linear transformation). Then, by (A), \( f(0, 1) = 1 \). Thus, by (*), we have (recall that 0 \( \leq U_1, U_2 \leq 1 \)):
   \[
   f(U_1, 0) = U_1f(1, 0) = U_1 \quad \text{and} \quad f(0, U_2) = U_2f(0, 1) = U_2 \tag{**}
   \]

4. Consider the gamble \( L = \frac{1}{2}y \oplus \frac{1}{2}z \), where \( y \) and \( z \) are such that \( U(y) = (U_1, 0) \) and \( U(z) = (0, U_2) \). (Note: these \( y, z \) do not have to be options from the given set, they are just introduced to analyse general properties of the function \( f \). The values \( U_1, U_2 \) here are general values, which could also be equal to 0.)
   Then \( U_1(L) = \frac{1}{2}U_1(y) + \frac{1}{2}U_1(z) = \frac{1}{2}U_1 \), and similarly \( U_2(L) = \frac{1}{2}U_2 \), so \( U(L) = (\frac{1}{2}U_1, \frac{1}{2}U_2) \). Hence, by (*), we have
   \[
   W(L) = \frac{1}{2}f(U_1, U_2) \tag{iii}
   \]
   Alternatively, (**) implies that \( W(L) = \frac{1}{2}W(y) + \frac{1}{2}W(z) = \frac{1}{2}(U_1 + U_2) \), which together with (iii) gives
   \[
   f(U_1, U_2) = U_1 + U_2
   \]
   □