4 On Reconciling Arrow's Theory of Social Choice with Harsanyi's Fundamental Utilitarianism

Peter J. Hammond*

1 INTRODUCTION AND OUTLINE

The last paragraph of Kenneth Arrow's Nobel Lecture is characteristic of the man:

> The philosophical and distributive implications of the paradox of social choice are still not clear. Certainly there is no simple way out. I hope that others will take this paradox as a challenge rather than as a discouraging barrier (Arrow 1973a; 1974).

Here I take up the challenge, though I will not claim to resolve the paradox, but will suggest that it may not be insoluble. The approach I shall present involves using cardinalizations of both individual and social welfare measures based on behaviour in risk-taking situations, as in the theory Arrow (1951c; 1965; 1971) did much to clarify, even if von Neumann and Morgenstern (1944) were the originators. This approach can, I shall argue, be justified by analysing sequential decisions in decision trees, a topic that relates closely to one of Arrow's earliest publications (Arrow, Blackwell and Girschik 1949). In particular, I shall make much use of the 'consequentialist axiom' of

*It is a particular pleasure on this occasion to express my sincere thanks for the constant encouragement, support and helpful comments I have received from Kenneth Arrow since we first met over fourteen years ago. Research support from the National Science Foundation is gratefully acknowledged.

rational choice, discussed elsewhere (Hammond 1983a; 1985b; 1986). Its motivation is briefly discussed in Section 2 and the definition of consequentialism is given in Section 3.

Thereafter, the bulk of the chapter proceeds through several logical steps to derive a form of utilitarianism closely akin to that of Vickrey (1945; 1960) and, more especially, of Harsanyi (1953; 1955; 1976; 1977a; 1978). The latter sections consider how the contradictory postulates of Arrow's Impossibility Theorem can be modified to accommodate this form of utilitarianism. That may help to explain the title of this chapter.

More specifically, Section 3 recapitulates an argument first presented in Hammond (1977a) that shows how Arrow's (1959) characterization of ordinal choice functions can be used to establish that consequentialism implies ordinality. That is, behaviour must maximize a complete reflexive and transitive preference ordering defined on the space of consequences. Section 4 introduces risky consequences, described by (simple) probability distributions, and claims that behaviour that is consequentialist and also weakly continuous in probabilities must maximize the expected value of a von Neumann-Morgenstern utility function defined on the space of consequences.

To have a theory of social choice based on 'individual values' (Arrow 1951a), or on 'individualistic ethics' (Harsanyi 1955), two steps are taken. First, 'interpersonal consequences' are defined in Section 5 in a way that gives each individual his own consequence space consisting of pairs of personal characteristics and personal consequences. Then individual norms are defined in Section 6 as social norms for 'personal' decision trees where only the individual's own consequences are affected by behaviour in the tree. By their very construction, an individual norm thus coincides with the social norm in any personal decision tree. This is effectively an 'ethical' form of 'liberalism' (Sen 1970a; 1970b) where each individual has rights within personal decision trees. 'Ethical liberalism' is also a form of 'private' liberalism that actually *implies* the strong Pareto principle rather than conflicting with it (cf. Hammond 1982b; Coughlin 1984). And it accords with Arrow's (1967) expressed view that it is 'personal welfare' that counts in social choice theory. From these postulates, it then follows from an argument of Harsanyi (1955) that prescribed behaviour must maximize the expected value of a positively weighted sum of individuals' von Neumann-Morgenstern utility functions whose expected values represent the respective individual norms. This is the topic of Section 7. At this point we have reached an idealized

form of Vickrey-Harsanyi utilitarianism based on individual behaviour norms when consequences are risky.

One weakness of Vickrey-Harsanyi utilitarianism, noticed by Harsanyi (1955) and Pattanaik (1968), among others, is that it is not obvious how the positive welfare weights should be determined. A possible procedure – suggested, indeed, by Harsanyi (1955; 1977a; 1977b) and also, more specifically by Arrow (1977a), although Arrow did not consider risky consequences – is to define a consequence space of Gorman-Lancaster characteristics so broadly that all possible individuals share the same individual norm upon it. Each person's characteristic is then defined as the mapping from individual consequences to this space of Gorman-Lancaster characteristics. It follows that all individuals will share the same 'fundamental preferences' (Kolm 1972) on the common space of personal characteristic and consequence pairs. This is discussed in Section 8, which also shows how an obvious anonymity condition serves to give every individual's fundamental von Neumann-Morgenstern welfare function the same weight. So the social welfare function is identical to the expected utility of any individual in an 'original position' with an equal chance of being in the position of any individual in the existing society. This is (Harsanyi's) 'fundamental utilitarianism'.

The hardest condition of Arrow's social choice theory for utilitarianism to meet is, of course, independence of irrelevant alternatives (IIA). The next section shows how even a weakened version of this condition is likely to be violated by utilitarianism, unless 'ethical liberalism' is weakened sufficiently to allow a dictator whose individual norm decides everything, or unless the domain of admissible individual norms is severely restricted. However, independence of irrelevant alternatives *is* consistent with existence of a cardinal, though dictatorial, von Neumann-Morgenstern social welfare function provided that risky consequences are allowed. To reconcile utilitarianism with IIA, however, some weakening of IIA is clearly required. Indeed, this is the crucial step in my attempted reconciliation of Arrow's social choice theory with utilitarianism. I propose a significant weakening of IIA to 'independence of ethically irrelevant mixed consequences'. Specifically, consequences that are 'relevant' to prescribed behaviour in a decision tree are those that are possible consequences of acts in that tree. But consequences that are 'ethically relevant' are extended to include consequences derived from relevant consequences by having individuals exchange both personal characteristics and personal consequences. In addition, all decision trees

whose consequences are probability mixtures of ethically relevant pure consequences are allowed into consideration. *This* weakened form of IIA *is* satisfied by the 'fundamental utilitarianism' of Section 8. It may also be more appropriate for ethical social choice than Arrow's original condition, as I have even presumed to suggest in my choice of terminology. It is somewhat related to Strasnick's (1977, p. 675) idea of considering 'morally relevant' information about individuals' choices.

Another of Arrow's conditions violated by fundamental utilitarianism is unrestricted domain. For one thing, individual norms decide only in personal decision trees. For another, the existence of a fundamental individual norm is also restrictive, though an individual with personal characteristics can have unrestricted preferences over personal consequences as those characteristics vary. In Section 11 I argue that neither kind of domain restriction is unacceptable, and that indeed the domain restriction that individual norms decide only in personal decision trees is more acceptable in economic environments than an unrestricted domain seems to be.

This essentially completes my attempt to reconcile Arrow's theory of social choice with Vickrey-Harsanyi utilitarianism. There are some serious loose ends, however. It is rather doubtful if we could ever define a space of Gorman-Lancaster characteristic consequences broadly enough to ensure that all individual norms coincide in this space. Yet this is what underlies fundamental utilitarianism. So we should perhaps recognize that different individuals have different ethical opinions – different versions of the fundamental individual norm, specifically – and try to reconcile them. Then, however, we are back with Arrow's original social choice problem, in effect, and it is hard to escape the conclusion that the fundamental individual norm has to be dictated. In other words, different ethical views can be reconciled only through agreeing about the fundamental individual norm. This defect does not seem to be fatal to fundamental utilitarianism, however, since most other ethical theories face similar problems, if not worse.

Finally, a matter on which my own views may appear to have changed. Previously, in Hammond (1977b, p. 63; 1981; 1982a, p. 196) I have clearly accepted the view expressed by Arrow (1951a; 1952; 1963, p. 10; and, in particular, 1983a, p. 48) that:

> the theory of cardinal utility constructed by von Neumann and Morgenstern to explain risk-bearing behavior has no relevance to the

> present discussion . . .; there is no reason to accept this particular utility scale for social decisions; to do so would amount to determining the distribution of income by the tastes of individuals for gambling.

Indeed, I even quoted a passage from Arrow (1963, p. 10) to the same effect not just once but *twice*. Yet, at first sight, it seems that here, in propounding fundamental utilitarianism – a form of Vickrey–Harsanyi utilitarianism – I am indeed contradicting this earlier view. In Section 13 I try to show that there is no contradiction, provided that the fundamental individual norm that lies behind fundamental utilitarianism is allowed to differ from each individual's *actual* behaviour, and from gambling behaviour in particular. This, of course, leads to a form of paternalism. Rights will be infringed, but if this matters, rights should be included among consequences. There are also violations of the 'consumer sovereignty' assumption that lends normative significance to the 'basic theorems of classical welfare economics' as in Arrow (1951b) and Debreu (1954). This is especially true when the allocation of risk bearing is considered. If individuals' attitudes to risk and tastes for gambling are not to determine the distribution of income, then Arrow's (1953; 1963) complete contingent security markets no longer serve to produce welfare optimal allocations. Nor do contingent lump-sum transfers generally serve as a substitute for complete contingent security markets when individuals cannot be expected to make optimal decisions in security markets. Exceptions are those considered by Starr (1973), and generalized by Hammond (1981). They have the very special property that, with welfare optimizing contingent lump-sum transfers, neither poor information nor poor decisions in security markets can have any adverse impact on the allocation in a sequence of competitive spot markets. Section 15 contains concluding remarks.

2 MOTIVATION FOR CONSEQUENTIALISM

A fundamental tenet of normative behaviour is that only consequences should matter. As Arrow (1951c, p. 404) put it:

> There is a set of conceivable actions which an individual could take, each of which leads to certain consequences . . . Among the actions actually available, then, that action is chosen whose consequences are preferred to those of any other available actions.

Later, Arrow (1965; 1971, p. 50) formulated an assumption he called 'valuation of actions by consequences'. Savage (1954) actually *defines* an act as a mapping from (uncertain) states of the world to consequences. After presenting his well-known omelette example, he concludes:

> If two different acts had the same consequences in every state of the world, there would from the present point of view be no point in considering them two different acts at all. An act may therefore be identified with its possible consequences. Or more formally, an *act* is a function attaching a consequence to each state of the world (Savage 1954, p. 14).

The 'consequentialist' tenet thus holds that, in determining what act should be chosen, it is enough to consider the consequences of all the acts that are possible, and to choose an act which has the best possible consequences. It is a fundamental postulate of modern decision theory. Modern moral philosophers, especially Williams in Smart and Williams (1973) and with Sen in the introduction to Sen and Williams (1982), have produced sophisticated arguments against consequentialism. Nevertheless, as I have tried to argue in Hammond (1986), many of the objections of Williams and others, based on ideas like heroism and personal integrity, appear not to apply to public decisions such as those which determine economic policy. Hare (1981) goes further and suggests that virtually all examples produced by the anticonsequentialists are the kind of hard case that makes bad law. Moreover, consequentialism is very elastic: many objections arise because consequences are defined too narrowly, e.g. by excluding feelings of personal integrity and self-esteem. If consequences are defined broadly enough, most objections lose their force, although, of course, consequentialism also becomes less able to make precise prescriptions. In a nutshell, the space of consequences should include everything of consequence.

3 CONSEQUENTIALISM AND ORDINAL CHOICE

It is in any case as a theory of normative behaviour that I want to discuss consequentialism, and I shall do so when the *space of consequences* Y is fixed. Then, when consequentialism is applied to all finite decision trees whose outcomes give rise to consequences within

the fixed space Y, it has strikingly powerful implications, as shown in Hammond (1985b). Before discussing these implications, however, let me explain how the consequentialist behaviour 'pre-axiom' is formulated. Loosely, weak consequentialism is defined to mean that, whenever each of two decision trees gives rise to identical sets of available consequences which result from the various possible acts in each tree, then behaviour in each tree should also give rise to identical sets of 'chosen' consequences. Behaviour thus corresponds to the choice of consequences.

A *behaviour norm* is then a mapping $\beta(T,\gamma,n)$ defined on triples consisting of finite decision trees T, consequence mappings γ from the sets of acts $A(T)$ in T to the space of consequences Y, and *decision nodes* n at which the agent has to decide what node to move to next. The image $\beta(T,\gamma,n)$ must be a non-empty subset of the set $N_{+1}(n)$ of all nodes of T which immediately succeed n in T. Thus $\beta(T,\gamma,n)$ represents the set of possible moves to a node n' which succeeds n in T, according to the norm β applied to the tree T with consequence mapping γ. Of course, β depends on γ whenever behaviour is affected by consequences.

Weak consequentialism of a behaviour norm is equivalent to there being a *consequence choice function* $C: \mathscr{F}(Y) \to \mathscr{F}(Y)$, where $\mathscr{F}(Y)$ is the set of all non-empty finite subsets of Y, and where $C(Z) \subset Z$ for all $Z \in \mathscr{F}(Y)$. This consequence choice function must have the property that, in a decision tree whose acts lead to consequences in the feasible set of consequences Z, $C(Z)$ must be the set of possible consequences of behaviour. In Hammond (1985b) it is shown formally that:

> *Proposition 1:* A behaviour norm is weakly consequentialist if and only if it corresponds to a consequence choice function.

The second part of 'consequentialism' is closely related to the requirement that Strotz (1956) called 'harmony' or 'consistency', and which Selten (1965; 1973) called '(subgame) perfectness' (to be distinguished from '(trembling-hand) perfectness' which Selten (1975) introduced later). The requirement is applied to '*continuation subtrees*' of each decision tree – that is, to the set of subtrees $T(n)$ with initial node one of the nodes n of the original decision tree T, and which include all the other nodes of T which succeed n. A behaviour norm $\beta(T,\gamma,\cdot)$ specifies behaviour at all decision nodes of $T(n)$, because these are decision nodes of T. But $T(n)$ is itself a decision tree, with a set of acts $A(T(n))$ consisting of restrictions to the set of decision nodes in $T(n)$ of the acts

in $A(T)$. There is also a mapping $\gamma(n)$: $A(T(n)) \to Y$ from acts to consequences in Y which is naturally induced by applying γ: $A(T) \to Y$ to the set of restricted acts $A(T(n))$. The behaviour norm β thus prescribes both $\beta(T,\gamma,n')$ and $\beta(T(n),\gamma(n),n')$ at any decision node n' of the continuation subtree $T(n)$. Unless these two behaviour sets are equal, there is a clear inconsistency or imperfection in the behaviour norm. So it is natural to require the behaviour norm to be *dynamically consistent* or *continuation perfect* in the sense that, given any decision tree T with consequence mapping γ, and given any node n of T:

$$\beta(T,\gamma,n') = \beta(T(n),\gamma(n),n')$$

for all decision nodes n' of the continuation tree $T(n)$. The only assumption here is that all continuation decision trees are to be treated as decision trees in the domain of β. Moreover, if $\beta(T,\gamma,\cdot)$ represents intended behaviour, then $\beta(T(n),\gamma(n),n)$ is likely to be actual behaviour when decision node n is reached, because there is no good reason for actual behaviour to depart from intended behaviour at n in the continuation subtree $T(n)$. Then $\beta^*(T,\gamma,n) := \beta(T(n),\gamma(n),n)$ represents actual behaviour at every decision node n of T. The behaviour β^* is then automatically perfect or dynamically consistent.

Combined with weak consequentialism, continuation perfection imposes very strong restrictions indeed on the consequence choice function. Let us call a weakly consequentialist and continuation perfect behaviour norm just *consequentialist* for short. Then it is not too difficult to show that a consequentialist behaviour norm corresponds to an *ordinal* consequence choice function, in the sense that the preference relation R mentioned above must be an *ordering* – i.e., complete, reflexive and transitive – and also the choice function always selects the set of all those consequences which maximize the ordering R over the feasible set. That is, for every $Z \in \mathcal{F}(Y)$:

$$C(Z) = \{y \in Z \mid yRz \text{ for all } z \in Z\}$$

A direct proof that consequentialism is equivalent to ordinal choice of consequences is given in Hammond (1985b). An indirect proof, making use of Arrow's (1959) characterization of ordinal choice, can also be given, along the lines of Hammond (1977a).

One makes use of:

Proposition 2: (Arrow's characterization of ordinal choice – see

*C*4 (p. 123), Theorem 2 (p. 125), and Theorem 3 (p. 125) in Arrow (1959)).

The choice function $C: \mathscr{F}(Y) \to \mathscr{F}(Y)$ is ordinal if and only if, whenever Z, $Z' \in \mathrm{F}(Y)$ with $Z' \subset Z$, and $Z' \cap C(Z) \neq \emptyset$, then $C(Z') = Z' \cap C(Z)$.

To verify that this characterization is implied by consequentialism, construct the following decision tree T as in Figure 4.1 so that $\gamma[A(T)] = Z$ and $\gamma(n_1)[A(T(n_1))] = Z'$. Suppose $y \in Z' \cap C(Z)$. Then y is a possible consequence of behaviour in T. This implies that $n_1 \in \beta(T,\gamma,n_0)$ because otherwise one would have $C(Z) \subset Z \setminus Z'$. Then, however, it must be true that (with slight abuse of notation):

$$\beta(T,\gamma,n_1) = C(Z) \cap Z'$$

because $\beta(T,\gamma,n_1)$ has consequences in the subset of Z' consisting of possible consequences in the set $C(Z)$ consistent with behaviour passing through node n_1. Then continuation perfection or 'dynamic consistency' implies that:

$$\beta(T,\gamma,n_1) = \beta(T(n_1),\gamma(n_1),n_1)$$

(cf. Hammond (1976a; 1977a)). But weak consequentialism implies that:

$$\beta(T(n_1),\gamma(n_1),n_1) = C(Z')$$

(with the same slight abuse of notation as above). Therefore:

$$C(Z') = Z' \cap C(Z)$$

as required.

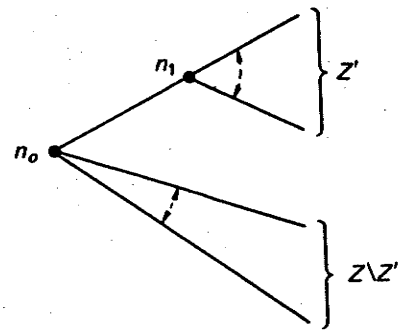

Figure 4.1 Decision tree T

It is also possible to verify by backward recursion that ordinal choice of consequences implies continuation perfect weakly consequentialist behaviour. Thus:

> *Proposition 3:* A behaviour norm is consequentialist (i.e., both continuation perfect and weakly consequentialist) if and only if it corresponds to an ordinal consequence choice function (i.e., one which maximizes a complete, reflexive and transitive preference ordering).

4 CONSEQUENTIALISM AND EXPECTED UTILITY

Up to now, the space of consequences Y has been entirely abstract, without any special structure at all. When there is risk, however, it is natural to make Y a space of probability measures. For behaviour in finite decision trees, it is enough to consider *simple* probability measures – i.e., probability measures which attach probability one to some finite set of sure consequences. Y will thus denote the set of sure or *pure* consequences, and the '*mixed*' *consequence* space will be $\mathscr{M}_0(Y)$ – or $\mathscr{M}$ for short – the set of all simple probability measures on Y. So, if $\mu \in \mathscr{M}$, then there exists a finite set $F \in F(Y)$ such that $\mu(F) = 1$. Also, for every $y \in Y$, there is some non-negative probability $\mu(\{y\})$ of y occurring, given the measure μ.

It has become common in economics to assume that behaviour under risk maximizes the expected value of some (von Neumann-Morgenstern (1944)) utility function which is unique up to an increasing linear transformation. The assumption is justified by appealing to a number of axioms – eight, for instance, in von Neumann and Morgenstern (1953, p. 26), although Samuelson (1952) backed up by Malinvaud (1952) noticed that an extra 'strong independence axiom' had slipped in implicitly. Marschak (1950), in the presentation which Arrow (1951c) chose in his survey, simplified the axioms greatly, and Herstein and Milnor (1953) were finally able to present just three simple axioms, of which their third is the independence axiom in a somewhat weakened form.

This independence axiom can be stated (in an unnecessarily strong form) as follows. Let λ, μ, $\nu \in \mathscr{M}$ be any three simple probability measures, and π any real number such that $0 < \pi \leqslant 1$. Let R denote the preference ordering on $\mathscr{M}$. Then:

$$[\pi\lambda + (1-\pi)\nu] R [\pi\mu + (1-\pi)\nu] \Leftrightarrow \lambda R \mu$$

Here $\pi\lambda+(1-\pi)\nu$, $\pi\mu+(1-\pi)\nu$ denotes the probability measures formed by multiplying every λ (resp. μ) probability by π and every ν probability by $(1-\pi)$ and adding. The result is a new simple probability measure in $\mathscr{M}$, as is easily verified, so that $\mathscr{M}$ is a 'mixture' set. The axiom is called the 'independence' axiom because the preference between $\pi\lambda+(1-\pi)\nu$ and $\pi\mu+(1-\pi)\nu$ is independent of ν.

I have already shown how consequentialism implies the existence of a preference ordering, which is Herstein and Milnor's first axiom. Hammond (1985b) also shows how it implies the above independence axiom as well, for reasons which perhaps Arrow (1951c, p. 425; 1971, pp. 27–8) came as close as anybody to seeing.

Consequentialism therefore implies two of Herstein and Milnor's (1953) three axioms. But only two: it is not difficult to exhibit consequentialist behaviour (based upon 'lexicographic' preferences) which violates their continuity axiom and is inconsistent with expected utility maximization.[1] In Hammond (1985b) an extra axiom of *weak continuity in probabilities* is added. It simply requires that behaviour in any decision tree T should vary continuously as the probabilities at its chance nodes vary. It leads to:

> *Proposition 4:* When the space of consequences consists of simple probability measures, a behaviour norm is consequentialist and weakly continuous in probabilities if and only if it corresponds to maximizing the expected value of a von Neumann-Morgenstern utility function defined on the space of pure consequences.

5 SOCIAL CHOICE OF INTERPERSONAL CONSEQUENCES

So far there has been no mention of whether the behaviour norms considered in Sections 3 and 4 are for individuals, or whether they are for societies on whose behalf public decisions are being made. This ambiguity was deliberate: the same arguments apply equally to both individual and to social norms and motivate expected utility maximization for individuals, or the maximization of an expected 'von Neumann-Morgenstern Bergson' social welfare function (NMBSWF) for a society.

What does distinguish individual norms from social norms, however, is the relevant space Y of pure consequences. For a society, I claim that the relevant Y space consists of pairs (θ^M, x^M) where M denotes the fixed set of members of the society, assumed to be finite,

$\theta^M = (\theta_i)_{i \in M}$ denotes the profile of individuals' personal characteristics, and $x^M = (x_i)_{i \in M}$ denotes the profile of 'individual consequences' in a sense to be explained below. Such a space will be called the space of *interpersonal consequences*.

The difference between θ_i and x_i, for an individual $i \in M$, is not very precise. Economists may want to think of θ_i as determining tastes and x_i as representing i's consumption – including consumption of personalized versions of public goods, of environmental externalities, and all other aspects of the public or social environment which economists are used to thinking of as arguments of an individual's utility function. However, θ_i must also include all other personal circumstances which help to determine, for example, how much income should be allocated to a θ_i-person. Complications arise because tastes and deserts, which are included in θ_i, tend to be closely related to some aspects of the social environment as well as to the individual's own consumption, both of which are included in x_i.

6 INDIVIDUAL NORMS AND ETHICAL LIBERALISM

One of the key ideas of Arrow (1951a; 1963) was to relate social choice to individual 'values'. Here, the comparable relation is between social norms and individual norms. I have just argued that the appropriate domain of pure consequences for social norms is the space $\Theta^M \times X^M$ of interpersonal consequences (θ^M, x^M). For individual i, I claim that the appropriate domain $\Theta_i \times X_i$ consists of all possible pairs (θ_i, x_i) of personal characteristics and personal consequences. This is really a matter of defining the sets Θ_i and X_i of possible θ_i and x_i broadly enough to include everything which is legitimately of consequence to individual i. In particular, all external effects should be included in x_i. i's individual norm is thus on the domain of pure consequences (θ_i, x_i).

Now it is natural to suppose that both social and individual norms are consequentialist on the space of simple probability measures over the domain of relevant pure consequences, and that they are also weakly continuous with respect to probabilities. By Proposition 4, this implies that both social and individual norms correspond to the expected value of a von Neumann-Morgenstern utility (or 'welfare') function on the domain of pure consequences. The social norm is thus represented on $\mathcal{M}_0(\Theta^M \times X^M)$ by:

$$\text{Max } Ew(\theta^M, x^M)$$

and each individual i's norm is represented on $\mathcal{M}_0(\Theta_i \times X_i)$ by:

$$\max Ev_i(\theta_i, x_i)$$

where E denotes 'expected value' in each case.

So far no link between the social and individual norms has been made. To establish one, consider the restricted domain of decision trees which lead to probability mixtures of pure interpersonal consequences of the form (θ^M, x^M) where, for some fixed $\bar{\theta}_j$, $\bar{x}_j$, one has $\theta_j = \bar{\theta}_j$, $x_j = \bar{x}_j$ (all $j \in M \setminus \{i\}$). Only i's personal characteristics and consequences are thus affected by decisions in these *personal decision trees*. In the literature on 'liberalism' or 'libertarianism' and rights, such decision trees are described as involving only 'personal issues'. A natural restriction then is that i's individual norm alone should determine all such personal issues, since i alone is affected. This is another assumption which I call *ethical liberalism*. It grants each individual the right to a personal identity, represented by θ_i, and the freedom to choose x_i whenever that freedom does not conflict with somebody else's freedom. Indeed, one can well argue that each individual's norm should be *defined* as the social norm for personal decision trees concerning just that individual, in which case ethical liberalism is certainly satisfied.

Notice that what counts as a personal consequence is determined normatively, and so therefore are the issues over which an individual has rights. This allows one to avoid the well-known liberal paradoxes due to Sen (1970b) and to Gibbard (1974), of which a recent discussion is by Coughlin (1984). Sen's (1970b) 'liberal paradox', based on the 'lewd', the 'prude', and one copy of the book *Lady Chatterley's Lover*, is thus resolved in one of two possible ways.

The first possible resolution is when both the lewd's claim to benefit from having the prude read the book, and the prude's claim to be harmed by having the lewd read the book, are dismissed as of no consequence, in the normative sense that such consequences should not count for ethical reasons. Then there is no conflict between the Pareto principle and each person exercising his freedom to choose whether or not to read the book. If either the lewd's claim to benefit, or the prude's claim to be harmed, are accepted as of consequence, however, the benefit and/or the harm are counted as externalities. Then whoever creates the externality loses his automatic right to choose whether or not to read the book. Of course, the freedom to choose may be valued for its own sake, in which case it becomes part

of the description of each consequence, along the lines discussed by Rowley and Peacock (1975).

Gibbard's (1974) paradox can be based on one nonconformist facing a conformist (Hammond 1982b), or more colourfully, on two girls each choosing to wear a red dress or a green dress, one wanting to be like the other and the other different (Sen 1976). Its resolution is similar to that of Sen's paradox. The desire to conform is either of no consequence normatively or else there is an externality. Similarly for the desire not to conform. So, if there are no externalities, there is no paradox, because normative preferences must be '(privately) unconditional' on the space of *relevant* consequences (cf. Gibbard 1974; Hammond 1982b). Alternatively, if there are externalities, then there are no automatic rights, and so again the paradox disappears.

Arrow himself has not ventured into the debate created by the Sen and the Gibbard liberal paradoxes, except in his review of Nozick (Arrow 1978). However, in his discussion (Arrow 1967) of one aspect of Brandt (1967), he concludes that 'strictly speaking, the preference scales that I deal with are those relating to personal welfare only'. In the formulation I am suggesting, individual i's 'personal welfare' must be represented by individual i's consequentialist norm on the space of mixed consequences $\mathscr{M}_i = \mathscr{M}_0(\Theta_i \times X_i)$, and this norm is in turn represented by the expected value of the von Neumann-Morgenstern utility function $v_i(\theta i, x_i)$. The essence of Arrow's approach to social choice, of course, is to make the social welfare ordering a function of individuals' 'preference scales'. Given that the social norm is consequentialist for the space of mixed interpersonal consequences $\mathscr{M}_0(\Theta^M \times X^M)$, and given weak continuity in probabilities, the social norm takes the form:

$$Ew(\theta^M, x^M) \equiv F(< Ev_i(\theta_i, x_i) >_{i \in M})$$

for some function F of individual expected utility levels, when this Arrow approach to social choice is extended to risky consequences. Such an approach is also consistent with the 'ex ante' notion of efficiency used in Arrow (1953; 1963).

In his work on social choice, Arrow (1963) has usually invoked only the weak Pareto principle, which requires F to increase when all individuals' expected utilities increase together. 'Ethical liberalism', however, as formulated above, implies that F must be strictly increasing in each variable separately, as with the strong Pareto principle. This may appear to be a minor difference. However, as will be seen in

Section 10 below, it does rule out dictatorial social norms which turn out to be the only ones which, for sufficiently general sets of personal characteristics, satisfy Arrow's independence of irrelevant alternatives condition as well as consequentialism, weak continuity, and the weak Pareto principle. So whether the difference is important depends upon one's attitude to independence of irrelevant alternatives (IIA). If this condition is important, ethical liberalism is indeed strong, because IIA is then inconsistent with the other axioms I have presented. But if only non-dictatorial social norms are of interest, then IIA has to be dispensed with anyway and ethical liberalism may not be unacceptably strong. This is the position I am going to take for the rest of the chapter, and I shall resume the discussion of IIA below.

7 AN 'IDEALIZED' VICKREY-HARSANYI UTILITARIANISM

When ethical liberalism is imposed, then as has been seen, there must exist a function F: $\mathbb{R}^M \times \mathbb{R}$ which is increasing in each of its $\#M$ arguments, such that:

$$Ew(\theta^M, x^M) \equiv F(< Ev_i(\theta_i, x_i) >_{i \in M})$$

Taking the case when there are pure consequences, it follows also that:

$$w(\theta^M, x^M) \equiv F(< v_i(\theta_i, x_i) >_{i \in M})$$

so that we may write $W \equiv F(v^M)$, where W is the von Neumann-Morgenstern measure of social welfare, and where $v^M \in \mathbb{R}^M$ is the M-vector of individual utilities. Then $EF\ (v^M) \equiv F(Ev^M)$ where $Ev^M \equiv < Ev_i >_{i \in M}$ is the M-vector of expected utilities. As was shown by Harsanyi (1955; 1977a, pp. 66–8), such a functional equation implies that F is linear. Therefore there exist constants ω_i, δ_i $(i \in M)$ such that:

$$F(v^M) \equiv \sum_{i \in M} \omega_i v_i + \delta_i$$

It follows that the von Neumann-Morgenstern social welfare function w must have the additive form:

$$w(\theta^M, x^M) \equiv \sum_{i \in M} [\omega_i v_i(\theta_i, x_i) + \delta_i]$$

Because of ethical liberalism, all the 'welfare weights' ω_i must actually be positive. Of course, this is the kind of welfare function originally presented by Vickrey (1945; 1960; 1961), Fleming (1952) and Harsanyi (1953; 1955). So:

> *Proposition 5:* Suppose that the social norm is both consequentialist and weakly continuous in probabilities on the space of simple probability measures over interpersonal consequences, and that each individual's norm has the same properties for personal consequences. Then ethical liberalism of the social norm implies that it corresponds to maximizing the expected value of a von Neumann-Morgenstern Bergson social welfare function which is the positively welfare weighted sum of von Neumann-Morgenstern individual welfare functions whose expected values represent each individual's norm.

Classical utilitarianism suggests maximizing the *unweighted* sum of individual 'utilities' or welfare functions. A similar form of welfare function can be derived here, simply by renormalizing each individual's welfare function. Indeed, define $\bar{v}_i(\theta_i,x_i) := \omega_i v_i(\theta_i,x_i) + \delta_i$ for all $i \in M$, $\theta_i \in \Theta_i$, $x_i \in X_i$. Because ω_i is positive, by ethical liberalism, $E\bar{v}_i$ represents i's individual norm just as well as Ev_i does. And, of course:

$$W(\theta^M,x^M) \equiv \sum_i \bar{v}_i(\theta_i,x_i)$$

so that:

> *Proposition 5':* Suppose that all the conditions of Proposition 5 are satisfied. Then the von Neumann-Morgenstern Bergson social welfare function is the *unweighted* sum of suitably normalized individual von Neumann-Morgenstern welfare functions.

This is a new version of classical utilitarianism, however, because of the interpretation of individual norms in the 'ethical liberalism' assumption. Each individual's norm is the social norm for personal decisions trees, and may be quite unrelated to the individual's happiness, preferences, or actual behaviour. In fact, I have presented a possible formalization of G. E. Moore's (1903) 'ideal utilitarianism' (cf. Fleming (1952) and Harsanyi (1955)). Indeed, the term 'liberalism' here may be somewhat inappropriate, in that individuals have

only the 'right' to behave according to the social norm defined on their personal issues. This is no 'right' at all, and can be very paternalistic, unless the social norm adapts to individual behaviour rather than the individual norm being tailored to fit the social norm. This issue will arise again in Sections 13 and 14 below.

Another difference from classical utilitarianism will be the determination of the welfare weights ω_i ($i \in M$) in Proposition 5, or equivalently, in determining the normalized utilities, of Proposition 5′. As Harsanyi (1955) and Pattanaik (1968; 1971) pointed out, the Vickrey and first Harsanyi (1953) approaches to utilitarianism do not resolve this issue entirely satisfactorily, because there is no reason to suppose that all individuals will agree on the weights, even if they do honestly place themselves in the 'original position' before choosing these weights. Whereas in classical utilitarianism – or what Sidgwick (1907) calls 'universalistic hedonism' – it seems to be presumed that 'total happiness' is an unambiguous quantity – see especially Edgeworth (1881).[2]

The approach to determining welfare weights which I shall present next also gives plenty of scope for individuals to disagree over what weights result. Moreover, there is even room for disagreement over an individual's welfare ordering. This is because the individual norms can be paternalistic, as I mentioned before, and so quite unrelated to individual behaviour. Individual norms are the subject of explicit value judgements, in fact.

8 THE FUNDAMENTAL INDIVIDUAL NORM AND HARSANYI'S UTILITARIANISM IN IDEALIZED FORM

One promising approach to the problem of determining welfare weights is to postulate the existence of 'fundamental preferences', to use a term due to Kolm (1972) – although he attributes the idea to Tinbergen (1957), and it can also be found in Harsanyi (1955). The idea is that a person is no more than a collection of personal characteristics, and that if characteristics are defined broadly enough, then all our preferences are the same. This combines ideas regarding the unimportance of personal identity which Arrow (1983a, p. 160)[3] finds in Pascal's *Pensées*, with the saying I have seen attributed to Confucius, 'All men are born equal: it's their habits that make them different.' But Arrow (1977a, p. 224; 1983a, p. 160) himself puts it very well for my purposes:

> Formally, we may suppose a space Y which defines the range of possible implications of a social state for an individual. Since the state defines, for every individual, everything that characterizes his satisfactions, the space Y is the same for all individuals. It includes goods, tastes, and the reactions of others to the extent that individuals care about one another. All individuals have the same preferences over Y. Let $u(y)$ be the ordinally defined utility indicator. Each state in X defines implications for every individual. Let $G_i(x)$ $(i=1,...,n)$ be for each individual a mapping from X to Y, expressing these implications. Then we can identify $u(x,i)=u[G_i(x)]$ . . .

In the consequentialist context with the individual norms which have been considered here, the space Y consists of all relevant individual consequences, including riskiness and attitudes to risk. All individuals share the same norm over decision trees with consequences in this space, and this must be representable by the expected value of a von Neumann-Morgenstern utility function $\bar{v}$ defined on Y. In my notation Arrow's space X becomes X_i, and the mapping G_i: $X_i \to Y$ can be regarded as the personal characteristic θ_i of individual i. Then, given the obvious extension to risky consequences, the von Neumann-Morgenstern utility function common to all individuals takes the form $v(\theta_i,x_i) \equiv \bar{v}(\theta_i(x_i))$. This function is unique up to transformations of the form:

$$\tilde{v}(\theta_i,x_i) \equiv \xi v(\theta_i,x_i) + \eta$$

where $\xi > 0$ and η are both independent of i and θ_i. So we have comparability of both utility levels and units for different individuals and different personal characteristics. The function v also allows welfare comparisons as tastes change, as considered in Harsanyi (1954). And the idea of using such fundamental preferences can be found in Harsanyi (1955; and especially 1977a, pp. 57–60; 1977b).

So, with fundamental preferences, there are common spaces Θ and X of personal characteristics and consequences respectively such that every individual i has a norm represented by expected values of the common von Neumann-Morgenstern individual welfare function $v(\theta,x)$ defined on $\Theta \times X$. Indeed, this is how the *fundamental individual norm* is defined.

Even with such a fundamental norm, the positive welfare weights in the welfare function:

$$w(\theta^M, x^M) \equiv \sum_i \omega_i v(\theta_i, x_i)$$

(where v now represents the fundamental norm) are still not determined. Now, however, it seems almost perverse not to make them all equal. Also, one can appeal directly to an *anonymity* axiom which would require the social norms to be identical in two societies whenever they differ only because individual consequences have been permutated. Formally, anonymity requires that whenever $\sigma: M \to M$ is a permutation, and $\tilde{\theta}_i = \theta_{\sigma(i)}$, $\tilde{x}_i = x_{\sigma(i)}$ (all $i \in M$, $x \in X^M$), then the two von Neumann-Morgenstern welfare functions:

$$w(\theta^M, x^M) \equiv \sum_i \omega_i v(\theta_i, x_i)$$

$$w(\tilde{\theta}^M, \tilde{x}^M) \equiv \sum_i \omega_i v(\theta_{\sigma(i)}, x_{\sigma(i)})$$

should represent identical social norms. This requires that, except for those unconcerned individuals with $v(\theta_i, x_i)$ constant for all $x_i \in X$, all individuals must share the same value of ω_i. For unconcerned individuals, too, it makes no difference what welfare weight ω_i they are given, so we may as well take $\omega_i = 1$ for all $i \in M$, or $\omega_i = 1/\#M$. The latter alternative recovers Harsanyi's (1953) welfare function precisely, with social welfare equal to the expected value of individual welfare when any individual has equal chances of becoming any one of the individuals in the society. A similar idea can be found in Lerner (1944). For the expected individual welfare from such a lottery, of course, is:

$$\sum_{i \in M} v(\theta_i, x_i)/\#M$$

Let us call this a *fundamentally utilitarian* social welfare function. To summarize:

> *Proposition 6:* Suppose that all the assumptions of Proposition 5 are satisfied, that there is a fundamental individual norm common to all individuals, and also that anonymity is satisfied. Then the von Neumann-Morgenstern Bergson social welfare function is the mean of all the individuals' welfares when these are measured according to the common von Neumann-Morgenstern individual welfare function representing the fundamental norm – i.e., it is fundamentally utilitarian.

9 INDEPENDENCE OF IRRELEVANT ALTERNATIVES AND DICTATORSHIP

A crucial and much discussed axiom in Arrow's theory of social choice is IIA. In Hammond (1977a; 1986) this condition is shown to be derivable from consequentialism under the assumption that, in any decision tree, the social norm depends only upon individual norms in decision trees with consequences which are possible in the original decision tree. Consequences which cannot result from acts in the original decision tree are treated as 'irrelevant alternatives'. Indeed, independence of such irrelevant alternatives, in the case of certainty, is part of the definition of an '(Arrow) consequentialist social norm' in the 1986 paper.

A weakened implication of IIA is that, in any society with fixed personal characteristics θ^M, the social norm depends only upon the individuals' *behaviour norms* which θ^M determines, and not upon the particular von Neumann-Morgenstern welfare functions used to represent them. Specifically, if ζ_i ($i \in M$) is any set of positive constants, then both $Ev_i(\theta_i, x_i)$ and $E\zeta_i v_i(\theta_i, x_i)$ represent i's behaviour norm in decision trees with pure consequences in the set $\{\theta_i\} \times X_i$ with the personal characteristic θ_i fixed. Suppose now that there exist $\bar{\theta}_i \in \Theta_i$ and constants φ_i ($i \in M$) such that:

$$v_i(\bar{\theta}_i, x_i) \equiv \zeta_i v_i(\theta_i, x_i) + \varphi_i \quad (\text{all } x_i \in X_i)$$

Then the individual norms for all decision trees with pure consequences in the set $\{\theta^M\} \times X^M$ are equivalent to those for all decision trees with pure consequences in the set $\{\bar{\theta}^M\} \times X^M$. So IIA, in combination with the conditions of Proposition 5 above, requires that both $E\Sigma_{i \in M} \omega_i v_i(\theta_i, x_i)$ and $E\sum_{i \in M} \omega_i v_i(\bar{\theta}_i, x_i)$ should represent the same social norm for decision trees with pure consequences in X^M. There must thus exist a positive constant $\zeta > 0$ and an arbitrary constant φ such that, on X^M:

$$\sum_{i \in M} \omega_i [\zeta_i v_i(\theta_i, x_i) + \varphi_i] \equiv \zeta \sum_{i \in M} \omega_i v_i(\theta_i, x_i) + \varphi$$

This requires that for all $i \in M$ (except unconcerned individuals for whom $v_i(\cdot, \cdot)$ is constant on $\{\theta_i\} \times X_i$ – these I ignore), either $\zeta_i = \zeta$ or $\omega_i = 0$ (or both). Thus, when the ζ_i ($i \in M$) above can be all different – as I shall assume – then there can be at most one individual, labelled $d = d(\theta^M)$, for whom $\omega_d > 0$. So:

$$w(\theta^M, x^M) \equiv \omega(\theta^M) v_d(\theta_d, x_d)$$

must be the form of the von Neumann-Morgenstern social welfare function. Since at most one individual's welfare weight can be positive for any fixed θ^M, ethical liberalism is violated. The Pareto principle, however, can be satisfied. It requires $\omega(\theta^M)$ to be positive. Then $d(\theta^M)$ is the *dictator* in the society with personal characteristics θ^M, in the sense that this individual's norm coincides with the social norm for decision trees with pure consequences in the set $\{\theta^M\} \times X^M$. Moreover, the dictator d must be the same for all possible θ^M to ensure that the Pareto principle is satisfied. For if there exist θ_1^M and θ_2^M satisfying:

$$d(\theta_1^M) = d' \neq d'' = d(\theta_2^M)$$

and

$$w(\theta_1^M, x^M) \equiv \omega_1 v_{d'}(\theta_{d'}, x_{d'})$$

$$w(\theta_2^M, x^M) \equiv \omega_2 v_{d''}(\theta_{d''}, x_{d''})$$

then one can find probabilities π_1, π_2 of having θ_1^M, θ_2^M respectively (with $\pi_1 + \pi_2 = 1$), and a variation in x^M which increases $\pi_1 v_i(\theta_{1i}, x_i) + \pi_2 v_i(\theta_{2i}, x_i)$ for every $i \in M$, and yet decreases $\pi_1 \omega_1 v_{d'}(\theta_{d'}, x_{d'}) + \pi_2 \omega_2 v_{d''}(\theta_{d''}, x_{d''})$. This contradicts the Pareto principle. So the von Neumann-Morgenstern social welfare function takes the form:

$$w(\theta^M, x^M) = v_d(\theta_d, x_d)$$

for some dictator d who is independent of θ^M. (The positive multiplicative constant $\omega(\theta^M)$ must also be independent of θ^M, and so can be ignored.) Thus:

> *Proposition 7* (cf. Arrow's impossibility theorem): Suppose that both the social norm and each individual's norm are consequentialist and weakly continuous in probabilities. Suppose, too, that the social norm for consequences which all include any fixed profile θ^M of personal characteristics depends only upon individual behaviour norms for the fixed personal characteristics θ_i ($i \in M$), and that the Pareto condition is satisfied. Suppose also that, for all $\theta^M \in \Theta^M$, there exist $\bar{\theta}^M \in \Theta^M$, all different positive constants ζ_i ($i \in M$), and additive constants φ_i ($i \in M$) such that:

$$v_i(\bar{\theta}_i,x_i) \equiv \zeta_i v_i(\theta_i,x_i) + \varphi_i$$

for all $x_i \in X_i$ and all $i \in M$. Then there is a dictator whose individual norm completely determines the social norm, and indeed the social norm simply takes account of the dictator's personal characteristics and consequences and then coincides with the dictator's individual norm.

Thus, even in the weakened form which I have been using in this section, independence of irrelevant alternatives (IIA) indeed seems to be the chief culprit in Arrow's impossibility theorem, as Arrow himself has often asserted. Notice, however, that IIA does *not* rule out the *existence* of preference intensities of cardinal utilities when the domain includes risky consequences and when preferences satisfy both independence and weak continuity. The dictator's von Neumann-Morgenstern welfare function does depend upon such preference intensities. The *construction* of these relative preference intensities may involve irrelevant alternatives, but IIA is still satisfied.

10 INDEPENDENCE OF ETHICALLY IRRELEVANT CONSEQUENCES

We have just seen that the social welfare function:

$$\sum_{i \in M} v(\theta_i,x_i)/\#M$$

of Section 8 and Proposition 6 violates independence of irrelevant alternatives (IIA). Indeed, the social norm in any decision tree with θ^M fixed does not just depend upon each individual i's norm with θ_i fixed, which is rather a strong violation of IIA.

One way of circumventing this dependence on irrelevant alternatives would be to postulate that, for all ethical decisions, there must be an original position in which each individual's personal characteristic is still entirely unknown. Lerner (1944) and Harsanyi (1953; 1955; 1977a; 1977b) consider what happens when each individual i has an equal chance of having any other individual j's personal characteristic and consequence (θ_j,x_j), and derive the social welfare function of Proposition 6. Rawls (1957; 1958; 1971; 1972) postulates maximin behaviour in an original position 'behind the veil of ignorance', which leads to a social welfare function of the form $\min_{i \in M} u_i(\theta_i,x_i)$ for some

set of individual utility functions u_i whose levels are interpersonally comparable. Such a function violates consequentialism and its implications for 'rationality' when there is risk, as Harsanyi (1975a; 1975b; 1977b) has emphasized, but that is not the main point I want to make here.

Original positions avoid the dependence on irrelevant alternatives pointed out above because they exclude from consideration decision trees in which θ^M is fixed. The only decision trees one is allowed to consider, in effect, commence with a chance node which determines who has which personal characteristic. Later decisions then determine individual consequences. Original positions, however, are never a part of any actual decision tree, so to insist on introducing one artificially is to be avoided if possible.

An alternative to the original position is to question whether consequences in which individuals have changed places really are irrelevant, even in decision trees with θ^M fixed. They may not be feasible consequences in the existing tree, but they may still be ethically relevant. Hare (1953; 1963; 1981) argues that ethics is about universalizable prescriptive statements, using 'universalizable' in a sense which most philosophers recognize from the works of Kant. One kind of universalizability might apply to consequences, and require that whenever (θ^M, x^M) is a relevant pure consequence which can occur in a given decision tree, and whenever σ: $M \to M$ is a permutation of individuals, then the consequence $(\tilde{\theta}^M, \tilde{x}^M)$ with:

$$\tilde{\theta}_i = \theta_{\sigma(i)}, \tilde{x}_i = x_{\sigma(i)} \quad \text{(all } i \in M\text{)}$$

is also *ethically relevant*. In other words, consideration must be given to interpersonal consequences which arise when individuals change their individual consequences (including personal characteristics).

A consequentialist social norm will then be described as *independent of ethically irrelevant mixed consequences* if, in any decision tree with θ^M fixed, the behaviour it prescribes is determined as a function only of the behaviour prescribed by the individuals' norms applied to all decision trees with mixed ethically relevant consequences. In particular, the social norm in a given tree is unaffected if the individual norms change in a way that affects prescribed behaviour only in decision trees with ethically irrelevant consequences.

This significant weakening of the usual IIA condition is satisfied by the fundamental utilitarian social norm which maximizes expected total individual welfare when there is a fundamental individual norm,

as in Section 8. For the individual welfare function $v(\theta,x)$ on $\Theta \times X$ can only be transformed to a new welfare function satisying:

$$\tilde{v}(\theta,x) \equiv \zeta v(\theta,x) + \psi \quad \text{(for some constants } \zeta > 0 \text{ and } \psi)$$

whenever $(\theta,x) \in \{(\theta_i,x_i) \mid i \in M\}$ for some relevant interpersonal consequence (θ^M,x^M). Otherwise the individual norm will change in some decision tree with mixed ethically relevant consequences. Then $\Sigma_{i \in M} \tilde{v}(\theta_i,x_i)$ and $\Sigma_{i \in M} v(\theta_i,x_i)$ are (cardinally) equivalent von Neumann-Morgenstern social welfare functions, so the social norm is indeed independent of ethically irrelevant mixed consequences.

As mentioned in the introduction, weakening IIA to independence of ethically irrelevant mixed consequences is *the* crucial step of my attempt to reconcile Arrow's social choice theory with Harsanyi's utilitarianism. I certainly intend this crucial step to be discussed elsewhere. For the moment, however, let me quote one of Arrow's (1973a; 1974, p. 270) more recent defences of IIA, culminating in a rhetorical question:

> To see what is at stake, suppose that a society has to make a choice among some alternatives and does so. After the decision is made, an alternative which has not previously been thought of is mentioned as a logical possibility, although it is still not feasible. The individuals can expand their preference orderings to place this new alternative in its place on their ranking; but should this preference information about an alternative which could not be chosen in any case affect the previous decision?

One difference between Arrow's IIA condition and my suggested weakening arises when the 'new alternative' involves individuals exchanging their personal characteristics and personal consequences. IIA claims that all preference information about such an alternative is irrelevant, whereas I claim that universalizability implies that such preference information is ethically relevant. This concept of ethical relevance is related to but rather different from Strasnick's (1977) 'moral relevance'. A second obvious difference is that I allow dependence on behaviour in decision trees with *mixed* consequences, even if there is no risk in the actual decision tree. The latter appears to be crucial: if the social behaviour norm in a decision tree without risk is allowed to depend only on individuals' behaviour norms in decision trees with *pure* consequences, it seems that dictatorship is once again

hard to avoid (unless weak continuity with respect to probabilities is relaxed – this remains to be investigated). What Arrow (1973c; 1983a, p. 142) calls:

> the somewhat peculiar property that social choices among decisions where there may be no uncertainty are governed by attitudes toward risk bearing

thus appears inescapable if dictatorship is to be avoided.[4]

It has yet to be shown what are the other implications of the new independence condition, although it is closely related to the 'independence' condition of d'Aspremont and Gevers (1977) which might better be called 'independence of irrelevant utilities' (see also Arrow 1977b, p. 621; 1983a, p. 173)). And its ethical appeal requires much closer examination, since it is explicitly an ethical condition motivated by universalizability. It looks as though it may be less objectionable than an original position, however.

11 DOMAIN RESTRICTIONS IMPLIED BY ETHICAL LIBERALISM AND THE FUNDAMENTAL INDIVIDUAL NORM

In discussing the version of Harsanyi's fundamental utilitarianism presented in Section 8, I have focused on Arrow's independence condition because it is violated in its original form and because it has indeed been the most controversial. It should be noted that, of the other conditions of Arrow's impossibility theorem, the Pareto rule is obviously satisfied – actually, in its strict form – and also there is no dictator. Indeed, anonymity was used to derive fundamental utilitarianism from the fundamental individual norm and liberalism. That leaves only the unrestricted domain condition, to which I turn next.

There are in fact two ways in which the domain of permissible individual norms is restricted, yet I shall argue that both restrictions are natural in the context of fundamental utilitarianism and so the unrestricted domain condition survives in spirit, if not in form.

The first restriction arises because of ethical liberalism and the definition of interpersonal consequences and interpersonal norms, back in Sections 5 and 6. Indeed, individual i's norm *is* the social norm for personal decision trees in which only consequences in the space $\Theta_i \times X_i$ are affected by behaviour. Individual i's possible norms are

thus restricted in the sense that they cannot by themselves prescribe behaviour in decision trees where only consequences outside this space are affected. The corresponding von Neumann-Morgenstern welfare function $v_i(\theta_i,x_i)$ is independent of θ_j and x_j for all $j \neq i$. This is a domain restriction. But since Θ_i and X_i are defined to include all consequences of relevance to individual i, it is a harmless restriction. Without it, in fact, individual norms lose much of their relevance. Perhaps one could instead postulate a fundamental individual norm for 'clone' societies in which all individuals always share the same characteristic (as in Hammond 1984) and have the same personal consequences as well. But I leave this for later: it seems less appealing to me than does ethical liberalism.

In addition, there are advantages to not having to postulate an unrestricted domain of individual norms, but having these norms restricted to individual consequences only. Arrow's impossibility theorem and its many successors require a domain which is sufficiently unrestricted to allow certain preference profiles to be constructed during the course of their proof. In 'economic' domains, however, with some private goods, and with preferences restricted so that each individual is unconcerned about the consumption of other individuals' private goods, these profiles are hard to construct and proofs have to be adapted – see, for instance, Border (1983). Whereas I have constructed the space of interpersonal consequences and individual norms so that all the economic domain restrictions are automatically satisfied. All the propositions of this chapter thus apply straight away to economic domains.

The second violation of the unrestricted domain condition arises because there is a fundamental individual norm, for decision trees with mixed consequences in the space $\Theta \times X$, which has to be the same for all individuals. But, in Arrow's original sense, there can be an unrestricted domain in that there can be unrestricted individual norms for each individual $i \in M$ on the mixed consequences in $X_i (= X)$, provided only that these norms are consequentialist and weakly continuous in probabilities. For, given any such norm β_i for individual i, one can ensure that the space of personal characteristics Θ is expanded to include some θ such that the expected value of $v(\theta,\cdot)$ represents β_i on $M_0(X_i)$. With this extension of personal characteristics, then, there is a sense in which the unrestricted domain condition is also satisfied.

So, by relaxing independence of irrelevant alternatives to independence of ethically irrelevant mixed consequences, Arrow's impossibi-

lity theorem is turned into a possibility result, provided that in addition the unrestricted domain condition is modified to admit ethical liberalism (with each individual i concerned only about consequences in $\Theta_i \times X_i$), and to admit the existence of a fundamental individual norm. However, no restriction (other than consequentialism and weak continuity) is placed on individual i's norm in personal decision trees in which individual i's personal characteristic is fixed, and the consequences for individuals other than i are not influenced by behaviour. So there is essentially an unrestricted domain in the sense used in Arrow (1950; 1951a) (which never considered choices of personal characteristics) except that, of course, individual's norms are limited to their own characteristics and consequences.

12 ETHICAL DICTATORSHIP OF THE FUNDAMENTAL INDIVIDUAL NORM

In Section 8 there was presumed to be an unambiguous 'fundamental individual norm' on which to base fundamental utilitarianism. Yet, even if there was agreement that *some* fundamental individual norm should be used, it might be impossible to agree *which* norm was the right one. In Section 8 it was suggested, following Arrow (1977), that one might construct a space of individual 'Gorman-Lancaster characteristic' consequences so broadly that all could agree what the norm should be for these consequences. And that a personal characteristic θ could be defined as a mapping from the more conventional space of consequences to this space of Gorman-Lancaster characteristics.

Earlier Arrow (1951a; 1963, Chapter 7) himself suggested that if 'individual values' were considered in a suitable context, one might obtain sufficient agreement to restrict the domain of possible individual values and so to allow the existence of a non-dictatorial social welfare function. The fundamental norm is perhaps as likely as any other form of 'individual values' to permit such agreement, especially if unanimity is being sought. Yet it is far from clear that the Gorman-Lancaster characteristic space could be defined broadly enough to allow 'universal' common preferences to be constructed upon it in a way which everybody could agree to. Or that, if such a commonly agreed construction were possible, there could be agreement about how actual individuals' ordinary consequences should be mapped via their personal characteristics into this Gorman-Lancaster characteristic space. This is especially true when one considers that there must be

agreement about the fundamental individual norm to be used for risky consequences, and so about the proper attitude to risk (cf. Pattanaik, 1968; 1971). So one asks what can be done in the absence of agreement over the fundamental individual norm, or over the two ingredients used in the Arrow approach to constructing it (as adapted to deal with mixed consequences).

Suppose, then, that different ethical views are parametrized by η in the space H. For each $\eta \in H$, all the assumptions of Proposition 6 of Section 8 are assumed to be satisfied. So there is a fundamental individual norm for decision trees with mixed consequences in the space $\mathscr{M}_0(\Theta \times X)$, and this norm corresponds to the expected value of the von Neumann-Morgenstern utility function $v(\theta,x;\eta)$ defined on $\Theta \times X$. The views of the various members of society are given by the profile $\eta^M = (\eta_i)_{i \in M}$. The NMBSWF of the society with ethical views η^M takes the form:

$$w(\theta^M,x^M;\eta^M) \equiv \sum_{i \in M} v^*(\theta_i,x_i;\eta^M)$$

as in Section 7, for suitable normalized individual welfare functions $v^*(\cdot,\cdot;\eta^M)$ defined on $\Theta \times X$.

Now an additional assumption is made, related to ethical liberalism as considered in Section 6, but extended to allow for different ethical views. Suppose, that, for a personal decision tree concerning some individual $i \in M$, the individual norm for i on the space of relevant mixed consequences is the same according to all ethical views η_j ($j \in M$). Then *extended ethical liberalism* requires that this common individual norm should be the social norm, too. In other words, when $Ev(\theta_i,x_i;\eta_j)$ ($j \in M$) represents the same norm for a set of mixed consequences in $\mathscr{M}_0(\Theta \times X)$, then each expected welfare function also represents the social norm in any personal decision tree concerning individual i alone.

Arguing as in Section 7, it then follows that for each fixed $\eta^M \in H^M$, there is a function $F(\eta^M;.)$ defined on the space $IR^{M \times M}$ of expected utility vectors such that:

$$Ew(\theta^M,x^M;\eta^M) \equiv F(\eta^M;\ \ll Ev(\theta_i,\ x_i;\ \eta_j) \gg_{ij \in M})$$

and

$$w(\theta^M,x^M;\eta^M) \equiv F(\eta^M;\ \ll v(\theta_i,\ x_i;\ \eta_j) \gg_{ij \in M})$$

Thus, using Harsanyi's lemma (Section 7) again, it follows that $(F\eta^M;.)$ must be linear, and there exist weights $\omega_{ij}(\eta^M)$ $(i,j \in M)$ such that (after ignoring irrelevant additive constants):

$$w(\theta^M, x^M; \eta^M) \equiv \sum_{i \in M} \sum_{j \in M} \omega_{ij}(\eta^M) v(\theta_i, x_i; \eta_j)$$

The weights $\omega_{ij}(\eta^M)$ must all be non-negative because of extended ethical liberalism. Recall that each individual's ethical views are assumed to satisfy all the conditions of Proposition 6, including anonymity. So $\omega_{ij}(\eta^M) = \bar{\omega}_j(\eta^M)$, independent of i, for all possible j *and* η^M. Unlike in Section 7, however, not all the weights need be positive; instead, it is enough that there should exist at least one $\in M$ for which $\bar{\omega}_j(\eta^M)$ is positive. So:

$$v^*(\theta, x; \eta^M) = \sum_{j \in M} \bar{\omega}_j(\eta^M) v(\theta, x; \eta_j)$$

The classical problem of social choice theory reappears; the individuals' ethical views, which are orderings on $\mathscr{M}(\Theta \times X)$ represented by $Ev(\theta, x; \eta_j)$ $(j \in M)$, must be aggregated into a social ordering on $\mathscr{M}$ z)$\Theta \times X)$ represented by $Ev^*(\theta, x; \eta^M)$. The only difference from the original Arrow problem is that the utility functions $v(\theta, x; \eta_j)$ $(j \in M)$ are cardinal, but this case is dealt with in Sen (1970a). As he shows, Arrow's 'impossibility' theorem remains true, and there will be a dictator, even with cardinal utility functions, if: (1) there is an unrestricted domain of cardinal utility functions; (2) there is independence of irrelevant alternatives; (3) the Pareto condition is satisfied. In the framework of this section, independence of irrelevant alternatives means that, given any subset Z of $\Theta \times X$, the ordering induced on Z by $v^*(\cdot, \cdot; \eta^M)$ should depend only on the orderings induced on Z by $v(\cdot, \cdot; \eta_j)$ $(j \in M)$. This is clearly true if one extends independence of ethically irrelevant mixed consequences in an obvious way to allow individuals' different conceptions of the fundamental norm to vary. The Pareto condition is also satisfied because each $\bar{\omega}_j(\eta^M)$ is non-negative. So one only needs to add the unrestricted domain assumption. Then Sen's result tells us that the ordering induced by $v^*(\theta, x; \eta^M)$ depends only upon one utility function $v(\theta, x; \eta_d)$ representing the ethical views of individual d. So d's ethical views alone determine the social norm *completely*, and he is *fully* an 'ethical dictator'.

Proposition 8: Suppose that each individual $j \in M$ has ethical views consistent with fundamental utilitarianism, including anonymity,

as described in Proposition 6 of Section 8. Suppose that the social norm satisfies extended ethical liberalism, and the extended form of independence of ethically irrelevant mixed consequences. Suppose finally that, given any von Neumann-Morgenstern individual welfare function $\bar{v}$ representing a view of the fundamental individual norm, defined on the space of pure consequences $\Theta \times X$, there exists $\eta \in H$ such that $v(\theta, x; \bar{\eta})$ is cardinally equivalent to $\bar{v}$ on $\Theta \times X$.

Then there exists an 'ethical dictator' $d \in M$ whose fundamental individual norm $v^d(\theta,x)$ completely determines the social norm; indeed, the social norm corresponds to expected values of the NMBSWF:

$$w^M(\theta^M,x^M) \equiv \sum_{i \in M} v^d(\theta_i,x_i)$$

with θ^M fixed and x^M variable.

So, in effect, the fundamental individual norm used in a fundamentally utilitarian NMBSWF has to be dictated. The need for some kind of dictatorship, despite the presence of interpersonal comparisons of utility, was foreseen by Arrow during our respective oral discussions of Phelps (1977) at the World Congress of the Econometric Society at Toronto in 1975. However, this is a much weaker kind of dictatorship than that in Arrow's impossibility theorem – that reconsidered in Section 10 above, for instance. In the original dictatorship, only the welfare of the dictator determines the social norm, and there is no anonymity or universalizability of any form whatsoever. Here, on the other hand, there is a kind of anonymity or universalizability; the dictator's *ethical views* determine the social norm, but these views must be anonymous or universalizable and correspond to some fundamental individual norm that takes into account all possible types of individual. That is why I have called it an 'ethical dictatorship'.

13 CONSUMER SOVEREIGNTY, ATTITUDES TO RISK, AND THE DISTRIBUTION OF INCOME

In Section 8 it was seen that fundamental utilitarianism leads to an NMBSWF of the form:

$$w(\theta^M,x^M) \equiv \sum_{i \in M} v(\theta_i,x_i)$$

where v is the von Neumann-Morgenstern individual welfare function whose expected value represents the fundamental individual norm. In this section, some implications of using such a norm are considered in the special case where there is only a single good called 'real income' which is of consequence to any individual. Thus x_i can be replaced by y_i, the real income of individual i.

Assume that a θ-individual's risk-taking behaviour corresponds to maximizing the expected value $Eu(y;\theta)$ of the von Neumann-Morgenstern utility function $u(\cdot;\theta)$. Consumer sovereignty, and also liberalism in the usual sense, both require the NMBSWF to respect each θ-individual's behaviour, as represented by $Eu(y;\theta)$. As in Section 7, it follows that the NMBSWF must be equivalent to:

$$\tilde{w}(\theta^M, y^M) \equiv \sum_{i\in M} \omega_i(\theta^M) u(y_i;\theta_i)$$

in expected values, for some positive welfare weights ω_i which can depend on θ^M. $Eu(y_i;\theta_i)$ and $Ev(\theta_i,y_i)$ must thus be equivalent as functions on probability mixtures of real income y_i, for every $i\in M$ and every $\theta_i\in\Theta$. So $Ev(\theta,y)$, as well as $Eu(y;\theta)$, must represent a θ-individual's risk-taking behaviour.

This is what lies behind Arrow's (1952; 1983a, p. 48) remark that the distribution of income is determined 'by the tastes of individuals for gambling'. One may quibble that gambling is enjoyed for its own sake as much as for the probability distribution of real income that it generates, in which case $Eu(y;\theta)$ on its own does not determine gambling behaviour. The consequence space needs to be expanded to allow for the excitement, thrills, and other non-pecuniary effects of gambling. Nevertheless, this is just a quibble, and the point that the norm depends upon individuals' attitudes to risk in a way that may not be ethically acceptable is well taken. I have discussed it at some length in Sections 5.4 and 6 in Hammond (1983a).

To avoid what may be an ethically unacceptable dependence of the social norm on individuals' risk-taking behaviour, the only escape is to relax consumer sovereignty. After all, consumer sovereignty *is* a value judgement, just as the construction of *any* fundamental individual norm must be. The function $\Sigma_{i\in M} v(\theta_i,y_i)$ represents the social norm when the pure consequences are income distributions y^M together with personal characteristic θ^M. Also, $Ev(\theta,y)$ represents the fundamental individual norm and, in particular, enshrines the ethical attitude to real income risk of a θ-individual. When consumer sovereignty is relaxed, $Eu(y;\theta)$, which describes behaviour, is given no

ethical significance at all, except in so far as actual risk-taking behaviour helps to determine an individual's personal characteristic.

When $Ev(\theta,y)$ is imposed as the norm of a θ-individual, rather than $Eu(y;\theta)$, then not only is citizen's sovereignty abandoned; in addition, 'liberalism' in the usual sense is clearly violated, in that appropriate interference with individuals' risk-taking behaviour is viewed as desirable when it is possible. This is a natural implication of consequentialism, however. Indeed, it has frequently been asserted that consequentialism necessarily ignores rights and recommends policies which violate them. This is not entirely accurate because, if rights are of ethical significance, then they should be included in the descriptions of consequences together with measures of the extent to which rights are violated. As I claimed at the end of Section 2, 'the space of consequences should include everything of consequence' – including, of course, rights, if these really do matter. So, if liberalism is violated in an important sense by imposing the norm $Ev(\theta,y)$ for a θ-individual's risk-taking, the fault lies not with consequentialism or with fundamental utilitarianism, but rather with too narrow a description of ethically relevant consequences. Indeed, consequentialism is even consistent with extreme libertarianism, in which the *only* relevant consequences are violations of individual rights, and such violations are to be minimized.

The idealized form of Vickrey-Harsanyi utilitarianism considered in Sections 7 and 8 thus does *not* entail having income distribution determined by risk-taking behaviour, let alone gambling behaviour. The argument that it does rests on the implicit assumption of consumer sovereignty. In this section, I have argued that consumer sovereignty should be abandoned rather than the use of 'idealized' utilities based upon behaviour norms for risky consequences. And also that this is not necessarily inconsistent with a respect for individual rights, despite frequent assertions to the contrary. Rather, where rights matter, they can be included among consequences, as Harsanyi himself (1977b) has for one suggested.

14 RISK-TAKING BEHAVIOUR AND THE BASIC THEOREMS OF CLASSICAL WELFARE ECONOMICS

So, if the distribution of income is not to be determined by individuals' risk-taking behaviour, then consumer sovereignty needs to be relaxed. This has important implications for the ethical significance of

the 'basic theorems of classical welfare economics', as presented by Arrow (1951b) and Debreu (1954) and extended to deal with risk by Arrow (1953; 1963), Debreu (1959) and Guesnerie and Jaffray (1974). For if individuals' risk-taking market behaviour diverges from the fundamental individual norm, one loses the link between competitive markets and the appropriate ethical concept of Pareto efficiency.

First, however, let me discuss the ethical significance of the two 'basic' theorems of 'classical' welfare economics, which I actually prefer to call the two *efficiency theorems*. The first efficiency theorem holds whenever individuals have locally non-satiated preferences and states that then any Walrasian equilibrium in complete competitive markets (without externalities) must be Pareto efficient. By itself, however, this first theorem has little ethical significance because Pareto efficiency is a very limited ethical criterion in any economy with diverse individuals. Pareto efficiency fails to exclude dictatorship or extremes of poverty and inequality. Indeed, it even fails to ensure that all healthy individuals receive enough food to survive (see Coles and Hammond 1986). So, while many writers have often used the term 'Pareto optimal', there is nothing very 'optimal' about some of the allocations, which I therefore prefer to call 'Pareto efficient'. In addition, 'Pareto efficiency' requires maximizing a vector or quasi-ordering, just as production efficiency does.

The second efficiency theorem has in any case more ethical significance. It requires assumptions such as convexity of feasible sets, continuity of preferences, and also convexity of preferences (or alternatively a 'continuum' of individuals in the sense of Aumann (1964; 1966)). Let us say that an allocation is *interior* when each consumer's allocation is interior to his feasible set, so ruling out the 'exceptional case' first noticed in Arrow (1951b, p. 528; 1983b, p. 39). Then, under the assumptions stated above, any interior Pareto efficient allocation is a Walrasian equilibrium in complete competitive markets *provided* that suitable lump-sum transfers are made which enable each consumer to afford his allocation at market clearing prices. That is the second efficiency theorem.

Suppose there is a Bergson social welfare function which is 'Paretian' in the sense that social welfare increases as any one individual's utility increases while the others' utilities all remain constant. Then any welfare optimal allocation must be Pareto efficient, so the second efficiency theorem applies. Complete markets can be used to decentralize the welfare optimal allocation. There are several crucial provisos, however, before this second efficiency theorem can be used to justify

complete markets on ethical grounds. Obviously the continuity, convexity and interiority assumptions have to be met, and the Arrow exceptional case has to be avoided; indeed, unless these assumptions are satisfied, there is no guarantee of existence of Walrasian equilibrium even if all consumers are initially endowed with claims to the goods they are due to receive in the welfare optimal allocation. Nor will complete markets achieve the welfare optimum if there are impediments to reaching the appropriate Walrasian equilibrium in those markets. Also very important is the need for suitable lump-sum transfers to be made. Lerner (1944) and Graaff (1957) realized that the information needed to carry out such transfers is all too rarely available. Indeed, as I have tried to explain in the past (Hammond 1979; 1983b; 1985a), individuals have every incentive to distort the information required to implement such transfers (cf. Arrow 1978; 1983a, p. 184).

An even more important and fundamental proviso is all too rarely discussed. It is supposed that the Bergson social welfare function is increasing in individuals' utilities for *utility functions which correspond to market behaviour*. This presumes consumer sovereignty, or an orthodox form of liberalism which is perhaps better called libertarianism. As Arrow (1973b; 1983a, p. 97) puts it:

> philosophers have been more prone to analyze what individuals should want, where economists have been content to identify 'should' with 'is' for the individual (not for society).

Both *ethical* liberalism, however, and fundamental utilitarianism, make no such presumption. Indeed, Section 13 argued that there were good reasons for abandoning the sovereignty of consumers' risk-taking behaviour. An even stronger case arises when consumers are misinformed or underinformed, though I lack the space to discuss this properly because it would require considering asymmetric information.

Once consumers' sovereignty is relaxed, the possibility arises that some consumers may be making 'bad' market decisions – in the sense that they go against their own 'interests' as represented by the fundamental individual norm. Then there is no way that the second efficiency theorem can be used to justify complete markets. There may be alternative ethical justifications. For libertarians 'free markets' are bound to be optimal because any other economic system infringes

individuals' rights, which to them are the only relevant consequences. A better ethical justification of free markets may be to show that all attempts to interfere in their operation are self-defeating, as they will be if, for instance, all commodities can be exchanged freely in an 'underground' economy which cannot be monitored at all (see Hammond 1979, Section 7; 1983b). This is an extreme case, however. Neither it nor the libertarian argument have anything whatsoever to do with the second efficiency theorem – or, indeed, with the first.

Generally, then, relaxing consumer sovereignty completely undermines the ethical significance of the second efficiency theorem of welfare economics. The first efficiency theorem, I argued, never has ethical significance anyway (unless all consumers happen to be identical). The violations of consumer sovereignty considered in Section 13, however, were not general; they concerned risk-taking behaviour specifically. At first sight, inappropriate risk-taking may seem to be a special violation of consumer sovereignty which is rather easily amenable to corrective treatment. As has been seen, lump-sum transfers are almost always required to bring about a welfare optimum through complete markets. In an intertemporal economy with risk, what scope is there for 'complete' contingent lump-sum transfers at each date-event pair to enforce an optimal allocation of risk-bearing, despite the inappropriate market behaviour of individuals? After all, in any exchange economy which happens to have only one physical commodity, contigent lump-sum transfers at each date-event pair suffice to bring about *any* particular desired feasible allocation, regardless of individuals' security market transactions, because such transactions can always be completely offset. So an optimal allocation especially can be brought about. The divergence between the fundamental individual norm and individuals' market behaviour is thus of no consequence when there is only one physical commodity; corrective treatment of individuals' inappropriate behaviour is indeed fairly straightforward.

With many physical commodities, however, or with private production, corrective treatment is no longer always possible, as shown in Hammond (1981). In the terminology of that paper, a welfare optimum according to the fundamentally utilitarian NMBSWF is an "ex-post" welfare optimum. When consumer sovereignty is not satisfied, the condition for a combination of complete spot markets, contingent lump-sum income transfers, and arbitrary security markets (which are actually irrelevant) to bring about such an optimum are quite stringent. Consumers' von Neumann-Morgenstern utility

functions, as functions of the time stream $(x_t)_{t=1}^T$ of net demand vectors, must take the special 'backwardly separable' form:

$$u((x_t)_{t=1}^T) \equiv \varphi_T(\varphi_{T-1}(\varphi_{T-2}(\ldots(\varphi_3(\varphi_2(\varphi_1(x_1),x_2),x_3),\ldots,x_{T-2})x_{T-1})x_T)$$

and the fundamental individual norm $v((\varphi_t,x_t)_{t=1}^T)$ must be ordinally equivalent to this function. Producers' production sets must be in the form of a Cartesian product, so that production decisions at different times are entirely independent of each other. There can be no capital or durable goods, in fact. A special case considered by Starr (1973) is when u is additively separable:

$$u((x_t)_{t=1}^T) = \sum_{t=1}^T \varphi_t(x_t)$$

and production sets are restricted as before.

In particular, corrective treatment is almost always possible only if there is no gain in welfare at the optimal allocation when more information becomes publicly available. This should not be too surprising: if there is to be no loss of welfare when individuals are poorly informed, there can be no gain from informing them more fully, and so how can there be any gains at all to more information?

The conclusion is that when consumers' risk-bearing behaviour diverges from the fundamental individual norm, the basic theorems of classical welfare economics lose all their ethical force, except in those obviously rather special economies for which better information is of no value. Market allocations are almost never optimal, however well schemes of contingent lump-sum income transfers may be drawn up.

15 CONCLUSION

Thirty years ago, Harsanyi (1955, p. 32) concluded his famous article with the following two paragraphs (quoted *passim*):

> There is here an interesting analogy with the theory of statistical decisions . . . [It] has been shown that a rational man . . . must act *as if* he ascribed numerical subjective probabilities to all alternative hypotheses, even if his factual information is insufficient to do this on

> an objective basis – so in welfare economics we have also found that a rational man ... must likewise act *as if* he made quantitative interpersonal comparisons of utility, even if his factual information is insufficient to do this on an objective basis.
>
> Thus if we accept individualistic ethics ... our social welfare function will always tend to take the form of a sum (or mean) of individual utilities; but whether the weights given to these individual utilities have an objective basis or not will depend wholly on the extent of our factual (psychological) information.

This chapter reaches a very similar conclusion from assumptions that initially appear much weaker. The main difference with Harsanyi comes in using an idealized normative utility for the fundamental individual norm of Section 8. Harsanyi (1977b) makes explicit his rejection of Moore's (1903) ideal utilitarianism; my proposal may possibly meet his legitimate criticisms of Moore by making it true by definition that behaviour *should* maximize the utility of its consequences. Also, without some such kind of idealized utility, it is hard to answer Pattanaik's (1968; 1971) concern that impersonality has little operational significance when different individuals have different attitudes to risk.

Arrow's social choice theory can thus be reconciled with an 'ideal' version of Harsanyi's 'fundamental' utilitarianism, but at three very significant costs I have identified: first, independence of irrelevant alternatives must be weakened to independence of ethically irrelevant mixed consequences; second, the fundamental individual norm must be dictated in the event of unresolvable differences of opinion over what it should be; third, consumer sovereignty must be abandoned if attitudes to risk and tastes for gambling are not to be the arbiters of trade-offs between total income and equality.

These 'costs' also represent major departures from Arrow's theory of social choice. They may be inevitable: that remains for future work to confirm or refute. Nevertheless, despite the departures from *Social Choice and Individual Values*, most of the steps in my argument relate closely to key ideas found in Arrow's work: (1) the concept of consequentialist behaviour in decision trees in Sections 2 and 3; (2) the characterization of ordinal choice in Section 3; (3) the arguments for the independence axiom and for expected utility maximization in Section 4; (4) the use of fundamental preferences representing 'extended sympathy' (though for risky consequences here) in Section

8; (5) the need to weaken Arrow's independence of irrelevant alternatives condition that becomes apparent in the version of Arrow's impossibility theorem presented in Section 9; (6) the need for an 'ethical dictator' shown in Section 12, which Arrow foresaw in 1975; (7) the relation between attitudes to risk and the distribution of income in Section 13; and (8) the discussion of the efficiency theorems of welfare economics and the role of securities markets as opposed to contingent lump-sum transfers in Section 14. A wealth of key ideas, most of which have stood the test of time for over thirty years already. With the main challenge in social choice theory, through the consideration of risky 'personal' consequences, coming from ideas due to Arrow's former student, Harsanyi! And, contrary to my suggestion in Hammond (1976b), it now does seem clear that independence of irrelevant alternatives is the crucial condition in Arrow's social choice theory, as he has consistently asserted.

NOTES

1. See, for instance, the papers by Hausner and by Thrall in Thrall, Coombs and Davis (eds) (1954), and also Skala (1975), Fishburn (1982, Chapter 4).
2. But notice too Arrow's (1952; 1983a, p. 48) use of Mitchell's quotation from Bentham.
3. This part is omitted from Arrow (1977a).
4. This conclusion is to a large extent supported by the results in Hammond (1985c).

REFERENCES

Arrow, K. J. (1950) 'A Difficulty in the Concept of Social Welfare', *Journal of Political Economy*, 58: 328–46; reprinted as Chapter 1 of Arrow (1983a).

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

Arrow, K. J. (1951b) 'An Extension of the Basic Theorems of Classical Welfare Economics', in J. Neyman (ed.) (1951) *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley.: University of California Press) pp. 507–32; reprinted as Chapter 2 of Arrow (1983b).

Arrow, K. J. (1951c) 'Alternative Approaches to the Theory of Choice in Risk-Taking Situations', *Econometrica* 19: 404–37; reprinted as Chapter 1 of Arrow (1971).

Arrow, K. J. (1952) 'Le principe de rationalité dans les décisions collectives', *Économie Appliquée* 5: 469–84; trans. as Chapter 3 of Arrow (1983a).

Arrow, K. J. (1953; 1963). 'The Rôle of Securities in the Optimal Allocation of Risk-Bearing', *Review of Economic Studies*, 31: 91–6 (first published in translation as 'Le rôle des valeurs boursières pour la répartition la meilleure des risques', *Econometrie* (Colloques Internationaux du Centre National de la Recherche Scientifique) 11: 41–47); reprinted as Chapter 4 of Arrow (1971) and as Chapter 3 of Arrow (1983b).

Arrow, K. J. (1959) 'Rational Choice Functions and Orderings', *Economica* (N.S.) 26: 121–7; reprinted in Arrow (1984a).

Arrow, K. J. (1965) *Aspects of the Theory of Risk-Bearing* (Helsinki: Yrjö Jahnsson Foundation).

Arrow, K. J. (1967) 'The Place of Moral Obligation in Preference Systems' in S. Hook (ed.) (1967) *Human Values and Economic Policy* (New York: New York University Press) pp. 117–19; reprinted as Chapter 5 of Arrow (1983a).

Arrow, K. J. (1971) *Essays in the Theory of Risk-Bearing* (Amsterdam: North-Holland).

Arrow, K. J. (1973a; 1974) 'General Economic Equilibrium: Purpose, Analytic Techniques, Collective Choice', in *Les Prix Nobel en 1972* (1973) (Stockholm: The Nobel Foundation) pp. 206–31; reprinted in *American Economic Review*, 64: 253–72, and as Chapter 9 of Arrow (1983b).

Arrow, K. J. (1973b) 'Some Ordinalist–Utilitarian Notes on Rawls's Theory of Justice', *Journal of Philosophy*, 70: 245–63; reprinted as Chapter 8 of Arrow (1983a).

Arrow, K. J. (1973c) 'Formal Theories of Social Welfare', in P. P. Wiener (ed.) (1973) *Dictionary of the History of Ideas* (New York: Scribner) vol. 4, pp. 276–84; reprinted as Chapter 9 of Arrow (1983a).

Arrow, K. J. (1977a) 'Extended Sympathy and the Possibility of Social Choice', *American Economic Review, Papers and Proceedings*, 67 (1): 219–25; reprinted as Chapter 11 of Arrow (1983a).

Arrow, K. J. (1977b) 'Current Developments in the Theory of Social Choice', *Social Research*, 44: 607–22; reprinted as Chapter 12 of Arrow (1983a).

Arrow, K. J. (1978) 'Nozick's Entitlement Theory of Justice', *Philosophia*, 7: 265–79; reprinted as Chapter 13 of Arrow (1983a).

Arrow, K. J. (1983a) *Collected Papers of Kenneth J. Arrow*, vol. 1: *Social Choice and Justice* (Cambridge, Mass.: Harvard University Press).

Arrow, K. J. (1983b) *Collected Papers of Kenneth J. Arrow*, vol. 2: *General Equilibrium* (Cambridge, Mass.: Harvard University Press).

Arrow, K. J. (1984a) *Collected Papers of Kenneth J. Arrow*, vol. 3: *Individual Choice Under Certainty and Uncertainty* (Cambridge, Mass.: Harvard University Press).

Arrow, K. J. (1984b) *Collected Papers of Kenneth J. Arrow*, vol. 4, *The Economics of Information* (Cambridge, Mass.: Harvard University Press).

Arrow, K. J. (1985) 'The Potentials and Limits of the Market in Resource Allocation', in G. R. Feiwel (ed.) (1985) *Issues in Contemporary Microeconomics and Welfare* (London: Macmillan) Chapter 2, pp. 107–24.

Arrow, K. J., Blackwell, D. and Girschik, M. A. (1949) 'Bayes and Minimax Solutions of Sequential Decision Problems', *Econometrica*, 17: 213–44; reprinted as Chapter 1 of Arrow (1986b).

d'Aspremont, C. and Gevers, L. (1977) 'Equity and the Informational Basis of Collective Choice', *Review of Economic Studies*, 44: 199–209.

Aumann, R. J. (1964) 'Markets with a Continuum of Traders', *Econometrica*, 32: 39–50.

Aumann, R. J. (1966) 'Existence of Competitive Equilibria in Markets with a Continuum of Traders', *Econometrica*, 36: 1–17.

Border, K. C. (1983) 'Social Welfare Functions for Economic Environments with and without the Pareto Principle', *Journal of Economic Theory*, 29: 205–16.

Brandt, R. B. (1967) 'Personal Values and the Justification of Institutions', in S. Hook (ed.) *Human Values and Economic Policy* (New York: New York University Press) pp. 22–40.

Coles, J. L. and Hammond, P. J. (1986) 'Walrasian Equilibrium without Survival: Existence, Efficiency and Remedial Policy', Stanford University Institute for Mathematical Studies in the Social Sciences, Economics Technical Report No. 483.

Coughlin, P. J. (1984) 'Rights and the Private Pareto Principle', mimeo., University of Maryland, forthcoming in *Economica*.

Debreu, G. (1954) 'Valuation Equilibrium and Pareto Optimum', *Proceedings of the National Academy of Sciences of the U.S.A.*, 40: 588–92.

Debreu, G. (1959) *Theory of Value: An Axiomatic Analysis of General Equilibrium* (New York: Wiley).

Edgeworth, F. Y. (1881) *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London: Kegan Paul).

Feiwel, G. R. (ed.) (1985) *Issues in Contemporary Microeconomics and Welfare* (London: Macmillan).

Fishburn, P. C. (1982) *The Foundations of Expected Utility* (Dordrecht: Reidel).

Fleming, M. (1952) 'A Cardinal Concept of Welfare', *Quarterly Journal of Economics*, 66: 366–84.

Gibbard, A. (1974) 'A Pareto-Consistent Libertarian Claim', *Journal of Economic Theory*, 7: 388–410.

Graaff, J. de V. (1957) *Theoretical Welfare Economics* (Cambridge: Cambridge University Press).

Guesnerie, R. and Jaffray, J.-Y. (1974) 'Optimality of Equilibrium of Plans, Prices and Price Expectations', in J. H. Drèze (ed.) (1974) *Allocation Under Uncertainty: Equilibrium and Optimality* (London: Macmillan) Chapter 5, pp. 71–86.

Hammond, P. J. (1976a) 'Changing Tastes and Coherent Dynamic Choice', *Review of Economic Studies*, 43: 159–73.

Hammond, P. J. (1976b) 'Equity, Arrow's Conditions and Rawls' Difference Principle', *Econometrica*, 44: 793–804.

Hammond, P. J. (1977a) 'Dynamic Restrictions on Metastatic Choice', *Economica*, 44: 337–50.

Hammond, P. J. (1977b) 'Dual Interpersonal Comparisons of Utility and the Welfare Economics of Income Distribution', *Journal of Public Economics*, 7: 51–71.

Hammond, P. J. (1979) 'Straightforward Individual Incentive Compatibility in Large Economies', *Review of Economic Studies*, 46: 263–82.

Hammond, P. J. (1981) 'Ex-Ante and Ex-Post Welfare Optimality Under Uncertainty', *Economica*, 48: 235–50.
Hammond, P. J. (1982a) 'Utilitarianism, Uncertainty and Information', in Sen and Williams (eds) (1982) *Utilitarianism and Beyond* (Cambridge: Cambridge University Press) Chapter 4, pp. 85–102.
Hammond, P. J. (1982b) 'Liberalism, Independent Rights, and the Pareto Principle', in L. Cohen, J. Łoś, H. Pfeiffer and K. P. Podewski (eds) (1982) *Logic, Methodology and the Philosophy of Science VI* (Amsterdam: North-Holland) pp. 607–20.
Hammond, P. J. (1983a) 'Ex-Post Optimality as a Dynamically Consistent Objective for Collective Choice Under Uncertainty', in P. K. Pattanaik and M. Salles (eds) (1983) *Social Choice and Welfare* (Amsterdam: North-Holland) Chapter 10, pp. 175–205.
Hammond, P. J. (1983b) 'Multilateral Incentive Compatibility in Continuum Economies', Stanford University Institute for Mathematical Studies in the Social Sciences, Economics Technical Report no. 435.
Hammond, P. J. (1985a) 'Welfare Economics', in G. R. Feiwel (ed.) (1985) *Issues in Contemporary Microeconomics and Welfare* (London: Macmillan) Chapter 13, pp. 405–34.
Hammond, P. J. (1985b) 'Consequentialist Behaviour in Decision Trees and Expected Utility', Stanford University Institute for Mathematical Studies in the Social Sciences, Economics Working Paper no. 112.
Hammond, P. J. (1985c) 'Independence of Irrelevant Personal Consequences', presented at the 5th World Congress of the Econometric Society.
Hammond, P. J. (1986) 'Consequentialist Social Norms for Public Decisions', in W. P. Heller, R. M. Starr and D. A. Starrett (eds) (1986) *Social Choice and Public Decision Making: Essays in Honor of Kenneth J. Arrow*, Vol. I (Cambridge: Cambridge University Press) chapter 1, pp. 3–27.
Hare, R. M. (1952) *The Language of Morals* (Oxford: Clarendon Press).
Hare, R. M. (1963) *Freedom and Reason* (Oxford: Clarendon Press).
Hare, R. M. (1981) *Moral Thinking: Its Levels, Method and Point* (Oxford: Clarendon Press).
Harsanyi, J. C. (1953) 'Cardinal Utility in Welfare Economics and the Theory of Risk-Taking', *Journal of Political Economy*, 61: 434–35.
Harsanyi, J. C. (1954) 'Welfare Economics of Variable Tastes', *Review of Economic Studies*, 21: 204–13.
Harsanyi, J. C. (1955) 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility', *Journal of Political Economy*, 63: 309–21.
Harsanyi, J. C. (1975a) 'Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls' Theory', *American Political Science Review*, 69: 594–606.
Harsanyi, J. C. (1975b) 'Nonlinear Social Welfare Functions: Do Welfare Economists Have a Special Exemption from Bayesian Rationality?' *Theory and Decision*, 6: 311–32.
Harsanyi, J. C. (1976) *Essays in Ethics, Social Behaviour, and Scientific Explanation* (Dordrecht: Reidel).
Harsanyi, J. C. (1977a) *Rational Behavior and Bargaining Equilibrium in Games and Social Situations* (Cambridge: Cambridge University Press).
Harsanyi, J. C. (1977b) 'Morality and the Theory of Rational Behavior',

Social Research, 44: 623–56; reprinted as Chapter 2, pp. 39–62 of Sen and Williams (1982).

Harsanyi, J. C. (1978) 'Bayesian Decision Theory and Utilitarian Ethics', *American Economic Review, Papers and Proceedings*, 68: 223–8.

Herstein, I. N. and Milnor, J. (1953) 'An Axiomatic Approach to Measurable Utility', *Econometrica*, 21: 291–7.

Hook, S. (ed.) (1967) *Human Values and Economic Policy* (New York: New York University Press).

Kolm, S.-C. (1972) *Justice et Équité* (Paris: Editions du Centre National de la Recherche Scientifique).

Lerner, A. P. (1944) *The Economics of Control* (London: Macmillan).

Malinvaud, E. (1952) 'Note on von Neumann–Morgenstern's Strong Independence Axiom', *Econometrica*, 20: 679.

Marschak, J. (1950) 'Rational Behavior, Uncertain Prospects, and Measurable Utility', *Econometrica*, 18: 111–41.

Moore, G. E. (1903) *Principia Ethica* (Cambridge: Cambridge University Press).

Neumann, J. von and Morgenstern, O. (1944; 3rd edn, 1953) *Theory of Games and Economic Behavior* (Princeton: Princeton University Press).

Pattanaik, P. K. (1968) 'Risk, Impersonality and the Social Welfare Function', *Journal of Political Economy*, 76: 1152–69.

Pattanaik, P. K. (1971) *Voting and Collective Choice* (Cambridge: Cambridge University Press).

Phelps, E. S. (ed.) (1973) *Economic Justice: Selected Readings* (Harmondsworth: Penguin).

Phelps, E. S. (1977) 'Recent Developments in Welfare Economics: Justice et Équité', in M. D. Intriligator (ed.) (1977) *Frontiers of Quantitative Economics* (Amsterdam: North-Holland) vol. III B, Chapter 16, pp. 703–30.

Radner, R. (1972) 'Existence of Equilibrium of Plans, Prices and Price Expectations in a Sequence of Markets', *Econometrica*, 40: 289–303.

Rawls, J. (1957) 'Justice as Fairness', *Journal of Philosophy*, 54: 653–62.

Rawls, J. (1958) 'Justice as Fairness', *Philosophical Review*, 67: 164–94.

Rawls, J. (1971; 1972) *A Theory of Justice* (Cambridge, Mass.: Harvard University Press; Oxford: Oxford University Press).

Rowley, C. K. and Peacock, A. T. (1975) *Welfare Economics: A Liberal Restatement* (London: Martin Robertson).

Samuelson, P. A. (1952) 'Probability, Utility and the Independence Axiom', *Econometrica*, 20: 670–8.

Savage, L. J. (1954) (2nd rev. edn, 1972) *The Foundations of Statistics* (New York: Wiley; Dover).

Selten, R. (1965) 'Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit', *Zeitschrift für die gesamte Staatswissenschaft*, 121: 301–24.

Selten, R. (1973) 'A Simple Model of Imperfect Competition, where 4 are Few and 6 are Many', *International Journal of Game Theory*, 2: 141–201.

Selten, R. (1975) 'Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games', *International Journal of Game Theory*, 4: 25–55.

Sen, A. K. (1970a) *Collective Choice and Social Welfare* (San Francisco: Holden Day).
Sen, A. K. (1970b) 'The Impossibility of a Paretian Liberal', *Journal of Political Economy*, 78: 152–7.
Sen, A. K. (1976) 'Liberty, Unanimity and Rights', *Economica*, 43 (August): 217–45.
Sen, A. K. and Williams, B. A. O. (eds) (1982) *Utilitarianism and Beyond* (Cambridge: Cambridge University Press).
Sidgwick, H. (1893) (7th edn, 1907) *The Method of Ethics* (London: Macmillan).
Skala, H. J. (1975) *Non-Archimedean Utility Theory* (Dordrecht: Reidel).
Smart, J. J. C. and Williams, B. A. O. (1973) *Utilitarianism: For and Against* (Cambridge: Cambridge University Press).
Starr, R. (1973) 'Optimal Production and Allocation under Uncertainty', *Quarterly Journal of Economics*, 87: 81–95.
Strasnick, S. (1977) 'Ordinality and the Spirit of the Justified Dictator', *Social Research*, 44: 668–90.
Strotz, R. H. (1956) 'Myopia and Inconsistency in Dynamic Utility Maximization', *Review of Economic Studies*, 23: 165–80.
Thrall, R. M., Coombs, C. H. and Davis, R. L. (eds) (1954) *Decision Processes* (New York: Wiley; London: Chapman and Hall).
Tinbergen, J. (1957) 'Welfare Economics and Income Distribution', *American Economic Review Papers and Proceedings*, 47: 490–503.
Vickrey, W. S. (1945) 'Measuring Marginal Utility by Reactions to Risk', *Econometrica*, 13: 319–33.
Vickrey, W. S. (1960) 'Utility, Strategy, and Social Decision Rules', *Quarterly Journal of Economics*, 74: 507–35.
Vickrey, W. S. (1961) 'Risk, Utility and Social Policy', *Social Research*, 28: 205–17; reprinted as Chapter 11 of Phelps (1973).