Arrow’s Impossibility Theorem:
The Logic of Dictatorship and the Meaning of Social Choice

Paul Anand
The Open University
Milton Keynes MK7 6AA
United Kingdom
p.anand@open.ac.uk

ROUGH Draft – Please do not cite or further circulate

To be given Baton Rouge, University of Louisville, USA
February 2004

Bio: Dr Paul Anand is Reader in Decision Science and Public Policy at The Open University and Research Associate at the Health Economics Research Centre, Oxford University. His main current interests are in economics and philosophy and the interplay between foundations and applications of decision theory and social choice.
Arrow’s Impossibility Theorem:  
The Logic of Dictatorship and the Meaning of Social Choice

Abstract
The paper pays critical tribute to Arrow’s (1951) analysis of social choice by focusing on a number of the foundational, conceptual and interpretational issues to which it gives rise. It begins with a discussion of justifications of other formal requirements, namely unrestricted domain, Pareto optimality and independence of irrelevant alternatives, before proceeding to a detailed analysis of the way in which Arrow’s famous Theorem formalises the concept of dictatorship. The term and the formalisation still play a central role in contemporary social choice theory though we shall argue that, for a variety of reasons, the formalisation provides neither necessary nor sufficient conditions for the existence of an agent with powers of inappropriate influence. Subsequently, the paper examines the claim that Arrow believed had been accepted by critics of his theory, namely that social choice is fundamentally about preference aggregation and suggests that preference aggregation should only be regarded as a special case of social choice. (In this regard, I closely follow critics of utilitarianism and of its economic implementation in conventional welfare economics, and offer some empirical evidence which helps to illustrate the criticisms theorists have made and make sense of people’s objections to social choice procedures such as QALY maximisation.) Finally, I conclude that, whilst not decisive in conceptual terms, the Theorem is an important contribution to C20th thought, offers a platform for mathematical, philosophical and even empirical work pertaining to a broadening conception of social choice.

Key words: Arrow’s Theorem, dictator, social choice, empirical ethics, economic theory and philosophy
1. Introduction

One of the central contributions to C20th thought was the idea, formalised apparently in Arrow’s Theorem (1951), that social choice was deeply problematic. Though Kenneth Arrow is a Nobel economist, and a fine mathematical one with a rare and wide grasp of policy issues at that, his famous possibility Theorem, more popularly described as an impossibility theorem, has had very substantial impacts in philosophy and political science as well. However, to say that the Theorem has been the subject of misunderstanding and controversy would be to understate the case by a substantial margin. Over fifty years after it first appeared in monograph form, it is possible to hear researchers express views about the theorem’s significance that range from ‘devastating for the possibility of capitalist democracy’ to suggesting that it is ‘just a theorem’. Widely divergent views can be heard whenever the topic arises and there are even disagreements about whether there could be reasonable disagreement about the implications of the famous theorem.

Yet at some level, the interpretation of a theorem, particularly an axiomatic one designed at least in part to make issues transparent, is just as important as its technical details. For example, if one thinks of Riemennian geometry in which parallel lines can meet, then it is of value to know that this geometry describes the behaviour of straight lines on a sphere. Historically, this application or interpretation helped many mathematicians to accept the geometry’s validity and indeed Arrow’s own discussions are replete with comments and pointers to issues of foundation, justification and interpretation.

In this paper, I want to pick up on two such issues which I think are important in their own right but which also generate technical, or at least, logical and conceptual issues as well as help us to evaluate the significance of Arrow’s insights. The first concerns the justification and interpretation of axioms and draws parallels with changes in decision theory, broadly since 1980, as a model. Up to that point in time, the axioms of expected utility were regarded as cannons of rationality though since then, normative arguments, technical knowledge, and empirical evidence have all pushed us away from viewing the assumptions of completeness,

---

1 My interest in the theorem discussed here owes much to informal comments about its origins Ken Arrow made to a small group after a seminar in Bielefeld, Germany. For discussions on this and related ideas at workshops and seminars in Oxford, Norwich, Caen and Lund, I am especially grateful to Robin Cubitt, David Kelsey, Clemens Puppe, Martin van Hees, Keith Dowding and Philippe Mongin.
transitivity and independence as essential or characteristic conditions of rational choice. Arrow never claimed that his ‘conditions’ were axiomatic for all social choices but he came fairly close when he argued that they were apparently highly desirable over a very wide range of situations. Each of his conditions has been the subject of some discussion and to the extent that this suggests limits on universality claims, so the worrying implications of his Theorem are mitigated. Subject to less discussion than other conditions is the non-dictatorship requirement, possibly because it is self-evident that we employ voting systems for a variety of reasons not least being the fact that we wish to eschew dictatorship. However, even if we take this for granted, there remains open the question with this condition, as with any other axiom, as to whether the formalisation is adequate and in if so in what so sense, and it is in these regards that I think there is a weakness that has gone under-acknowledged in the literature. Specifically, I shall argue that the formalisation of dictatorship is rather questionable as it provides neither a necessary nor a sufficient condition for the existence of a dictator. Indeed I shall suggest that it in one sense, it points in the wrong direction. The formalisation is therefore different in nature and be contrasted with, say, that offered by Pareto optimality which, when applied to voting, seems to provide necessary and sufficient conditions for the state of unanimity that it purports to formalise.

The second related, issue concerns more broadly and less formally, the extent to which we can, following Arrow’s Theorem, view social choice as being about the aggregation of preferences. It is held that this view is too modally flat and, following Sen (1979) *inter alia*, that a broader of conception of social choice can be articulated. The elements of this broader conception are sketched, and illustrated with empirical evidence, before concluding with a few remarks on attempts to interpret mathematics outside the traditional, physical applications which some believe are privileged. The rest of the paper is structured as follows. Section 2 reviews normative arguments for unrestricted domain (U), Pareto optimality (PO) and independence of irrelevant alternatives (IIA) whilst section three carries the main analysis of dictatorship and its formalisation. Section 4 addresses the conception of social choice whilst section 5 concludes with remarks about the value and interpretation of mathematics in economic theory.

---


3 One paper that gives some support for the view analysed here is Dowding (1997).

4 The phrase I owe to J R Lucas and is taken from one of his wonderful lectures on the philosophy of mathematics.
2. The Limits of U, PO and IIA.

Arrow’s theorem has, at its core, the idea of a social welfare function (SWF) which maps a set or space of individual orderings (usually called a profile) of options which may be political candidates or more generally states of affairs, onto a single, social ordering of same. By imposing constraints on the form of the SWF, Arrow demonstrates that eventually, there is a contradiction that indicates that there is no logical space left for a SWF to satisfy the desiderata he proposes simultaneously. Any subset of his conditions is mutually compatible so his conditions specify boundaries of what is possible.

Though the use of functions in economics is widespread – some attribute this to the unificatory activities of the group of French mathematicians known collectively as Bourbaki – the framework for Arrow’s use of functions in social choice already excludes certain social choice processes which theorists have argued can be fair. For example, the idea that social choice can be represented as the mapping from a set of orderings to the social ordering excludes the possibility that the function is multi-valued (in the co-domain). Yet, this is exactly what is needed to represent outcomes that arise only with a particular probability. When it comes to the allocation of scarce discrete goods (kidney dialysis machines for example), it is often claimed that the only fair non-market social procedure is one which randomises allocation to recipients. Yet such a process cannot be described within the usual interpretations of Arrow’s framework. This exclusion provides a simple example of a basic and fundamental limitation of the conceptual framework Arrow provides – though one which attracts little attention to itself. To pursue the idea of clear but possibly unobvious conceptual limits further, I shall make some remarks, before discussing the formalisation of dictatorship, about the normative desirability of other assumptions which Arrow made to restrict the form of the SWF.

It is often said that the appeal, or defence, of U derives from a core feature of democratic systems namely that voters should be allowed to give vent to their preferences, whatever they are (eg. Saari (1994)). However, it is also a feature of democratic systems that we constrain the formation and articulation of preferences which are at odds with the existence (and to a degree the dignity) of others. Many countries, for example, regulate advertising aimed at children, constrain the grounds on which employers can treat workers differentially or ban political candidates who would propose certain kinds of extremist policies. These actions and

---

5 If we confined our attentions to weak rankings or strict rankings, it would be natural to think of this as the collapse of a set of permutations onto a single element of the set.

6 I don’t discuss here connectedness or transitivity but the interested reader may consult a number of the references in footnote 2.
ones like them are not captured well, if at all, by U and to the extent that the Theorem has been used to derive implications about social choice in capitalist democracies, the use of U is limiting. Of course, one could say that the alternatives in Arrow’s Theorem are just those that are left when the unacceptable ones have been filleted out. But that seems to undermine the libertarian justification of U. Furthermore, appealing to prior restrictions on preferences limits the extent to which the framework deals with social choice in the round, as opposed to, say, a final selection stage. However, it is perfectly possible to read Arrow’s Theorem as applying to any group choice where there are at least three options that it is acceptable to rank in any way possible. In most cases, if there are three legitimate options, it is difficult to think of substantive and plausible objections to any particular rankings of those options. Read in this way, the Theorem tells us something valuable about group choice where there are at least three reasonable options: the limited relevance for voting and social choice derives from the fact that the options before us may not always be socially acceptable. So even U can have quite divergent justifications and not all of these will be equally acceptable.

PO has traditionally been presented as a, if not the, bedrock of the new welfare economics even though its limits have been known for some time. The idea behind the principle is just that if something is better for all individuals, it will be better for society – often described as unanimity - and the principle is one a small number of such principles in economics whose appeal derives, it might be argued, from their deceptive proximity to tautology. However, it has been known for some time that PO cannot cope with important distributional issues, Sen (1979). More recently, experimental work concerning choice in games such as the ultimatum game has provided empirical evidence that also questions the acceptability of PO improvements. In ultimatum games, where a distribution of payoffs to players of is a Pareto improvement over the status quo, most players fail to accept offered distributions that depart too far from equality. For example, if individual i offers j a division which allocates to each £80 and £20 respectively, many j players prefer to accept a distribution in which both players get nothing. Though this experimental finding is generally offered as an example of a situation in which players demur from their Nash equilibrium strategies, it also provides evidence of social choices in which unbridled faith in PO is frustrated by a sense of (distributive) fairness. Even though the principle might apply, there is little reason to suppose that it must do so if at least one person wishes to reject it. And in real social choice or policy applications there is perhaps even more reason to think that the condition might not be universal. If the income of low-income groups rises marginally whilst those on very high incomes grows significantly, people across the spectrum could reasonably find this
objectionable. Admittedly, engineers use Pareto analysis extensively and when one considers problems involving choice of manufacturing methods that yield, say, high quality products at lower cost, Pareto dominance seems uncontroversial. However, the same cannot be said of applications where distributional sensitivities matter as they do in many social choice problems.

Of all four conditions, it was the independence of irrelevant alternatives assumption that Arrow (19xx) thought should be dropped. The view appears to have been influential within econometrics where the estimation of choice functions that allow violations of IIA has been the object of some research. The reason Arrow gives for this view is just that the introduction of additional options might provide new information about the existing options in a choice set. This seems fair, and certainly one can see how it might be inevitable if the new option were a new political candidate offering voters his/her own views about the existing candidates. But it does suggest that the primitive options in Arrow’s theorem cannot be taken to be ‘materially complete’ descriptions as Savage (1954) suggests they must be if axioms of decision theory (like transitivity) are to be appealing. Certainly it is widely accepted that what plausibility any such axioms might have in the presence of materially complete option descriptions usually evaporates as soon as the requirement is not met. (In fact, it is not clear that any of the axioms of decision theory could apply coherently to incomplete option descriptions. See for example Anand (1993a)). So if one accepts that social choice is committed to the analysis of choice over materially incomplete descriptions, there is just as much reason to question the requirement that the social welfare mapping from the space $R^n_1 \times R^n_2 \times \ldots R^n_m$ to $R^n$ be transitive. Dropping both I and T, in the contexts that Arrow had in mind are possible responses to the Theorem but doing so for this reason raises and then leaves open the question as to whether the collection of Arrow’s conditions would be (more) compelling, as a set, if we imposed Savage’s materially complete descriptions requirement.

2. When is a Dictator Not A Dictator?

Any reading of the literature scattered over a range of disciplines would find it hard not to conclude that particular significance is given to the requirement that SWF not be dictatorial. The formal literature, including relatively recent work that offers simpler proofs of Arrow’s results, seems to accept that the avoidance of dictatorship is an important consideration for designers of social choice mechanisms. Indeed, it is difficult to imagine the grounds on which any reasonable person might demure. Of course, and as Arrow reminds us, his contribution deals only with formal issues, which may not exactly mirror the concept of dictatorship in
natural language but even then, his characterisation of dictatorship raises some serious questions.

To see the force of these questions, we should recall that Arrow (1963 p30) defines a SWF to be dictatorial if:

\[ \exists i \in S: \forall xy \in O, xP_i y \rightarrow xPy \tag{1} \]

where \( i \) is an individual in some set, \( S \), of \( m \) individuals, \( x \) and \( y \) are social options from a set \( O \) of \( n \) options, \( P_i \) is \( i \)'s preference relation and \( P \), society’s preference relation. The existence of a situation in which a single individual can determine the social ordering for all possible pair-wise choices regardless of the wishes of others, certainly seems to be something we should avoid. But the question I wish to pose concerns the extent to which (1) is an adequate formalisation of this constraint on the social choice process.

A little reflection shows that it is far from obvious that (1) must be true even if a person with dictatorial powers, particularly ones exercised, belongs to \( S \). It may be, for instance, that there is a relatively large sub-set of \( O \), say \( \{p,q,\ldots,x,y\} = L \), for which the implication in (1) is true even if it fails to hold for all possible choices. Some dictators might not want their way on absolutely everything for which they have a preference, so the universal quantifier in (1) is too strong and therefore unwarranted. Likewise, an individual might exert disproportionate influence on a subset of all possible pairwise choices in a manner that we would find unacceptable though such a person would not be dictatorial in the sense of (1). A possible response is to point out, as some have\(^7\) that if, in the context of Arrow’s theorem, the formal statement of what it is to be a dictator could be weaker, then Arrow’s insight is even stronger than his theorem states. The observation is a good one though it does nothing to rebut the suggestion that the failure of necessity suggests that the formalisation is further away from the concept than we might suppose and that this gap naturally leads to further probing, particularly concerning questions of sufficiency.\(^9\)

\(^7\) None of the points I make turn on the distinction between strict and weak preference although they are more transparently made using the strict preference relations, \( P, P_i \) etc.. Arrow defines strict preference in terms of weak preference in the customary manner.

\(^8\) I am grateful to Philippe Mongin and a seminar participant in Basel for this observation.

\(^9\) One further point that could be made concerns the extent which dictators might manipulate other people’s preferences. In general we want to limit the shaping of other people’s preferences but the Arrowvian framework deals with fixed preferences so this issue cannot be discussed without further work. As it stands, the framework needs to be elaborated if is to provide a substantially correct account of the constraints that operate on social choice mechanisms. It might be that such accounts have not
If the formal statement in (1) is not a necessary condition for the existence of at least one person with dictatorial powers, might we at least regard this an evidential condition i.e. one the satisfaction of which can be taken as suggestive that something might be awry? It is tempting to think that at least we can give a positive answer to this question even if it appears to change, if not demote, the significance we attach to the presence of an Arrowvian dictator. Certainly if there were a single person who had the capacity to assert his or her will and did so in every possible choice where s/he had a preference, then (1) would be satisfied. But even from an evidential perspective, there are difficulties.

To see the point about sufficiency, suppose we rewrite the implication part of (1), suppressing the quantifiers, as a conditional probability thus:

$$ p(xP_{y} | xP_{i}y) = 1 \quad (2) $$

This formulation brings different considerations to mind. Even if there is an individual whose preferences we can use to predict the social ordering with perfect accuracy, which is one thing (2) suggests, correlation is not the same as causation. So we need to consider the possible causes that would give rise to the satisfaction of (2).

One possibility is that the equation in (2) holds true for one individual because it is true for all $i \in S$. In that case, for all pairs of options, if any person had a strong preference, then society would exhibit the same strong preference. The reason is this. If at least one person had a strong preference with respect to a pair of options, then anyone else who had a strong preference would have to have the same strong preference. This is true for all options so there would be unanimity over all strong preferences: weak preference and indifference are symmetric relations so unanimity over options for which there are strong preferences implies unanimity over all options (If there were a pair of options where no one had a strong preference, then the dictatorship axiom, taken on its own, allows society to express any preference). So (2) could be true if there were unanimity which is hardly disturbing and certainly not a state of affairs one would always want to rule out as the Arrowvian formalisation appears to.

---

been developed as they appear to be redundant from the viewpoint of generating impossibility results.
When Arrow motivates (1), he says a dictator is a person who can have reflected all their strong preferences in the social ordering, ‘regardless’ of other people’s preferences. Some people take this to be a counter to the point just made because the state of unanimity on which it relies seems not to address the ‘regardless’ aspect of the formalisation. However, this is not quite right as Arrow, in his discussions of unanimity, implicitly accepts. The Arrowvian argument is just that a conjunctive set of assumptions implies the truth of (1): the fact that (1) gives cause for concern in some situations means that we can only infer the existence of a problem for social choice in just those circumstances. Put another way, it is true that the Theorem establishes, for its assumptions, the existence of at least one individual who will always see their strong preferences mirrored in the social ordering but we cannot assume that this is, in every situation, something about which we should be distraught.

Another, related difficulty arises from the observation that the conditional relationship in (1) is logically equivalent to the following holding true for all elements of O:

\[ \neg xPy \rightarrow \neg xP_iy \quad (3). \]

(3) tells us that whenever society does not strongly prefer x over y say, there will be an individual who also does not strongly prefer x to y, which is hardly surprising. What is surprising is that there is at least one particular individual who makes this statement true for all option pairs. The fact is hardly obvious – if not quite paradoxical – but nor is it obviously worrying. Indeed, it is difficult to think of a way in which a person must be worried by (3) despite its logical relation to dictatorship as formalised in the Theorem.\(^{10}\)

Logic aside, one might challenge sufficiency not just abstractly but on more concrete grounds too. One possibility is that the Theorem just points to the inevitable existence of someone who is ‘lucky’; if a SWF exists which satisfies the three other axioms of Arrow’s theorem, then we know that at least one person would see their binary preference mirrored in the social ordering whatever choice was being considered. This description of dictatorship is consistent with the formalisation and the Theorem but underlines the fact that (1) is neutral with respect to causality. Such a description we might use to conclude that what Arrow shows is this: if his other conditions are satisfied, then any SWF will always guarantee that (at least) one person will get their way in any choice between two options. This, too, is not an

\(^{10}\) I suggest this poses a serious challenge for philosophers, which if resolved, would be invaluable to economic theory. The challenge is simply what should we make of consequences that clearly have different normative implications even they are logically equivalent.
obvious result as, intuitively, it would have been reasonable to conjecture that a SWF satisfying Arrow type conditions might yield the result that no one could expect to get their way in all possible choices.\textsuperscript{11} So in this positive light, we might take Arrow’s results to be more optimistic than the usual interpretation suggests. The worrying consequence for practice, as opposed to theory, has to do with equity – the difficulty being that a SWF guarantees perfect efficiency but only to one person.

Ultimately, it would be best if social choices gave everyone what they wanted but this is only possible trivially, as Arrow notes, in the situation where individuals are unanimous in their orderings of social options. Indeed, this is opposed to the assumption of unrestricted domain, an extreme assumption perhaps but one that does at least reflect the need to exclude the trivial situation in which everyone agrees. But from an efficiency viewpoint alone, there is still no reason why we should not treat the satisfaction of preference for as many people in as many choices as an asymptotic ideal – one to which we should move towards, other things equal, but one which cannot itself be achieved. That this amounts to proposing that desirable SWFs should maximise, \textit{inter alia}, the number of Arrowian dictators constitutes a substantial reason for rejecting the conventional significance attributed to the ‘dictatorial’ property.

Sometimes, difficulties in the formalisation of conditions like dictatorship arise from the context in which they are proposed, specifically the other principles that they are designed to accompany. Although Arrow himself discusses the implications of unanimity for his formalisation of dictatorship, it is worth noting that the opposite (complete diversity) can give rise to dictatorship – Arrowian style. To see this, suppose we have a set of six individuals, i to n, who collectively exhibit all the possible strict preference rankings over three social options, a, b and c – an assumption we might think of as ‘full realisation’ of Arrow’s unrestricted domain assumption. For ease of exposition (see below), the orderings are written out as ternary relations from which binary preferences can be derived in the obvious way.\textsuperscript{12}

\textsuperscript{11} Indeed, one can still conjecture that this is true for a particular set of assumptions similar, but different, to those that appear in Arrow’s seminal work. However, I am unaware of such results at present.

\textsuperscript{12} Binary preferences are derived from ternary preferences simply by dropping the unavailable alternative.
In this example, it follows that if there is SWF which maps the space to an element in the manner described earlier, then it must be dictatorial because for any social ordering there is always an individual who orders the options in the same way. We might put the point more generally in the form of a proposition about the minimum of individuals required to ensure that (1) all possible SWFs are dictatorial.

**Proposition.** If individuals are uniformly distributed over the domain of possible preference orderings, the ratio of individuals to options, \( \frac{m}{n} \), required for the existence of at least one Arrowvian dictator = \( \frac{n!}{n} = (n-1)! \).

The example above illustrates this proposition simply – because there are three options, we need to have a minimum of 3! (6) individuals to be sure that all possible SWFs are dictatorial. So where the number of individuals is large relative to the number of social options, or more directly, where unrestricted domain is fully realised because every possible ranking can be found to be held by someone, satisfaction of (1) cannot be regarded as evidence of dictatorship. This shows that U combined with a sufficiently large number of voters is enough to satisfy ND. Exit routes would be to constrain U, which might be arbitrary, expand the number of alternatives which is often desirable, or reduce the number of voters which in many cases would be absurd. From this perspective, the impact of PO and IIA is to reduce the number of voters and preference profiles that give rise to contradiction or dictatorship. The consequence for the formalisation of ND is that we have another example, this time an internal criticism from within the framework, which shows that ND can be violated for reasons that have nothing to do with the existence of a dictator. Indeed the example demonstrates the extent to which the term ‘regardless’ in the motivation for the formalisation of ND, can be misleading.  

\[13\] From Arrow (1963 p30) ‘A social welfare function is said to be dictatorial if there exists an individual such that for all [pairwise preferences a strong individual preference implies a strong social preference]… regardless of all individuals other than i.’ Emphasis added.
orderings, then for a social ranking, there will be one person whose preferences determine the social ordering, ‘regardless’ of the orderings held by other people. Though the term regardless is not incorrect in a sense, its use here is misleading.

There is a final problem that goes beyond sufficiency or necessity and concerns a theme that has been emerging in this discussion, namely the extent to which (1), even if we took it just to be an indicator rather than a necessary or sufficient condition, points in the correct direction. I suggest it does not. To see this, we begin by noting that the desirability (des) of a SWF is positively related to the extent to which it can meet people’s preferences. In Arrow’s Theorem, the idea is called positive association. Although I shall question this view subsequently, this should not be controversial for theories that hold social choice to be at root a matter of preference aggregation as Arrow’s Theorem does. We might formalise this idea by saying that:

\[
\text{des}(SWF) = f\left(\text{d}\right) \quad (4)
\]

where \(d\) is a distance function (and its value when there is no confusion) defined over the pair \((P, (P_1, P_2, \ldots, P_m))\), bounded from below by 0 at the points where all the Ps are the same, and where \(f\) is a strictly decreasing function of \(d\), (i.e. \(\partial f / \partial d < 0\)). (4) says that the desirability of a SWF is a function of the difference between the social ordering and the profile of individual orderings. (Of course, desirability might be a function of other things, including differences between orderings and other possible preference profiles, a possibility that can be represented with a little additional notation.)

Now consider a sequence of states characterised by the number of Arrowvian dictators in a society:

\[
\begin{align*}
\text{For 0 in S: } & \forall xy, xP_i y \Rightarrow xPy \\
\text{For 1 in S: } & \forall xy, xP_i y \Rightarrow xPy \\
\text{For 2 in S: } & \forall xy, xP_i y \Rightarrow xPy \\
& \vdots \\
\text{For all m in S: } & \forall xy, xP_i y \Rightarrow xPy
\end{align*}
\]

Moving from one dictator to none, or generally from a situation in which there are \(l\) dictators to \(l-1\) dictators, involves, at least changing the social ordering for one pair so that, for a person who was previously dictatorial, there now exists a pair of alternatives, \(x\) and \(y\), such
that \( xP_i y \land \neg xP_y \). This indicates an increase in \( d \) and therefore a SWF that has a lower des value. (Of course the net effect of such a change will, in any particularly situation, depend on the numbers of other members in society with similar or opposed preferences and the exact nature of the metric used.)

Situations exist where this is an unambiguously poor move. For example, suppose we have a unanimous population in which the final statement in the above list is true, i.e. that (1) is true because each individual is an Arrowvian dictator. In that case we could only reduce the number of individuals whose preferences satisfied (1), keeping tastes constant, by changing the social ordering so that it fails to reflect preferences in some respect where before, it did not. This is nonsensical but as I have indicated, reflects a larger problem, namely that the formalisation of dictatorship is fundamentally at odds with the idea that better SWFs are ones which find a tighter, rather than a looser, fit between the social ordering and the profile of individual preferences.\(^{14}\) In a rough but nonetheless fairly deep sense, we think of most social choice processes as needing to avoid allowing undue influence so understanding problems in the formulation of a special case, namely dictatorship, should help us understand something about how to specify the problem of social choice itself. But even if we ignore the difficulties that surround the formalisation of ND, we shall see there are further problems concerning the conceptualisation of social choice.

4. Social Choice: Integration of Claims or Aggregation of Preferences?

Towards the end of the second version of Arrow’s book, he summarises reactions to the intellectual challenges his theorem posed. In a section entitled ‘What is the Problem of Social Choice’ Arrow (op cit p103) concludes that ‘Upon close examination, all [critics of this approach] implicitly accept the essential formulation stated here: The social choice from any given environment is an aggregation of individual preferences.’. Arrow’s theorems certainly do not require that this interpretation be correct for it to be a remarkable suite of mathematical insights about group choice over three or more options. However, at this point in history, I would suggest that preference aggregation conception does not predominate across a range of disciplines from social choice theory through to moral and political philosophy. The reasons for this change, I want to suggest, are compatible with the view of social choice that can be seen as a generalisation of that which Arrow’s Theorem formalises.

\(^{14}\) The idea that social choice involves increasing the correlation (fit) between a social ordering and the preference profile is discussed at length, perhaps for in one of the first such discussions, by Craven (1996).
The view of social choice proposed, namely that it is fundamentally about the integration of competing and potentially incommensurable types of claims, is one that I have discussed elsewhere, Anand (2003). Here, I want to emphasise the extent to which it helps highlight the limits of social choice as preference aggregation though my aim is not to argue for one over the other so much as to indicate the substantive grounds on which one might make a choice between these two conceptions. Within the utilitarian framework on which traditional welfare economics was constructed, the strength of a claim in any social choice problem is measured by the utility that the meeting of that claim will produce. The fair society is just one that maximises utility so fairness and wellbeing stand and fall together – a point which Arrow’s Theorem seems to embody, albeit within an ordinal setting. When one moves away from utilitarianism, to other forms of consequentialism and to deontological approaches beyond, it becomes possible and often essential to distinguish between fairness and wellbeing and to thereby allow that social choice procedures might satisfy one of these, but not both simultaneously. I want to suggest that social choice as preference aggregation is open to question on both fairness and wellbeing grounds.

Firstly fairness. To focus exclusively, and exhaustively depending on the reading of U involved, on preferences implies either that fairness is irrelevant to social choice or that a claim on the social choice is legitimate if and only if it is a preference. However, it is possible to challenge both the urge to respect all preferences as well as the refusal to recognise claims that are not preference based. Suppose that the reason $P_{1;ab}$ is racist or sexist. Should we then take the preference into account? Whilst it might be an open question when the choice is political (we don’t after ask people to give reasons for the way in they vote in general elections), economic systems in many countries deny that such preferences should be recognised (e.g. a variety of human resource decisions).

Even benign preferences may have a limited role to play as a famous exchange between H L A Hart (1980) and Ronald Drown (1984), can be used to illustrate. Part of their exchange focuses on the case of a person, Sarah, who is particularly popular and whose projects many people are willing to support. For present purposes, suppose that Sarah is a musician and a choice has to be made who will perform at a televised party to celebrate the President’s birthday. In this case, Sarah’s popularity might quite properly be regarded as a legitimate consideration in the choice of performer. However, suppose the choice related to health, that Sarah was ill and needed access to treatment that was in strictly limited supply – in that case, should her popularity give extra weight to her claims on resources from any state funded
health services? Many would say not though a negative response appears to be incompatible with some of the more straightforward interpretations of utilitarianism. In this case, the preferences of others are somehow irrelevant – compared with the choice-of-performer situation where they are germane. Whilst it might be fruitful to unpick what makes preferences a source of legitimate claim in some circumstances and not in others - which could be a significant theoretical task - I think the example is sufficiently clear to indicate that such diverse situations exist and that the exclusion of preference information can be a key aspect of social choice.

Conversely, not all legitimate claims are preferential. Many accept that rights and responsibilities (duties) are important claims on social choice, whilst others emphasise the importance of procedures and procedural justice. Even Nozick (1974) who has done most in modern times to articulate the notion of a minimal state that might underpin free markets agreed that people had rights. In his account, the state essentially protects a thin set of rights leaving preference satisfaction to individuals, a view that might formalised as

\[ SWF: \{a, b, \ldots, l\} \rightarrow L \subset \{a, b, \ldots, l\} \]

i.e. one in which the SWF maps from a set of options to a proper subset of ‘legitimate’ options, \(L\), by disregarding those which cause, or are caused by, rights violations. Then there are process claims of the kind discussed by proponents of procedural justice. In extremes, if everyone has their say, there may be situations where the social choice is unimpeachable, because going through the process is what constitutes. I assume that whilst the evolutionary motive for process instincts is outcome based, in combination with social practices, they are sufficiently ingrained that they take on a life of their own. Both kinds of considerations are illustrated by the evidence against utilitarianism, or its incarnation in health economics, in health reported in the appendix.

The point about social choice and wellbeing is perhaps simpler. If we take actual preferences, as we might do when people vote, then we know that there are situations where people are substantially under-informed or prone to apparent pathologies, like addiction, where the following of actual preferences might be seen to be in the interests of the individual. The standard response is to say that we should be concerned with the counterfactual preferences that a person would have if they were better informed about their own interests. But this standard move still doesn’t say about issues like autonomy or the intrinsic value of choice. It

15 I don’t deny that utilitarianism has been so capable of reinterpretation that there must be some version compatible with this position too.

16 I say apparent because some researchers have argued that addictive behaviour can be rational. I would add that preference itself is fundamentally about forming an emotional bond between the agent and things that are good for it.
just suggests that we ensure that people get what is really in their interests rather than what they happen to prefer. Furthermore, the move is powerless against the Hannibal Lector syndrome. There may be people with anti-social preferences because they are, fundamentally anti-social and yet don’t wish to exclude their priorities because they would do their ill informed bearers harm but because their bearers are fully informed and intent on doing others harm. In short, neither is social choice about aggregation or about preferences alone – these points are reflected in the empirical evidence reported in the appendix which looks at attitudes to health-care rationing and the claims people believe are relevant. [For the purposes of the presentation, I have kept the theoretical part and the empirical parts of the paper distinct. My point is that the empirical evidence indicates that we cannot, in general, assume that preference aggregation is a good model of actual social choice for reasons that reflect the reasons philosophers and economists have given for rejecting utilitarianism or welfare economics. The evidence illustrates that that there are preferences or beneficial claims that many people would discount and concerns to do with rights, duties and procedures that do count. In practice these latter claims might be consequentially motivated even if they are not utilitarian. The empirical evidence presented is consistent with evidence concerning rejections of QALY maximisation in other studies and can reasonably provide input into the formulation of ethical theories even its role in their evaluation is, and should be, secondary.]

So what is left? I think three meta-ethical criteria for social choice emerge from Arrow’s Theorem, which any theory of social choice should take into account. In turn, these are efficiency, distributive fairness and decidability. Efficiency, in terms of meeting objectives seems central to social choice, though those objectives may not always be correctly articulated in preferences. Distributive fairness is something which seems important and appears in Arrow, though rather quietly and indirectly and not in a way that one might expect. One reason for objecting to a dictatorial agent is that one person gets everything they want whilst others do not – and this distributive reason seems to fit closely with the proofs. However, this is a non-parametric approach and one that one cannot distinguish between someone earning twice as much or twenty times as much as someone else. The cardinality point has been extensively discussed elsewhere but its absence and the difficulties for Arrow’s framework to which it gives rise need not be taken to mean that distributive fairness issues are excluded completely. Finally, the emphasis on a social ordering, suggests the importance of decisiveness or decidability in social choices, particularly one-shot choice problems. As I have hinted, I believe there are other approaches to decidability, use of aleatoric mechanisms for example, which when implemented give rise to a choice, even if
that particular choice is not defined when the individual preference orderings have been specified. If one considers sequences of social choices, which Arrow’s Theorem does not, then there are questions of stability and diversity arise which do not arise in his framework. One often wants decisions that are stable, which is perhaps part of the merit of majority voting. The way in which researchers use transitivity, as a defence against cycling, suggests that the concept of stability does indeed play a role. However, to parallel a line of argument developed by Saari (op cit), the imposition of a transitive social choice ordering on a population who are unanimously intransitive may not lead to a stable social choice. The concept of stability that I think is more important is one which that we might think of as long term support and is perhaps better analysed using a game-theoretic approach: transitivity I believe to a weak or mis-specification of the stability requirement. Finally, there is a question about diversity, which perhaps pulls in the opposite direction. We value diversity, if we do, because people are different and the future is unknown yet it is difficult to see how only a minor change to Arrow’s framework could incorporate such concerns.

3. Concluding Remarks
This paper has looked at two aspects of Arrow’s Theorem – the axioms themselves and the interpretation of social choice to which they give rise. Without attempting to summarise the detailed arguments, let me conclude with some observations about the role of mathematics in economic analysis, particularly economic theory. It used to be claimed that mathematics was just a language which brought rigor to proceedings and axiomatic theories have often been additionally justified on the grounds that they are transparent. That even qualitative methods can be useful and rigorous is a view that could reasonably be held and is so to a degree that varies between disciplines and within disciplines over time. A converse argument, which some decision theorists accept, is that the value of mathematics is in providing approximate arguments or qualitative insights. In this case, I have argued that there are strong parallels between the dictatorship condition of Arrowvian social choice and the reinterpretation of the relation between transitive preference and rational choice in decision theory. Specifically, I have argued that the formalisation of dictatorship in the Theorem is not a necessary or sufficient condition for the existence of a dictator. Furthermore, one can argue that it is directionally incorrect as the larger the number of Arrowvian dictators, other things equal, the better is the fit between the social ordering and the individual orderings. My point is emphatically not that axiomatic theories should be dispensed with but just that rigor and transparency are overplayed as justifications. Far more persuasive, to my mind, is the claim that because mathematics enables one to describe things more or less efficiently than natural
language, its principal benefit is often one to do with innovation insofar as it enables one to see and demonstrate things that would otherwise be impenetrable.

Second, but related, the value of Arrow’s contribution, like that of the contributions of Von Neumann and Morgenstern to decision theory, is not to be found in the correctness and definitiveness of his Theorem but rather in its technical insights, the range of concepts that it brought together not to mention his discussion of them and the subsequent research literatures that the Theorem inspired. Given the way in which literature has developed, incremental technical progress seems to be more rapid than conceptual progress. Indeed, formal theories can often be literally incorrect and yet provide a sharp indication of where the truth might lie. This is particularly true in any science where the formal model is not meant to be an exact replica of the phenomenon in question. Models of molecular structures are, for example, claimed to be exact whereas models of weather systems are not. Whilst one might not always agree with the detailed formalisation or normative essentialness of particular concepts in Arrow’s Theorem, the criteria of efficiency as measured by the metric of preference, fairness requiring the absence of undue influence, and the notion of decisiveness clearly should be core features of any account of social choice. Furthermore, his theorem speaks to different conceptions of, or bases for, social choice which include voting, preference maximisation and group choice. In some cases, where the group is a household or family for instance, certain axioms, like impartiality may be less appealing than they are in other conceptions of social choice, say political voting.

Finally, an issue arises concerning the lessons drawn from proofs by contradiction and impossibility theorems, not in principle the same though they stand and fall together in Arrow’s Theorem. In this case, the impossibility of satisfying a set of conditions and the possibility of satisfying any subset has a direct value. If you were, for example, to attempt the construction of a voting procedure which satisfied Arrow’s conditions for at least two agents and three alternatives, you would very be frustrated. But there other less direct messages as we have seen. If underlying values, like efficiency, fairness and decisiveness cannot be met completely and simultaneously, then we must accept that more of one will mean less of another. We may have views as to how best the trade-off should be made and this seems to involve an understanding of the different dimensions involved as well as where alternatives lie. Checking the robustness of Arrow’s result by altering the space of possibilities (social orderings) is a necessary response but it cannot tell us anything about human valuations of points within that space. Nor has it encouraged much work along these lines though,
(presumably) the payoff from economic analysis is knowing what is desirable, rather than what is merely possible. This trade-off interpretation of social choice is something that one can take from Arrow’s result, though it was not the interpretation most prominent in his own discussions. It is emerging formally in the work on rights and freedom but will, I venture to suggest, help us to see old issues in a new light – for example cardinality, as being a matter of practical efficiency rather than anything else – and it may even encourage some empirical work.

References (includes some works not cited)
Anand P 2003 The Integration of Claims to Health-Care: A Programming Approach, Journal of Health Economics, 22, 731-45
Barbera S, Pivotal Voters: A New Proof of Arrow’s Theorem, Economics Letters, 6, 13-6
Brown D J, 1975, Aggregation of Preferences, Quarterly Journal of Economics, 89, 456-469
Geanakopolos J 2001 Three Brief Proofs of Arrow’s Impossibility Theorem, New Haven, Cowles Foundation Discussion Paper 1123RRR, Yale University
Gendin S 199x, Why Preference is Not Transitive, *The Philosophical Quarterly*, **54**, 482-88
Miller D
Sen A K, 1979, Personal Utilities and Public Judgements: Or What’s Wrong with Welfare Economics, **89**, 537-58
Sen A K, 1985, *Commodities and Capabilities*, Amsterdam, Elsevier