SYNOPSIS

The problem of democratic aggregation is the problem of aggregating potentially divergent individual inputs (e.g. the 'preferences', 'welfare', or 'sets of beliefs or judgments' of individuals) into corresponding overall collective outputs (e.g. the 'preferences', 'welfare', or 'sets of beliefs or judgments' of a group as a whole), typically for the purpose of guiding collective decisions.

Condorcet's paradox and Arrow's impossibility theorem famously highlight a difficulty raised by democratic aggregation: if the domain of admissible individual preference input is unrestricted, it is possible for pairwise majority voting to generate cyclical collective preferences, and there exists no procedure for aggregating individual preference or welfare orderings in such an unrestricted domain into collective ones in accordance with a set of arguably undemanding minimal conditions -- conditions capturing the requirement, in essence, that collective preference or welfare orderings be both minimally responsive to individual ones and consistent.

A prominent view in social choice theory, defended in particular by William Riker and his followers (e.g. Riker, 1982), is that these and related impossibility results pose a severe threat to the foundations of democracy. It is held that the impossibility results of social choice theory demonstrate the impossibility, instability or arbitrariness of democratic decision procedures, and that they question our very idea of democracy, except under an extremely minimalist interpretation of democracy.

The aim of this thesis is to offer a social-choice-theoretic response to this negative view, and to reassess the implications of Arrowian social choice theory for the problem of democratic aggregation. The thesis focuses on three different 'arenas' in which the problem of democratic aggregation occurs, the first and second ones being the most commonly discussed ones in, respectively, the political science and economics literatures on social choice theory, and the third one being less commonly studied in social choice theory: the arena of 'preferences' (part I), the arena of 'welfare' (part II), and the arena of 'sets of beliefs or judgments' (part III).

The discussion of each arena of democratic aggregation begins with a presentation of the social-choice-theoretic impossibility results that are central to that arena, including several new results. These impossibility results are then, however, not interpreted in the classical 'impossibility' manner, as showing primarily what cannot be done. Rather, they are interpreted as our guides to what can be done, by demarcating the boundaries of the logical space of possibilities. Unlike the
kinds of responses to Arrow's theorem and related results that proceed by claiming either that the entire framework of Arrowian social choice theory is misguided or that some of the conditions underlying these results are simply irrelevant or uncompelling, the present approach strives to take the impossibility results seriously, to take as many of their conditions on board as possible, but nonetheless to identify those escape-routes, if any, to which the results point. In the case of each arena, the technical results are shown to highlight precisely those conditions that need to be met for meaningful and non-arbitrary democratic aggregation to be possible.

The arena of 'preferences', discussed in part I of this thesis, is the classical arena of social choice theory. The aggregation of preferences is exemplified by voting or by any form of committee decision making in which individuals explicitly express -- truthfully or not -- how they would rank the alternative options or candidates, or what they would like the decision to be, and the collective outcome is formed on the basis of those individual expressions of preference, for instance by means of a voting procedure.

In chapter 1, I present Condorcet's paradox, Arrow's original impossibility theorem and the Gibbard-Satterthwaite theorem on strategic manipulability, and discuss the challenge they pose to democratic aggregation in the arena of preferences, focusing especially on Riker's famous normative argument against the meaningfulness of democratic outcomes. I point out that the difficulties posed by aggregation depend crucially on how divergent the preferences expressed by the individuals are. Hence chapters 2 and 3 concentrate on one particular escape-route from the impossibility results: the escape-route through preference structuration. Since Duncan Black's seminal work (1948), it is known that, if the domain of admissible individual preference input is restricted to profiles of individual preference orderings across people that satisfy the structuration condition of single-peakedness, then cycles and the impossibility results disappear (more general conditions of preference structuration are also discussed).

In chapter 2, I argue that Black's result and related results on preference structuration draw our attention to an important distinction between two different concepts of agreement, agreement at a substantive level and agreement at a meta-level, and I suggest that preference structuration is a way in which agreement at a meta-level can surface. Having introduced preference structuration as a sufficient condition for meaningful democratic aggregation in the arena of preferences, I then ask how this condition might be brought about in democratic practice. To address this question, I present the hypothesis that processes of group deliberation, as advocated by deliberative democrats, can induce preference structuration (particularly through bringing about greater
agreement at a meta-level), and thereby open up an escape-route from the social-choice-theoretic impossibility results.

In chapter 3, I develop operational measures of preference structuration, and cite empirical results from a recent study by List, McLean, Fishkin and Luskin (2000) in support of the introduced hypothesis.

Empirical claims about the likelihood of preference structuration, such as the hypothesis that group deliberation can induce preference structuration, involve a deviation from a so-called "impartial culture" assumption, common in social choice theory. On this assumption, any logically possible preference ordering is as likely to be held by an individual as any other. Given an "impartial culture" assumption, the probability of cycles increases as the number of options increases and also as the number of individuals increases. While the hypothesis discussed in chapters 2 and 3 involves a substantial deviation from the situation of an "impartial culture", chapter 4 is concerned with the effects of very slight deviations from an "impartial culture". I present a probability-theoretic model which suggests (giving formal proofs for the three-option case) that, as soon as the distribution of preference orderings amongst individuals is suitably systematically skewed, however slightly, in favour of certain preference orderings and against others (a condition considerably less demanding than preference structuration, and possibly arbitrarily close to an "impartial culture"), then the probability of cycles and preference profiles leading to social choice problems converges to zero as the number of individuals increases.

In the arena of 'welfare', discussed in part II of this thesis, the concern is not the classical committee or voting decision, which is based on the individuals' explicit expression of their rankings of (or top-preferences amongst) the alternative options or candidates. Rather, the concern is the distribution of benefits and burdens in society, the allocation of resources, or the evaluation of alternative socio-economic policy alternatives, on the basis of the 'welfare' of the relevant individuals as assessed by reference to some normatively relevant evaluation standard. Examples of such evaluation standards, or 'currencies of welfare', are money, indices of resources, utility, or Rawlsian primary goods. Famous proposals on how to solve the problem of democratic aggregation in the arena of welfare are classical utilitarianism or Rawls's Theory of Justice.

In chapter 5, I argue that in the arena of welfare, too, Arrow's impossibility theorem at first sight seems to demonstrate the impossibility of finding an attractive general solution to the problem of
democratic aggregation. Arrow's theorem is, however, crucially dependent on excluding certain types of information from the input that is 'fed' into an aggregation procedure. In particular, Arrow's theorem depends on disallowing interpersonal comparisons of welfare or of the normatively relevant evaluation standard. Since Sen's work, it is known that, if sufficient interpersonal comparability is introduced, Arrow's minimal conditions are no longer mutually inconsistent, and utilitarian and Rawlsian aggregation procedures, amongst others, satisfy them. I review some of the central results on the escape-route from Arrow's theorem through introducing interpersonally significant information and discuss their significance for the problem of democratic aggregation in the arena of welfare.

In chapter 6, I show that Arrow's minimal conditions can already be satisfied under much weaker forms of interpersonal comparability than commonly assumed, namely as soon as an interpersonally significant 'zero-line' of welfare (interpretable, for instance, as a poverty line) is introduced.

In chapter 7, I offer a reassessment of the orthodox view in economics that such possibility results are irrelevant because interpersonal welfare information is (purportedly) meaningless. The proposed reassessment of the orthodox view draws on a parallel between the problem of interpersonal comparisons of welfare and the problem of translation of linguistic meaning, as explored by Quine and Davidson. On the orthodox view, interpersonal comparisons of welfare are empirically underdetermined, and "hence" meaningless and unsuitable for providing an escape-route from the impossibility results. Against this view, I argue that, even if we concede that such interpersonal comparisons are empirically underdetermined, this does not imply their impossibility, and I suggest that the underdetermination can be broken non-arbitrarily, albeit in a normative way.

Chapters 5 and 7 aim to show that, as soon as we accept that the choice of a welfare evaluation standard is a normative choice, the escape-route from Arrow's theorem through introducing interpersonally significant information becomes available.

Furthermore, most traditional models of democratic aggregation in the arena of welfare are based on an assumption of onedimensionality, namely the assumption that the welfare of each individual is expressible in terms of a single linear welfare ordering or a single welfare indicator such as income or utility. In chapter 8, I develop a new model which lifts that assumption and generalizes the Arrowian social choice framework, so as to capture the idea that a plurality of
welfare indicators is relevant to the assessment of an individual's welfare. I show that, like Arrow's original framework, the multidimensional generalization is affected by a version of Arrow's impossibility theorem, highlighting not only the threat of dictatorship of a single individual as in Arrow's original case, but also the threat of dominance of a single dimension. But it also turns out that, like Arrow's original impossibility result, its multidimensional generalization depends crucially on certain informational restrictions, namely restrictions on measurability, interpersonal comparability and, importantly, interdimensional commensurability of individual welfare. Again, this opens up an escape-route from the impossibility result through introducing sufficient information. I show that, given both sufficient interpersonal comparability and sufficient interdimensional commensurability, the impossibility result disappears and a broad range of possibilities emerges, some of which can be interpreted as social-choice-theoretic formalizations of prominent proposals in the philosophical literature on utilitarianism and Rawls's Theory of Justice.

Part III, consisting of chapter 10, addresses an arena much less commonly studied in social choice theory: the arena of 'sets of beliefs or judgments'. Often democratic decision making bodies are faced with the need to aggregate individual sets of beliefs or judgments over an entire set of logically interconnected propositions into corresponding collective ones. This need arises when complex policy systems, party or interest group programmes, ideologies or institutional structures are to be designed or chosen, where the various parts of such a system are interconnected and mutually constrain each other, and where consistency is of importance. Classical social-choice-theoretic models are unsuitable for studying this class of aggregation problems, because the classical concern is the aggregation of individual preferences (or welfare) over several separate and mutually unconstrained options or alternatives, with the aim of selecting a single winning outcome or determining a collective ranking of these options. The aggregation of sets of beliefs or judgments, by contrast, requires the simultaneous decision on whether to accept or reject each proposition amongst multiple interconnected propositions, where the acceptance or rejection of some of these propositions logically constrains the acceptance of rejection of others.

I first note that, like the aggregation of preferences or welfare, the aggregation of sets of beliefs or judgments is affected by a paradox as well, the so-called "doctrinal paradox", first identified by scholars in law and economics. I then develop a new general model for formalizing the aggregation of sets of beliefs or judgments, and present an impossibility result by List and Pettit (2000) showing that, if the domain of admissible individual sets of beliefs or judgments is unrestricted (assuming, however, that these individual sets of beliefs or judgments satisfy certain
consistency criteria), there exists no procedure for aggregating individual sets of beliefs or judgments in this domain into collective ones in accordance with a set of minimal conditions similar in spirit to Arrow's minimal conditions in the other two arenas. In analogy to the impossibility results in those arenas, the new impossibility result highlights a tension between two plausible requirements: the requirement that a collective set of beliefs or judgments be minimally responsive to the sets of beliefs or judgments held by individuals and that it be consistent.

A number of escape-routes from the impossibility problem are identified and discussed. As in the arena of preferences, structuration is the key to an escape-route from the impossibility problem. I point out that the difficulties posed by the new impossibility result depend on how divergent the sets of beliefs or judgments held by different individuals are, and I show that, while Black's concept of single-peakedness is not applicable to the arena of sets of beliefs or judgments, an alternative structuration condition can be devised for the latter arena. The new concept is called unidimensional alignment, and I prove that, if the domain of admissible individual belief or judgment input is restricted to profiles of individual sets of beliefs or judgments that satisfy unidimensional alignment, then "doctrinal paradoxes" and the impossibility result disappear. This possibility result also entails a new type of a median voter theorem, applicable to the arena of sets of beliefs or judgments. I also revisit the distinction between agreement at a substantive level and agreement at a meta-level.

Unlike in the case of single-peakedness in the arena of preferences, so far no empirical research has been done on whether processes of group deliberation can induce unidimensional alignment, or on whether there are plausible empirical situations in which the sets of beliefs or judgments held by different individuals satisfy this condition. The conclusions in the third arena must therefore remain more tentative and theoretical. However, as in the previous arenas, here, too, the impossibility problems of social choice theory are the consequence of certain conditions which may, but need not necessarily, obtain. A central condition is again a lack of structuration. To the extent that democratic practice can provide a way around these conditions, there is an escape-route from the impossibility problems.

In conclusion, the view taken by Riker and his followers is tenable only if conditions highly unfavourable to democracy hold, i.e. conditions which block the identified escape-routes from the impossibility results. In the arena of preferences, a substantial, and systematic, lack of preference structuration is necessary (and, as chapter 4 suggests, often not even sufficient) for the occurrence
of preference profiles that generate cycles and that are social-choice-theoretically problematic. Probabilistic reasoning suggests that, under arguably plausible assumptions, cycles and problematic preference profiles are not very likely, and empirical evidence supports the hypothesis that processes of group deliberation may induce the identified forms of preference structuration. In the arena of welfare, the impossibility problems occur only under severe restrictions on the types of individual welfare information that are considered admissible. If all interpersonally significant welfare information (or, in the multidimensional case, all interdimensionally significant welfare information) is disallowed, then we are faced with Arrow-type impossibility problems. But I defend the claim that the fact that such interpersonally (or interdimensionally) significant information is not purely empirical, but that it requires certain additional normative assumptions (which can itself emerge in a democratic process), does not render this information meaningless or arbitrary. And if we accept this claim, the impossibility results not only disappear, but a rich set of possible solutions to aggregation problems becomes available. Finally, in the arena of sets of beliefs or judgments, a lack of structuration is once again a key source of the impossibility problems. Here, too, a sufficient level of structuration, which may be brought about by agreement at a meta-level, provides an escape-route from these problems. However, given the nature and complexity of what is aggregated in this arena, namely sets of beliefs or judgments over an entire set of logically interconnected propositions, the requisite conditions of structuration are more demanding than their counterparts in the arena of preferences. This underlines the onus on group deliberation processes in bringing about these conditions in this third arena, although further empirical research is required to test the success of this escape-route.

An aggregation procedure cannot perform magic. The social-choice-theoretic impossibility results certainly show that if the input to the aggregation is either too diverse (i.e. lacking structuration), or too sparse (i.e. lacking information), the output may not meet all criteria of consistency and responsiveness to the input. The results therefore point precisely to the keys for successful democratic aggregation: structuration, in the case of the first and the third arenas, and information, in the case of the second one. Democratic aggregation is thus not inherently impossible or arbitrary, but its success depends on certain preconditions, particularly structuration or information, and the challenge for democratic practice is to ensure that these preconditions obtain.