

Two Concepts of Agreement¹

Christian List

A central problem of democracy is the aggregation of divergent individual inputs into overall collective decisions. Social-choice-theoretic impossibility results famously demonstrate the intractability of a large class of such aggregation problems. This paper develops a taxonomy of two concepts of agreement, agreement at a substantive level and agreement at a meta-level, and discusses the escape-routes these concepts open up from the impossibility problems of social choice. Specifically, two contexts of democratic aggregation are addressed: first, the familiar context of preferences, and second, the largely unexplored context of sets of judgments over multiple interconnected propositions. Drawing on some recent developments in social choice theory and democratic theory, I will defend the view that, when agreement is conceptualized in democratic theory and when it is sought in democratic practice, more emphasis should be placed on agreement at a meta-level than is commonly done. Finally, I will address the more general question to what extent it is acceptable for the stability of institutions for democratic aggregation to be dependent on specific empirical contingencies.

1. Introduction

A great challenge for democracy is the resolution of conflict between divergent individual preferences, views or interests. Sometimes it is held that democracy is about finding "the will of the people", but if the "wills" of different individuals bear little resemblance to each other, it may be hard to extract anything from the set of individual "wills" that can plausibly be considered "the will of the people".

Condorcet's famous paradox captures some of these problems. If there are three individuals, where one prefers option x to option y to option z , the second prefers option y to option z to option x , and the third prefers option z to option x to option y , then there exist a majority for x against y , a majority for y against z , and a majority for z against x . This means that, if pairwise majority voting is the method for aggregating individual preferences into collective ones, or more grandly, for extracting "the will of the people", then the resulting collective preferences can be cyclical and thus useless for reaching consistent collective outcomes.

More generally, Arrow's celebrated impossibility theorem (1951/1963) shows that, if the domain of admissible individual preference input is unrestricted (the universal domain condition), there exists no procedure for aggregating individual preferences in this domain into collective ones in accordance with a set of arguably undemanding minimal conditions (the weak Pareto principle, independence of irrelevant alternatives, non-dictatorship and transitivity of social orderings) -- conditions capturing the requirement, in a nutshell, that collective preferences be both minimally responsive to individual preferences and consistent.

The difficulties posed by aggregation depend crucially on how divergent the preferences, views or interests of individuals are. In the (rare) limiting case of unanimity the difficulties obviously disappear. If everybody had exactly the same preferences, views or interests, there

¹ Previous versions of this paper were presented at the 2000 conference of the New Zealand division of the Australasian Association of Philosophy, held at Victoria University, Wellington, December 2000, and at the conference *Deliberation and Decision*, held in Wittenberg, Germany, September 2001, and accepted for publication in *PEGS: The Good Society*. The author wishes to express his gratitude to Stephen Elkin for his encouragement to write this paper, to Geoffrey Brennan, John Dryzek, Robert Goodin, David Miller, Philip Pettit and the participants of the conference panel in Wellington for many useful discussions, as well as to the German National Merit Foundation for financial support. The work on this paper was done while the author was a Visiting Scholar at the Center for Basic Research in the Social Sciences at Harvard University and at the Department of Linguistics and Philosophy at the Massachusetts Institute of Technology. Address for correspondence: Christian List, Nuffield College, Oxford OX1 1NF, U.K.; E-mail christian.list@nuffield.oxford.ac.uk.

would be no conflict to resolve. But while unanimity is sufficient for the disappearance of the famous Condorcet and Arrow aggregation problems, it is not necessary. Since Duncan Black's seminal work (1948), it is well known that Condorcet's paradox can be traced back to a 'lack of structure' in the relevant set, also called *profile*, of individual preferences across individuals. Black himself proved that *single-peakedness*, a structure condition to be discussed more formally below, (jointly with the (harmless) technical condition that the number of individuals is odd) is a sufficient (though not necessary) condition for the avoidance of Condorcet paradoxes. A well-known corollary of Black's insight is that Arrow's impossibility result ceases to hold if the domain of admissible individual preference input is restricted to profiles of preferences satisfying single-peakedness.

At one level, Black's result seems only to confirm what we already think we know, namely that, if the disagreement between different preferences, views or interests stays within certain limits -- limits that are somehow transcended in situations like the one of the Condorcet paradox --, then familiar methods of democratic decision making, like pairwise majority voting, are available for reaching collective outcomes in reasonably defensible and logically consistent ways. But Black's result teaches us much more than that. It highlights an important distinction between two different concepts of agreement. The two concepts are what we might call *agreement at a substantive level* and *agreement at a meta-level*. It is this distinction that the present paper will explore. The paper discusses two contexts of democratic aggregation: first, the context of preferences, and second, the context of sets of judgments over multiple interconnected propositions. Sections 2 and 3 explore the two concepts of agreement in the first context, section 4 introduces the less familiar context of sets of judgments over multiple interconnected propositions, and sections 5 and 6 explore the two concepts of agreement in this second context. In particular, drawing on some recent developments in social choice theory and democratic theory, I will defend the view that, when agreement is conceptualized in democratic theory and when it is sought in democratic practice, more emphasis should be placed on the concept of agreement at a meta-level than is commonly done. Section 7, finally, takes a step back and addresses the question to what extent it is acceptable for the stability of institutions for democratic aggregation to be dependent on specific empirical contingencies.

2. Agreement at a Substantive Level and Unanimity

Two or more individuals agree *at a substantive level* to the extent that their preferences or views are the same. Perfect *agreement at a substantive level* is the case of identical preferences or views across different people.

In response to the problems of democratic aggregation, it is natural to argue that processes of political deliberation should be encouraged with the aim of somehow reducing the level of conflict between different people's preferences or views and bringing about greater agreement at a substantive level. In his account of the idea of deliberative democracy, Jon Elster summarizes this view quite succinctly: "The core of the theory [of deliberative democracy] ... is that rather than aggregating or filtering preferences, the political system should be set up with a view to changing them by public debate and confrontation. The input to the social choice mechanism would then not be the raw, quite possibly selfish or irrational, preferences ..., but informed and other-regarding preferences. Or rather, there would not be any need for an aggregation mechanism, since a rational discussion would tend to produce unanimous preferences." (Elster, 1986, p. 112)

If successful, the view outlined by Elster may seem attractive. There are, however, at least two problems with this view. The first, and practical, problem is that the idea that democratic

deliberation would produce unanimity may be unrealistic in many circumstances. People may agree on all relevant facts, and arguments, concerning different political options, and yet disagree on their most preferred option. They may agree on what the environmental effects of a new industrial development would be, and yet disagree on whether these effects should be accorded more weight than the expected economic benefits of the new development. In a recent empirical study, Cass Sunstein showed that deliberation, especially if it takes place in clustered homogeneous groups whose members reinforce each other's views, can sometimes even lead to polarization of opinion across different such groups rather than convergence (Sunstein, 2000). Although none of this is to deny the benefits of agreement at a substantive level -- *if it can be reached* --, it should be clear that democracy needs to have alternative resources for dealing with conflicts of preferences or views in those (plausible and potentially numerous) cases in which deliberation fails to produce unanimity.

The second, and somewhat more theoretical, problem with the view outlined by Elster is that it is unclear whether *convergence towards* agreement at a substantive level (falling short of perfect unanimity) is the most promising strategy for avoiding Condorcet and Arrow aggregation problems. William Gehrlein devised social-choice-theoretic measures of *social homogeneity* capturing basically the question of how closely a given profile of preferences across individuals approximates agreement at a substantive level. Using these measures, he showed that, although there is a positive connection between the level of social homogeneity and the avoidance of Condorcet problems, the connection is much weaker than what one might have hoped to find (Gehrlein, 2000).

3. Agreement at a Meta-Level and Single-Peakedness

Black's insight is to ask not whether two or more individuals have the same preferences over a set of options, but rather whether there exists an ordering of these options from 'left'-most to 'right'-most such that each individual has a most preferred position on that 'left'/'right' dimension and prefers options less and less as these options get more and more distant from his or her most preferred position. If there exists a 'left'/'right' ordering of the options with this property, called a *structuring dimension*, then we say that the given profile of preferences across individuals satisfies *single-peakedness*.

The terms 'left'/'right' are used here in a purely 'geometrical' sense. Any ordering of the options along which individual preferences are systematically aligned in the requisite way could serve as a structuring dimension, whether it orders the options from most urban to most rural, from most secular to most religious, from most architecturally avant-garde to most architecturally conservative, or in any other, however esoteric, way.

Black's concept inspires the following definition. Two or more individuals agree *at a meta-level* to the extent that they agree on a common dimension in terms of which an issue is to be conceptualized. They may reach perfect agreement at a meta-level while at the same time disagreeing substantively on what the most preferred position on that dimension is.

Single-peakedness may thus be an implication of *agreement at a meta-level*. *If* the individuals agree on a common structuring dimension along which each individual's preferences are systematically aligned in the requisite way, *then* the profile of preference orderings across these individuals satisfies single-peakedness. However, since single-peakedness is only a formal structure condition on a profile of preference orderings across individuals, single-peakedness is logically less demanding than agreement at a meta-level. The latter is sufficient, but not necessary, for the former. A profile of preference orderings across individuals may

happen to have the right *formal* structure for satisfying single-peakedness without the individuals *semantically* conceptualizing the issue in terms of the same common dimension.

Further, single-peakedness is not merely a consistency condition on individual preferences. Take the preference ordering of a single individual (over a finite set of options). Unless we make reference to a specific structuring dimension, the question of whether this preference ordering is single-peaked is vacuous or, to be precise, must vacuously be answered in the affirmative. For we can simply define this individual's most preferred option to be the 'left'-most option, his or her least preferred option to be the 'right'-most option, positioning intermediate options from 'left' to 'right' in a decreasing order of preference (supposing the preference ordering is strict). With respect to this artificially constructed structuring dimension, the given individual's preference ordering is clearly -- but of course uninformatively -- single-peaked. The concept of single-peakedness becomes non-vacuous only when applied to the preferences of multiple individuals, for it is in those cases that we can ask whether or not different individuals' preferences are systematically aligned *along the same common structuring dimension*. This is the sense in which single-peakedness truly captures a formal implication of *agreement*, albeit at a meta-level.

Now Black proved the following surprisingly simple, and yet ingenious result. Given a profile of preferences across individuals that satisfies single-peakedness, order the individuals (let us assume for simplicity that there is an odd number of them) from 'left'-most to 'right'-most in terms of their most preferred position (their 'peak') on the corresponding structuring dimension. With respect to this 'left'/'right' ordering of the individuals, the *median individual* is simply the one who has as many individuals to the left as he or she has to the right. Then the most preferred option of the median individual will beat, or at least tie with, all other options in pairwise majority voting. A simple corollary of this result is that, *if* the domain of admissible individual preference input consists only of profiles of preference orderings satisfying single-peakedness, *then* pairwise majority voting is a procedure for generating collective preferences in accordance with Arrow's minimal conditions of responsiveness and consistency. Moreover, studies by Niemi (1969) as well as Tullock and Campbell (1970) have shown that consistent social preferences, in accordance with these minimal conditions, are likely to exist if only 75% or even fewer of the individuals have preferences that are systematically aligned along the same common structuring dimension.

Black's result suggests an alternative response to the challenge of democratic aggregation. Rather than seeking convergence towards agreement at a substantive level, which may be hard to achieve, we might seek convergence towards agreement at a meta-level. A recent strand of deliberative democratic thinking advocates precisely this idea (see Miller, 1992, and Dryzek and List, 2000). The idea is that the key to a deliberative democratic response to Condorcet and Arrow lies not in the idea of deliberation-induced agreement at a substantive level, but rather in the idea of deliberation-induced agreement at a meta-level. As the proponents of this idea emphasize, it is much more realistic, if nonetheless still demanding, to expect processes of political deliberation to produce agreement on what the relevant *questions* are rather than on what the *answers* should be.

Specifically, the idea can be stated as a three-part hypothesis:

- (1) that group deliberation leads people to identify a single shared issue-dimension in terms of which the issue at stake is to be conceptualized;
- (2) that, for a given issue-dimension, group deliberation leads people to agree on the position of each (policy) option on that dimension; and

- (3) that, once an issue-dimension has been identified as relevant, group deliberation leads each individual to determine a most preferred position (his or her 'peak') on that dimension, with decreasing (dimension-specific) preference as options are increasingly distant from that most preferred position.

Neither part of the hypothesis is trivial, and since each part raises difficult social-psychological issues, I am not able to address these questions properly here. To give an intuitive illustration of the issues at stake, suppose in over-simplified terms that (i) the question of what issue-dimension is relevant to a given democratic decision problem is a normative question, (ii) the question of where options are located *on a given issue-dimension* is (often or at least sometimes) a factual question, and (iii) the question of what structure *dimension-specific* preferences of an individual should take is a question of rationality. Under this supposition, the success of part 2 depends on whether group deliberation can bring about agreement on factual matters, for instance by clarifying and supplying information; and the success of part 3 depends on whether group deliberation can induce greater rationality in individuals. While both of these requirements are not undemanding, they are clearly not completely implausible. The success of part 1, on the other hand, depends on whether group deliberation can bring about agreement on normative matters, a much more demanding requirement in many cases. The combination of parts 1, 2 and 3 is, in essence, the hypothesis that group deliberation can bring about agreement at a meta-level, which then surfaces in the form of single-peakedness.

In some cases, an appeal to public reasons and generalizable interests, characteristic of the kind of group deliberation deliberative democrats advocate, may lead to the identification of a single publicly relevant issue-dimension, and in consequence to agreement at a meta-level, the implication of which might be single-peakedness. But in other cases even the invocation of public reasons and generalizable interests may not have this effect, because individuals might still disagree about what is in the public interest, or whether, for instance, ecological integrity or economic growth should be given priority when such interests are in conflict.

Even rational choice theorists of a more traditional orientation are likely to agree that parts 2 and 3 of the hypothesis are relatively undemanding, and that the demanding and difficult part of the hypothesis is part 1. Mueller (1989, pp. 89-90), for example, argues, "[g]iven that we have a single-dimensional issue, single-peakedness does not seem to be that strong an assumption. What is implausible is the assumption that the issue space is one dimensional".

Obviously, the question of whether the hypothesis that deliberation induces single-peakedness is empirically adequate cannot be answered by pure theorizing. List, McLean, Fishkin and Luskin (2000) have used data from Fishkin and Luskin's deliberative polls (on this method, see Fishkin, 1997, and a Symposium in *PEGS*, 9 (2), 1999) to test the hypothesis empirically. They studied a set of deliberative polls on topics of energy provision in Texas as well as a recent poll on the future of the monarchy in Australia. In these polls, participants were first confidentially interviewed on their opinions, then invited to participate in a weekend of intense group deliberation, and finally confidentially interviewed again, being asked exactly the same questions as in the pre-deliberation interviews. The study showed that the post-deliberation levels of single-peakedness were either strictly greater than the corresponding pre-deliberation levels (in *every* Texan poll in the set) or at least on a par with them (in the Australian poll).

Of course, these findings are not the final word on the introduced hypothesis. There may be situations that are favourable to its success, such as situations in which people, on reflection,

agree that a certain single issue-dimension (for instance, an ecological one) is particularly salient and should guide their conceptualization of the decision problem. But there may also exist situations less favourable to its success, such as situations in which a decision problem is perceived to be so inherently multidimensional, or in which people's value priorities on different issue-dimensions are so fundamentally different, that they neither agree on what the correct answer is, nor even on how to look at the problem.

4. From Preferences to Sets of Judgements

The discussion so far has been centred around the problem of aggregating individual *preferences over a set of options* into corresponding collective ones. Often, however, democratic decision making bodies are faced with the need to aggregate individual *sets of judgements over multiple interconnected propositions* into corresponding collective ones. This need arises when complex policy systems or institutional structures are to be designed or chosen, where the various parts of a policy system or institutional structure constrain each other and consistency is of great importance.

A simple example will serve to illustrate the problem. Suppose the following three policy proposals are simultaneously under discussion:

- P : to introduce a special tax on gasoline (in order to provide an incentive to reduce the amount of 'unnecessary' driving and thereby to reduce the level of pollution);
 Q : to abolish a subsidized commuter railway system;
 R : to give special tax benefits to commuters (in order to compensate for their commuting expenses).

Suppose further that everyone accepts that proposal R should be accepted *if and only if* both proposal P and proposal Q are accepted -- formally, $(R \leftrightarrow (P \wedge Q))$ --, for the following reason: If the special gasoline tax is introduced (P) and the subsidized commuter railway system abolished (Q), then society will impose an exceptionally great financial burden on commuters, so that commuters should be entitled to some compensation (R). On the other hand, if the special tax on gasoline is not introduced, or if the subsidized commuter railway system remains in service, then nobody needs compensation.

For simplicity, let there be three individuals in the relevant decision making body, with the following sets of judgments on the three proposals (P , Q , R) and on the proposition describing their logical interconnection ($(R \leftrightarrow (P \wedge Q))$).

	individual 1	individual 2	individual 3
P	yes	yes	no
Q	yes	no	yes
R	yes	no	no
$(R \leftrightarrow (P \wedge Q))$	yes	yes	yes

Table 1

Note that each of the three individuals holds a perfectly consistent set of judgments in light of the proposition that R should be accepted if and only if P and Q are both accepted. Now if the individuals choose to determine their collective set of judgments on the basis of majority voting on each proposition, then they will accept both P and Q (each by majorities of 2 out of 3), as well as $(R \leftrightarrow (P \wedge Q))$ (the latter proposition unanimously), but reject R (by a majority

of 2 out of 3). But this is an inconsistency, given the unanimous agreement that the acceptance of both P and Q is a necessary and sufficient condition for the acceptance of R .

In short, propositionwise majority voting over a set of logically connected propositions can lead to inconsistent collective sets of judgements, even when all individuals hold perfectly consistent sets of judgements. This problem is a version of what is sometimes called the doctrinal paradox (see, for example, Kornhauser, 1992, Chapman, 1998, Pettit, 2001, and Brennan, 2000).

Generalizing this insight, List and Pettit (2001) have shown that, if the domain of admissible individual sets of judgements is unrestricted (so long as these individual sets of judgements satisfy basic consistency criteria) (the universal domain condition), there exists no procedure for aggregating individual sets of judgements in this domain into collective ones in accordance with a set of minimal conditions similar in spirit to those proposed by Arrow (anonymity, systematicity; and completeness, consistency and deductive closure of collective sets of judgments) -- conditions capturing the requirement that collective sets of judgements be both minimally responsive to individual ones and consistent.

Once again, the difficulties posed by this result depend on how divergent the sets of judgements held by different individuals are. I will now show that the two different concepts of agreement we have identified in the context of preferences can also be identified in the context of sets of judgements and that they here, too, point towards two different responses to the problem of aggregation. I will also argue that John Rawls's idea of an *overlapping consensus* (Rawls, 1993) can be seen as a special case of the concept of *agreement at a substantive level* and that it thus differs from the concept of *agreement at a meta-level*.

5. Agreement at a Substantive Level and Overlapping Consensus

Perfect *agreement at a substantive level* in the context of sets of judgements is the case of identical sets of judgements across different people. As before, one might try to solve problems of aggregation in the context of sets of judgements by encouraging processes of deliberation with the aim of bringing about greater agreement at a substantive level. We have noted in the context of preferences that agreement at a substantive level is a rather demanding condition that may be hard to attain in practice. Now, if we are dealing not with preferences over individual options, but with sets of judgements over an entire set of logically connected propositions, agreement at a substantive level may be even harder to attain.

But there is one special case of the concept of agreement at a substantive level that may seem somewhat more realistic, if still demanding. This special case is a version of John Rawls's concept of an *overlapping consensus*. Often the sets of judgments on fundamental moral and political issues held by different individuals are mutually incompatible, in so far as these sets of judgments represent genuinely different comprehensive views of morality, including different supporting reasons even for those judgments the individuals agree on (like "killing human beings is bad"). However, in a less fundamental and more pragmatic realm of issues, there may actually be much more agreement. Given someone's overall set of moral and political judgments, this overall set will contain some judgments that the individual considers fundamental, maybe even 'axiomatic', and others that he or she considers less fundamental, maybe more pragmatic and applied. Now it is perfectly possible that the different (possibly mutually inconsistent) fundamental judgments held by different individuals may nonetheless have certain *implications* in common at a less fundamental level. This is logically possible because, if there is a relation of logical entailment between someone's fundamental and his or

her less fundamental judgments, this relation is usually a one-way relation. The more fundamental judgments may be logically sufficient to determine the less fundamental ones, but not logically necessary. The former ones may logically entail the latter, but not the other way round. The less fundamental and more applied judgments are usually logically *insufficient* to tell us what the underlying more fundamental judgments are. For instance, consider how many different fundamental reasons might be adduced to support certain welfare provision arrangements. One might believe in certain values of socio-economic justice and equality for liberal reasons or, alternatively, for religious reasons. Or one might believe that the main objective of the state is to keep the streets safe, to prevent crime and to ensure social stability, and that welfare provision arrangements are the only way to achieve this. Or consider how many different fundamental reasons one might have for wanting to implement specific policies for protecting the environment. One might believe in certain rights of future generations; or one might believe that non-human animals have rights that ought to be respected, or that ecosystems have intrinsic value. Or one might believe that environmental disasters would ruin the economy and that the only way to secure a successful economy in the long term would be to implement environmental policies. Many other examples could be given. An overlapping consensus requires the identification of a certain set of policy propositions (usually a proper subset of the set of those propositions that would make up an entire comprehensive view of morality) such that, with respect to these propositions, it is feasible to reach agreement at a substantive level, even if different individuals endorse the same judgments for different underlying reasons.

It is an open question whether, and in what circumstances, processes of political deliberation can bring about an overlapping consensus that is sufficiently broad to give rise to a non-trivial collective set of judgments. In this brief discussion, however, two points should have become clear. First, an overlapping consensus in the sense of agreement at a substantive level over a *restricted* set of propositions is a less demanding condition than agreement at a substantive level over *all* propositions. Second, the concept of an overlapping consensus is different from the concept of agreement at a meta-level. In an overlapping consensus, the individuals agree on certain judgments, without necessarily agreeing on the supporting reasons for these judgments. They agree on certain answers, without necessarily agreeing on what the more fundamental issues or questions are. In a sense, they agree at a substantive level, albeit with respect to a restricted realm of issues, without necessarily agreeing on any meta-theoretical foundations for their substantive agreement.

6. Agreement at a Meta-Level and Unidimensional Alignment

Agreement at a meta-level is defined as agreement on a common dimension in terms of which an issue is to be conceptualized. In the context of preferences, I have argued that the structure condition of single-peakedness may be seen as an implication of *agreement at a meta-level*, in the sense that *if* the individuals agree on a common structuring dimension along which each individual's preferences are systematically aligned in the requisite way, *then* the profile of preference orderings across these individuals satisfies single-peakedness.

While Black's concept of single-peakedness is not straightforwardly applicable to the context of sets of judgments, we will now see that an alternative structure condition can be devised for the latter context (List, 2001), and further that the new structure condition, like single-peakedness, can be seen as an implication of agreement at a meta-level.

In analogy to the context of preferences, the question is not whether two or more individuals hold the same sets of judgements, but this time it is whether there exists a single ordering of

the individuals from 'left'-most to 'right'-most such that, for *every* proposition in the relevant domain of issues, the individuals accepting that proposition are either all to the left, or all the right, of those rejecting it. If there exists a 'left'/right' ordering of the individuals with this property, once again called a *structuring dimension*, then we say that the given profile of sets of judgements across individuals satisfies *unidimensional alignment*. It is easily seen that the judgements in table 1 above violate unidimensional alignment. No matter how the individuals are reordered from 'left' to 'right', it is impossible to get the required pattern of acceptance and rejection. By contrast, the judgements of the five individuals in table 2 below satisfy unidimensional alignment: there exists a single 'left'/right' ordering of the five individuals (namely: 3, 2, 5, 4, 1) with respect to which, for *every* proposition, the individuals accepting the proposition are all to the left (they could also be all to the right) of those rejecting it.

	individual 3	individual 2	individual 5	individual 4	individual 1
P	yes	no	no	no	no
Q	yes	yes	yes	no	no
R	yes	no	no	no	no
$(R \leftrightarrow (P \& Q))$	yes	yes	yes	yes	no

Table 2

Now, given a profile of sets of judgements across individuals that satisfies unidimensional alignment, order the individuals (again assume for simplicity that there is an odd number of them) on a structuring dimension. Then the set of judgements of the median individual with respect to the structuring dimension will be accepted in propositionwise majority voting (in the case of table 2, the judgements of individual 5). And provided that the set of judgements of each individual, most importantly the median individual, satisfies the relevant consistency criteria, so will the collective set. A corollary of this result is that, if the domain of admissible profiles of sets of judgements across individuals consists only of those profiles satisfying unidimensional alignment, then propositionwise majority voting is a procedure for generating collective sets of judgements in accordance with the minimal conditions of responsiveness and consistency used in the impossibility theorem by List and Pettit (2001).

The claim that unidimensional alignment is an implication of agreement at a meta-level may seem less straightforward than the analogous claim in the case of single-peakedness, but here is a way of making it plausible. Suppose, firstly, that there is a single common issue-dimension in terms of which all the propositions are conceptualized by the individuals, and suppose that each individual takes a certain position on that dimension. For simplicity, we will call it a 'left'/right' dimension, but a range of interpretations is possible. And suppose, secondly, that, for each proposition, the extreme positions on the 'left'/right' dimension correspond to either clear acceptance or clear rejection of this proposition and there exists an 'acceptance threshold' on the dimension (possibly different for different propositions) such that all the individuals to the left of the threshold accept the proposition and all the individuals to its right reject it (or vice-versa). If these two conditions are met, then we have a situation of unidimensional alignment. As in the case of single-peakedness, unidimensional alignment requires no agreement at a substantive level: in the case described by table 4, for example, individual 3 and individual 1 disagree about *every* proposition. Unidimensional alignment requires only the existence of a common 'left'/right' ordering of the individuals that systematically structures their pattern of acceptance and rejection over the various propositions.

Once again, the claim is only that unidimensional alignment may be an *implication* of agreement at a meta-level, not that unidimensional alignment is by itself sufficient for

agreement at a meta-level. A profile of sets of judgments across individuals may happen to have the right *formal* structure for satisfying unidimensional alignment without the individuals *semantically* conceptualizing all propositions in terms of the same common issue-dimension.

Moreover, like single-peakedness, unidimensional alignment is not merely a consistency condition on individual sets of judgments. The singleton profile, consisting only of the set of judgments of a single individual, always trivially -- but of course uninformatively -- satisfies unidimensional alignment. Like the concept of single-peakedness, the concept of unidimensional alignment becomes non-vacuous only when applied to the sets of judgments of multiple individuals. In this sense unidimensional alignment captures a formal implication of *agreement*, albeit again at a meta-level.

Unlike in the case of single-peakedness, no empirical research has been done on whether group deliberation can induce unidimensional alignment, or indeed on whether there are ever plausible real-world situations of sufficiently realistic complexity in which a profile of sets of judgments across individuals satisfies unidimensional alignment. But the mere observation that agreement at a substantive level may be hard, if not impossible, to attain in cases of complex and controversial issues and that unidimensional alignment is at least *less* demanding than full agreement at a substantive level should lead us to take the concept of unidimensional alignment seriously.

7. Empirical Contingencies and Institutional Design

A critic might be unconvinced by the claim that agreement at a meta-level and corresponding structure conditions such as single-peakedness or unidimensional alignment open up attractive escape-routes from the paradoxes and impossibility results of aggregation. Specifically, the critic might argue as follows. He or she would concede that, *if* empirical circumstances are such that individual preferences or sets of judgments satisfy the identified structure conditions (or approximate them to a sufficient degree), *then* aggregation procedures like pairwise majority voting or propositionwise majority voting will generate consistent collective outcomes. But, as soon as empirical circumstances are different, the very same aggregation procedures will fail to generate consistent collective outcomes. Social choice theorists can even make precise predictions about when such collective inconsistencies will occur. In other words, what we know about these aggregation procedures is that they work satisfactorily in some empirical circumstances (e.g. for certain profiles of individual preferences or sets of judgments) but not in others. The impossibility theorems further tell us that this problem is not just an artefact of pairwise or propositionwise majority voting, but that it is more general. It cannot be resolved by any aggregation procedure, except at the expense of some seemingly attractive minimal conditions. One would not like to design a house just on the basis of the observation that there are some empirical circumstances in which the house would be stable, while there are others in which the house would collapse. Rather, one would seek to design a house on the basis of solid physical calculations confirming its stability. Analogously, the critic argues, it is a risky *ad hoc* response to the problems of democratic aggregation to use aggregation procedures like pairwise or propositionwise majority voting and to rely on the observation that there are empirical circumstances -- like situations of agreement at a meta-level -- in which these procedures work satisfactorily, while deemphasizing the existence of other circumstances in which they fail to do so. Like a house, procedures for democratic aggregation should be designed so as to work satisfactorily in *all* relevant circumstances (e.g. for all logically possible profiles of individual preferences or sets of judgments). They should not have to rely on the hope that certain empirical circumstances rather than others obtain.

They should be guaranteed to generate consistent collective outcomes, no matter what the empirical circumstances are. So far the critic's objection.

The general question this objection highlights is the following. Should political institutions, such as aggregation procedures, be designed in such a way as to work robustly under all possible empirical circumstances (at least within a relevant domain of circumstances), or is it acceptable for such institutions to rely on specific empirical contingencies that are exogenous to, and not guaranteed by, them?

At first sight, the critic's objection clearly has some force. After all, institutions that rely on specific empirical contingencies may seem prone to erratic behaviour, just as the house whose stability can be guaranteed only under specific empirical circumstances may seem a dangerous structure. However, on closer inspection, the objection loses some of its force. Even the most well designed house, designed on the basis of solid physical calculations, will collapse under some circumstances, for example if there is a sufficiently strong earthquake. Strictly speaking, it is therefore not true that a well designed house will be stable in *all* circumstances. The critic might respond that building standards take that problem into account too. In earthquake zones like California, unlike in Britain, houses are required to meet particular building standards to withstand earthquakes. But, if an exceptionally strong earthquake were to occur, even a house built in accordance with those rigorous standards might collapse.

The observation that can be made from these points is this. Whenever something is to be designed, whether it is a house or an institution, there is a domain of possible circumstances the house or the institution may be confronted with. In some of these circumstances (call them type 1) the house or the institution may have perfect stability, while in others (call them type 2) the house may collapse or the institution may fail to work satisfactorily. Rather than requiring the design of a house or institution for which there are no circumstances of type 2, which may be difficult if not impossible, a reasonable strategy would be the following. First consider the probability distribution over the relevant domain of circumstances, and then ask whether the probability of circumstances of type 2 is sufficiently low (if still nonzero) for the risk of instability under those circumstances to be bearable. For example, even in Britain the probability of the occurrence of a massive earthquake is presumably nonzero. But as this probability is low, it is considered acceptable for houses in Britain not to be built to withstand massive earthquakes. In California, the probability distribution is different, and building standards are adjusted accordingly. Generally, if stability across *all* possible circumstances is difficult or impossible, the design of a house or an institution would have to be adjusted to the probability distribution over the relevant domain of circumstances. The idea would be to design a house or an institution in such a way that the probability of corresponding circumstances of type 1 is high and the probability of corresponding circumstances of type 2 is low.

If discursive conditions in a group of individuals are such that (a sufficient degree of) agreement at a meta-level and corresponding structure conditions are extremely probable, while the occurrence of profiles of preferences or sets of judgments leading to collective inconsistencies is extremely improbable, then the use of aggregation procedures like pairwise or propositionwise majority voting seems as defensible as the construction of houses in Britain which do not meet Californian building standards (under arguably undemanding assumptions about the probability distribution over profiles of preferences, it can be shown that the probability of collective inconsistencies decreases towards 0 as the number of individuals increases; see the appendix on the probability of cycles in List and Goodin, 2001).

If discursive conditions are different, on the other hand, then the use of such aggregation procedures is much less defensible. The defensibility of an aggregation procedure may thus be sensitive to the relevant probability distribution over the set of all logically possible profiles of preferences or sets of judgments. This probability distribution will itself depend on certain empirical facts about the relevant context and the relevant group of individuals. Moreover, additional institutions, like deliberation settings prior to the use of an aggregation procedure, might be designed with the aim of transforming that probability distribution so as to reduce the probability of circumstances of type 2. The defensibility of the suggested aggregation procedures, then, is no longer an *a priori* matter (as typically viewed by classical social choice theorists), but dependent on certain empirical contingencies, which might differ from context to context.

8. Concluding Remarks

The distinction between agreement at a substantive level and agreement at a meta-level was motivated by Black's response to Condorcet and Arrow. And, as we have seen in the context of preferences, the context of Black's own work, there are both theoretical and empirical results in support of the view that the 'meta-agreement' strategy -- the idea that a democracy should seek agreement at a meta-level through processes of political deliberation -- is a promising response to the challenge of resolving conflict between divergent individual preferences, given appropriate empirical circumstances.

In the context of sets of judgements over multiple interconnected propositions, on the other hand, our conclusions must remain much more tentative. Again, the two concepts of agreement can be identified, and their logical status, even in terms of avoiding impossibility problems of aggregation, is very similar to that of their more well known counterparts in the context of preferences. To determine whether a 'meta-agreement' strategy is practically available in this new context, however, remains an open question for democratic theory and a challenge for democratic practice.

In short, many important questions remain open. Most importantly, we will need to tackle the question of whether a certain 'base level of cohesion' -- for instance, in the form of (partial) single-peakedness or (partial) unidimensional alignment -- is absolutely necessary for any strategy of democratic conflict resolution to be successful, or whether democratic conflict resolution mechanisms can be made so robust as to cope with even the most extreme limiting cases of divergence. The impossibility theorems of social choice would certainly seem to apply in the latter cases.

Bibliography

- Arrow, K. (1951/1963) *Social Choice and Individual Values*, New York: Wiley.
- Black, D. (1948) "On the Rationale of Group Decision-Making," *Journal of Political Economy* 56, 23-34.
- Brennan, G. (2000), "Collective Coherence and Belief," *International Review of Law and Economics*, forthcoming.
- Chapman, B. (1998) "More Easily Done than Said: Rules, Reason and Rational Social Choice," *Oxford Journal of Legal Studies* 18, 293-329.
- Dryzek, J., and List, C. (2000) "Social Choice Theory and Deliberative Democracy: A Reconciliation", Proceedings of the American Political Science Association, September 2000.

- Elster, J. (1986) "The Market and the Forum," in Elster, J., and Hylland, A. (eds.) *Foundations of Social Choice Theory*, Cambridge: Cambridge University Press.
- Fishkin, J. S. (1997) *The Voice of the People: Public Opinion and Democracy*, New Haven/London: Yale University Press.
- Gehrlein, W. (2000) "Social Homogeneity and Condorcet Winners: A Weak Connection," paper presented at the annual meeting of the Public Choice Society, Charleston, South Carolina, March 2000.
- Kornhauser, L. A. (1992) "Modelling Collegial Courts. I. Path-Dependence," *International Review of Law and Economics* 12, 169 - 185.
- List, C. (2001) *Mission Impossible? A Social-Choice-Theoretic Investigation into the Problem of Democratic Aggregation*, DPhil-thesis, University of Oxford.
- List, C., and Goodin, R. E. (2001) "Epistemic Democracy: Generalizing the Condorcet Jury Theorem," *Journal of Political Philosophy*, forthcoming.
- List, C., McLean, I., Fishkin, J., and Luskin, R. (2000) "Can Deliberation Induce Greater Preference Structuration? Evidence from Deliberative Opinion Polls," Proceedings of the American Political Science Association, September 2000.
- List, C., and Pettit, P. (2001) "Aggregating Sets of Judgments: An Impossibility Result," *Economics and Philosophy*, forthcoming.
- Miller, D. (1992) "Deliberative Democracy and Social Choice," *Political Studies* 40 (special issue), 54-67.
- Mueller, D. (1989) *Public Choice II*, Cambridge: Cambridge University Press.
- Niemi, R. G. (1969) "Majority Decision-Making with Partial Unidimensionality," *American Political Science Review* 63, 488-497.
- Pettit, P. (2001) "Deliberative Democracy and the Discursive Dilemma," *Philosophical Issues* 11.
- Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.
- Sunstein, C. R. (2000) "The Law of Group Polarization," paper presented at the conference Deliberating about Deliberative Democracy, University of Texas at Austin, February 2000.
- Tullock, G., and Campbell, C. D. (1970) "Computer Simulation of a Small Voting System," *Economics Journal* 80, 97-104.