

Deliberation and Agreement

Christian List¹

Abstract. How can collective decisions be made among individuals with conflicting preferences or judgments? Arrow's impossibility theorem and other social-choice-theoretic results suggest that, for many collective decision problems, there are no attractive democratic solutions. In response, deliberative democrats argue that group deliberation makes collective decisions more tractable. How can deliberation accomplish this? In this chapter, I explore the distinction between two different types of agreement and discuss how they can facilitate collective decision making. Deliberative democrats have traditionally defended the hypothesis that group deliberation generates consensus on what decision option should be chosen: substantive agreement. But substantive agreement is not only empirically unrealistic, but also logically unnecessary for meaningful collective decision making. An alternative, less demanding hypothesis is that, under certain conditions, group deliberation generates consensus on how a decision problem should be conceptualized: meta-agreement. I assess the latter hypothesis, explain how it can be empirically tested, and cite some evidence consistent with it. My discussion addresses two contexts of democratic decision making: preference aggregation and judgment aggregation.

1. Introduction

A central problem of democracy is decision making among individuals with conflicting preferences or judgments. Democracy is sometimes thought to be about finding 'the will of the people', but if different individuals have radically different 'wills', it may be hard to extract from these 'individual wills' a consistent 'popular will'. To illustrate, imagine a group of people in which a third prefers option x to option y to option z, a second third prefers option y to option z to option x, and the last third prefers option z to option x to option y. Then a majority prefers x to y, a majority prefers y to z, and a majority prefers z to x – a 'cyclical' majority preference. So the plausible democratic procedure of pairwise majority voting fails to generate a consistent 'popular will' here. This is Condorcet's paradox.

More generally, Kenneth Arrow's impossibility theorem (1951/1963) shows that this problem is not just an artefact of majority voting. Consider a group of individuals, a committee, legislature, or perhaps entire society. Suppose we want to find a procedure for aggregating the preferences of these individuals into corresponding preferences for the group as a whole. And suppose we want our democratic procedure to satisfy some minimal conditions. First, the procedure should accept as its admissible input all possible combinations of individual preferences ('universal domain'). Second, whenever all individuals agree that option x is preferable to option y, the procedure should respect this unanimous preference ('the weak Pareto principle'). Third, the social preference over any two options x and y should depend only on individual preferences over x and y, and not on individual preferences over third alternatives, thus ruling out various forms of manipulation ('independence of irrelevant alternatives'). Fourth, there should be no dictator ('non-dictatorship'). And, fifth, social preferences should be consistent, in particular there should be no 'cycles' as in Condorcet's paradox above ('collective rationality'). In short, these five conditions require that the democratic procedure should work for all possible inputs that might

¹ Address for correspondence: C. List, Department of Government, London School of Economics, London WC2A 2AE, c.list@lse.ac.uk. This chapter presents earlier arguments and material from List (2002) in a substantially rewritten, redeveloped and extended form. I thank many colleagues for helpful feedback and discussions, including Geoffrey Brennan, Franz Dietrich, John Dryzek, Stephen Elkin, James Fishkin, Robert Goodin, Robert Luskin, David Miller and Philip Pettit.

arise in a pluralistic society, that its outputs should be democratically responsive to its inputs, and that these outputs should themselves be consistent. Can we find such a procedure? Arrow's theorem gives a negative answer to this question. When there are three or more decision options, no procedure will simultaneously satisfy Arrow's five conditions.

Of course, the difficulties posed by democratic decision making depend on how much disagreement there is between different individuals' preferences or judgments. In the rare case of unanimity the difficulties disappear. If everybody had exactly the same preferences or judgments, there would be no conflict to resolve. But although unanimity is sufficient for resolving the problems identified by Condorcet and Arrow, it is not necessary. Since Duncan Black's seminal work (1948), it is known that Condorcet's paradox can be traced back to a 'lack of structure' in the combination of preferences across individuals. Black proved that 'single-peakedness', a structure condition on preferences discussed formally below, is sufficient (but not necessary) for avoiding Condorcet paradoxes.² A well-known corollary of Black's result is that Arrow's impossibility theorem ceases to apply if the domain of admissible input to the decision procedure is restricted to combinations of individual preferences satisfying single-peakedness.

At first sight, Black's result may only seem to confirm what we already know, namely that, if disagreement in a group stays within certain limits – limits that are somehow transcended in the example of Condorcet's paradox – then familiar majoritarian procedures can be used for making consistent democratic decisions. But Black's result teaches us more than that. It highlights an important distinction between two different types of agreement. The two types are 'substantive agreement' on the one hand and 'meta-agreement' on the other. My aim in this chapter is to explore this distinction.

I discuss two different contexts of democratic decision making. The first is the familiar context of preference aggregation, the second the less familiar one of judgment aggregation. In sections 2 and 3, I explore the two types of agreement in the context of preference aggregation, and in section 4, I discuss the significance of meta-agreement for a deliberative democratic response to Condorcet's and Arrow's problems. In section 5, I introduce the context of judgment aggregation, and in sections 6 and 7, I explore the two types of agreement in that context. My main suggestion is that, when agreement is conceptualized in democratic theory and when it is sought in democratic practice, more emphasis should be placed on meta-agreement than is commonly done. In section 8, I take a step back and ask to what extent it is acceptable for the stability of democratic procedures to depend on special empirical contingencies. In section 9, I make some brief concluding remarks.

2. Substantive agreement

How can we define 'substantive agreement'? Two or more individuals are in substantive agreement to the extent that their preferences or judgments are the same. Perfect substantive agreement requires identical preferences or judgments across different individuals. In this section, I focus on preferences; I turn to judgments in section 6 below.

² Jointly with the (harmless) technical condition that the number of individuals is odd.

In response to Condorcet's and Arrow's problems, it is natural to suggest that collective decisions should be preceded by a period of democratic deliberation, so as to reduce the level of conflict between different people's preferences and to bring about consensus on what decision option should be chosen. Jon Elster summarizes this view 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 seems attractive. But there are at least two problems. The first, and practical, problem is that it is often unrealistic to expect democratic deliberation to produce unanimity. 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 given more weight than the expected economic benefits of the new development. Indeed, as discussed in more detail below, experiments on group deliberation have produced little evidence of post-deliberation unanimity (List, Luskin, Fishkin, McLean 2000/2005). Moreover, Cass Sunstein has presented striking empirical evidence that deliberation, especially in homogeneous groups whose members reinforce each other's views, can sometimes lead to polarization of opinion rather than convergence to a centrist position (Sunstein 2002). This is not to deny the benefits of substantive agreement, if it can be reached. But clearly democracy needs to have alternative resources for dealing with conflicts of preferences in those plausible and numerous cases in which deliberation fails to produce unanimity.

The second, and more theoretical, problem with the view outlined by Elster is that it is unclear whether convergence towards substantive agreement – falling short of perfect unanimity – is the most promising strategy for avoiding Condorcet's and Arrow's aggregation problems. William Gehrlein has devised measures of preference homogeneity capturing how closely a given combination of preferences across individuals approximates substantive agreement. Using these measures, he has shown that, although there is a positive correlation between preference homogeneity and the avoidance of Condorcet paradoxes, the correlation is weaker than one might have expected (Gehrlein 2000).

3. Meta-agreement

Black's insight is to ask not whether two or more individuals have the same preferences over a set of decision options, but rather whether their preferences can be rationalized in terms of the same underlying 'left-right' dimension. Suppose the decision options are somehow aligned on a single axis, such as from left-most to right-most. We say that an individual has 'single-peaked' preferences on this axis if he or she has a most preferred position somewhere on the axis with decreasing preference as options get more distant in both directions from the most preferred position. If *all* individuals' preferences are single-peaked on the *same* axis, then we say that the entire combination of preferences across individuals is 'single-peaked'. A

shared axis on which all individuals' preferences are single-peaked, if it exists, is called a 'structuring dimension'. Table 1 shows an example of two preference orderings over five options that are single-peaked on the left-right axis x, y, z, v, w . By contrast, table 2 shows a preference ordering that is not single-peaked on that axis. (In both tables, the options are plotted on the horizontal axis, and ordinal preference intensity is plotted on the vertical axis.)

Table 1. Single-peaked preferences

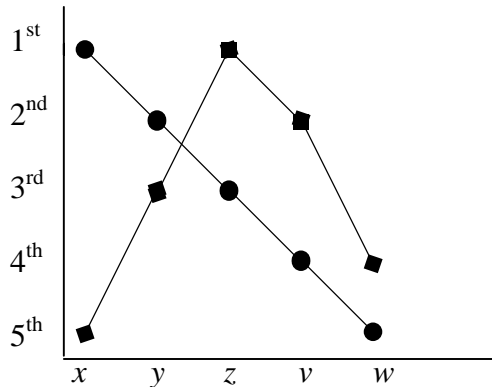
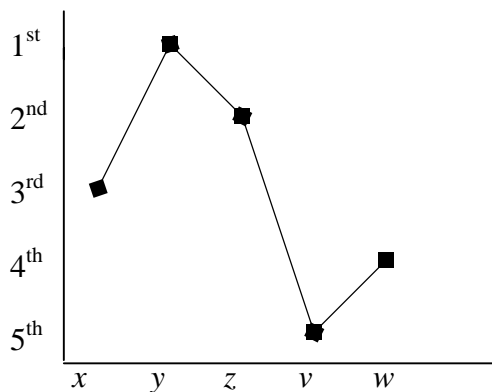


Table 2. Non-single-peaked preferences



The terms 'left' and 'right' are used here in a purely geometrical sense. Any one-dimensional alignment of the options could serve as a structuring dimension, whether it orders the options from most socialist to most capitalist, 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. While the term 'structuring dimension' refers only to a geometrical alignment of the decision options, we may use the term 'issue dimension' to refer to a broader interpretation underlying such an alignment, such as one of the interpretations just given.

Black's concept of single-peakedness inspires the following definition of 'meta-agreement'. Two or more individuals are in meta-agreement to the extent that they agree on a common issue dimension in terms of which a given decision problem is to be conceptualized – and in terms of which preferences are to be rationalized. They may reach perfect meta-agreement while at the same time disagreeing substantively on what the most preferred option is.

Single-peakedness may be an *implication* of meta-agreement. If the individuals agree on a common issue dimension and rationalize their preferences in terms of that dimension, then the resulting combination of individual preferences will satisfy single-peakedness, provided the common (semantic) issue dimension translates into a common (geometrical) structuring dimension. However, as single-peakedness is only a formal structure condition on individual preferences, single-peakedness is logically weaker than meta-agreement. A group of individuals may accidentally have single-peaked preferences on a certain (geometrical) structuring dimension without explicitly rationalizing their preferences in terms of a common (semantic) issue dimension.

Further, single-peakedness is not merely a consistency condition on individual preferences. Take the preferences of a single individual (over a finite set of options). Unless we refer to a specific structuring dimension, it is vacuously true that the individual has single-peaked preferences *on some dimension* (supposing the individual is able to rank the options in a clear order of preference). We can simply define the individual's most preferred option as the left-most option, his or her least preferred option as the right-most option, aligning all other options from left to right in the individual's order of preference. On this artificially constructed structuring dimension, the individual's preference ordering is clearly single-peaked, though uninformatively so, as the constructed structuring dimension has little independent meaning. The condition of single-peakedness becomes non-vacuous only when we *either* refer to a specific structuring dimension *or* apply the condition to several individuals' preferences. In the latter case, we can ask whether there exists at least one *common* structuring dimension on which *all* individuals' preferences are single-peaked. If individual preferences are as in Condorcet's paradox, then there exists no such dimension. Regardless of how we align the options from left to right on some axis in that example, some individuals' preferences will fail to be single-peaked on that axis. On the other hand, if only two of the three different preference orderings in Condorcet's example were held among the individuals (for example, just the first two but not the third), then the resulting combination of individual preferences would be single-peaked on a common structuring dimension. This is the sense in which single-peakedness captures an implication of *agreement*, albeit at a meta-level.

Now Black proved the following simple, yet remarkable result. Given a combination of preferences across individuals satisfying single-peakedness, align the individuals 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 alignment of the individuals, the 'median individual' is the one who has an equal number of individuals to the left and to the right (assuming, for simplicity, that the number of individuals is odd). 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 input to the democratic procedure consists only of individual preference combinations satisfying single-peakedness, then pairwise majority voting is guaranteed to generate collective preferences in accordance with Arrow's minimal conditions (except of course 'universal domain').

Moreover, several studies have shown that consistent majority preferences are still likely to exist even if not all individuals, but only a sufficiently large proportion of them (sometimes as few as 75%) have preferences that are single-peaked on a common structuring dimension (Niemi 1969, Tullock and Campbell 1970, Gehrlein

2004). So perfect meta-agreement, with single-peaked preferences among all individuals, is sufficient but not necessary for avoiding Condorcet's and Arrow's problems; partial meta-agreement, with single-peaked preferences among a sufficiently large proportion of the individuals, will often suffice.

4. Deliberation and meta-agreement

Black's result suggests an alternative response to the challenge posed by Condorcet and Arrow. Rather than seeking convergence towards substantive agreement through deliberation, which may be hard to achieve, we might seek convergence towards meta-agreement. A recent strand of thinking on deliberative democracy advocates precisely this idea (Miller 1992; Knight and Johnson 1994; Dryzek and List 2003, 2004; List, Luskin, Fishkin and Mclean 2000/5). On this view, the key to a deliberative democratic response to Condorcet and Arrow lies not in deliberation-induced substantive agreement, but rather in deliberation-induced meta-agreement. As the proponents of this view emphasize, it is more realistic, though still demanding, to expect deliberation to produce agreement on what the relevant dimension is than to produce agreement on what option should be chosen. Or, more figuratively, it is often easier to reach agreement on what the questions are than on how to answer them. The view can be stated as a three-part hypothesis:

- (1) Group deliberation leads people to identify a common (semantic) issue dimension in terms of which to conceptualize the decision problem at stake.
- (2) For a given such issue dimension, group deliberation leads people to agree on how the decision options are aligned from left to right with respect to that issue dimension; so people determine which (geometric) structuring dimension best represents the given (semantic) issue dimension.
- (3) Once a (semantic) issue dimension and a corresponding (geometric) structuring dimension have been identified as relevant, group deliberation leads each individual to determine a most preferred position (his or her 'peak') on that dimension, with decreasing preference as options get increasingly distant from the most preferred position.

The combination of parts 1, 2 and 3 is, in essence, the hypothesis that group deliberation brings about meta-agreement, which then surfaces in the form of single-peakedness. Neither part of the hypothesis is trivial. And each part raises difficult social-psychological issues, which I am not able to address here. But let me make some suggestions about the status of each part of the hypothesis, albeit in rather simplistic terms.

First, the question of what semantic issue dimension is relevant to a given democratic decision problem is a *normative* question. For example, if society has to choose between different industrial policy options, it is a normative matter whether the choice should be conceptualized in terms of a classical socio-economic left-right dimension, in terms of a trade-off between short-term economic growth and the environment, in terms of a different trade-off between urban and rural development, or in terms of some other issue dimension. This normative question may not have an independently 'true' answer. Rather, different societies might give different answers, depending on their shared values, histories, demographic and ideological circumstances and other factors.

Second, the question of how decision options are aligned from left to right with respect to a given issue dimension is a *factual* question, albeit one on which there may be considerable disagreement (and where the debate may not always be entirely value-free). For example, if society has agreed that its choice between different industrial policy options should be conceptualized in terms of a trade-off between short-term growth and the environment, it is then a complex factual matter how exactly the proposed policy options are positioned in terms of that trade-off, that is, which policy option is most environmentally friendly, which one leads to the greatest short-term growth, and what the relative positions of the intermediate options are in terms of their impact on growth and the environment.

Third, the question of whether an individual's preferences are single-peaked on a given structuring dimension is a question of *rationality*, provided (crucially) that the individual has recognized this structuring dimension (and its underlying semantic issue dimension) as the relevant one. Suppose an individual has come to accept that a given industrial policy decision should be conceptualized in terms of the trade-off between growth and the environment, and suppose further he or she has come to accept that, among different policy options, option x is the most growth-friendly one, option y the most environmentally friendly one, and other options lie in a particular order in between x and y. Would the individual then have single-peaked preferences on this structuring dimension? It would seem that, *if* the individual genuinely believes the given dimension to be the relevant one, *then* he or she should indeed rationalize his or her preferences in terms of that dimension, that is, the individual should determine his or her most preferred position on that dimension and then prefer options less as they get more distant from the most preferred position.

If my remarks are correct, what does the success of the three-part hypothesis depend on? Let me first address parts 2 and 3 of the hypothesis and then turn to part 1.³

The success of part 2 depends on whether group deliberation can bring about agreement on factual matters, for instance by clarifying and supplying information on the properties of the various decision options. The success of part 3 depends on whether group deliberation can induce rationality in individuals. Both of these requirements are not undemanding, but they are clearly not completely implausible.

The success of part 1, by contrast, depends on whether group deliberation can make one particular issue dimension sufficiently salient, so as to produce agreement on the relevance of that issue dimension. This in turn depends on whether deliberation can bring about agreement on what normative considerations are relevant for a given decision problem, a demanding requirement in many cases.

³ The three parts of the hypothesis concern independent mechanisms. It is *logically possible* for different individuals to agree on a semantic issue dimension without agreeing on how exactly options are geometrically aligned with respect to that dimension. For example, a certain industrial policy option (like a nuclear technology option) may be considered extremely environmentally friendly by some, and extremely environmentally unfriendly by others. Likewise, it is *logically possible* for someone to consider a certain geometrical structuring dimension (and its underlying semantic issue dimension) to be relevant without having single-peaked preferences on that dimension. For example, one might strongly disprefer centrist options on that dimension and strongly prefer extremist ones. However, it is not clear that one would hold such a non-single-peaked preference if one genuinely recognized the given dimension as relevant. Rather, holding such a preference might reflect a tacit appeal to a different dimension.

In some cases, an appeal to public reasons and generalizable interests – something that deliberative democrats advocate – may lead to the identification of a single publicly relevant issue dimension and thereby to meta-agreement, as required by part 1 of the hypothesis. But in other cases an appeal to such reasons or interests alone may not suffice, because individuals might still disagree about what is in the public interest or whether economic growth or the environment should be given priority when such interests are in conflict.

Even mainstream rational choice theorists are likely to agree that parts 2 and 3 of the hypothesis are relatively undemanding, and that the demanding part 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 deliberation induces single-peakedness is ultimately an empirical matter that cannot be settled by pure theorizing. In a recent study, List, Luskin, Fishkin and McLean (2000/2005) have used data from James Fishkin and Robert Luskin's deliberative polls to test the hypothesis empirically (on the method of deliberative polling, see Fishkin 1997). They have studied a set of deliberative polls on topics ranging from energy provision in Texas to the future of the monarchy in Australia. In these polls, participants were first interviewed on their preferences and opinions, then invited to participate in a weekend of group deliberation, and finally interviewed again, being asked exactly the same questions as in the pre-deliberation interviews. This research design allows the comparison of pre-deliberation and post-deliberation levels of single-peakedness, measured in terms of the proportion of individuals whose preferences are single-peaked on a common structuring dimension. The study has shown that the post-deliberation levels of single-peakedness were either strictly greater than the corresponding pre-deliberation levels or at least on a par with them. Moreover, the questionnaires also included factual questions revealing how well-informed participants were before and after deliberation about the issues at stake. Among those participants who emerged best informed from the deliberative process, there was a consistent increase in the level of single-peakedness in all deliberative polls under investigation. Finally, the deliberative process did not appear to lead to a consistent increase in the level of substantive agreement, which suggests that deliberation's effect on meta-agreement is more marked than its effect on substantive agreement.

Of course, these findings are not the final word on the introduced hypothesis. Some situations may be favourable to its success, such as when a certain issue dimension – for example, a trade-off between growth and the environment – is easily identified as the salient one. But other situations may be less favourable to its success, such as when a decision problem is perceived to be inherently multidimensional or when different people's values clash. In such cases, people neither agree on what the correct answer is, nor even on how to think about the problem. (On favourable and unfavourable conditions, see also List, Luskin, Fishkin and McLean 2000/2005.)

5. From preferences to judgments

So far my focus has been on preference aggregation. But sometimes democratic decision making bodies are faced with the task of judgment aggregation, that is, they

have to make collective judgments on multiple propositions on the basis of conflicting individual judgments on these propositions. Further, the propositions may be logically interconnected, so the judgments on some propositions may constrain the judgments that can consistently be held on others. Judgment aggregation problems arise, for example, when complex systems of policy or legislation are to be designed, where multiple issues are involved, with interconnections between these issues, and where consistency across different issues matters. (For recent discussions of judgment aggregation, see List and Pettit 2002, 2004; Chapman 2002; List 2004a, b; Pauly and van Hees 2005; Dietrich 2005.)

A simple example illustrates the problem. Suppose a multi-member government has to make judgments on three policy propositions:

- P: A budget deficit is affordable.
 If P then Q: If a budget deficit is affordable, then spending on education should be increased.
 Q: Spending on education should be increased.

For simplicity, let there be three individual government members, with judgments on the propositions as shown in table 3. Each member's judgments are individually consistent.

Table 3. A combination of individual judgments

	Individual 1	Individual 2	Individual 3
P	Yes	Yes	No
If P then Q	Yes	No	Yes
Q	Yes	No	No

If majority votes are taken on each of the three propositions, then a majority accepts P, a majority accepts 'If P then Q', and a yet majority rejects Q, an inconsistency. So proposition-by-proposition majority voting over a set of interconnected propositions can lead to an inconsistent set of collective judgments. Moreover, this can happen even when all individuals hold perfectly consistent individual judgments, as in the present example. This problem is sometimes called the 'discursive dilemma' (Pettit 2001; Brennan 2001; Chapman 1998; the present example is given in Dietrich 2005).

Just as Arrow's impossibility theorem shows that Condorcet's paradox is not just an artefact of majority voting, we may ask whether the present 'paradox' of judgment aggregation hints at a more general problem. A recent theorem by List and Pettit (2002, 2004) addresses this question. Again consider a group of individuals, and suppose we want to find a procedure for aggregating the judgments of these individuals into corresponding judgments for the group as a whole. Also, suppose we want the judgment aggregation procedure to satisfy some minimal conditions, similar in spirit to Arrow's conditions on preference aggregation. First, the procedure should accept as its admissible input all possible combinations of individual judgments, so long as these judgments are individually consistent ('universal domain'). Second, the procedure should give equal weight to all individuals' judgments ('anonymity'). Third, the collective judgments on each proposition should depend only on individual judgments on that proposition and different propositions should be treated equally

(‘systematicity’). And, fourth, collective judgments should be consistent, in particular there should be no collective inconsistencies as in the ‘discursive dilemma’ example (‘collective rationality’). As in the case of Arrow’s conditions, the essence of the conditions is that the judgment aggregation procedure should work for all possible inputs, that its outputs should be responsive to its inputs, and that these outputs should themselves be consistent. Can we find such a procedure? Again, the answer is negative. Except in special cases when the propositions under consideration are largely unconnected, no judgment aggregation procedure will simultaneously satisfy the four conditions just introduced.

Once again, the difficulties posed by this result depend on how much disagreement there is between different individuals’ judgments. I now suggest that the two different types of agreement I have identified in a preference aggregation context can also be identified in a judgment aggregation context, and that they here, too, point towards two different responses to the problems of aggregation. I also argue that Sunstein’s idea of an ‘incompletely theorized agreement’ can be seen as a special case of a substantive agreement (Sunstein 1994).

6. Substantive agreement revisited

In the context of judgment aggregation, two or more individuals are in substantive agreement to the extent that their judgments are the same on the given propositions. As before, one might try to solve judgment aggregation problems by encouraging a period of deliberation prior to forming collective judgments, so as to bring about greater substantive agreement among the individuals involved in the decision. I have already noted in the context of preference aggregation that substantive agreement may be hard to attain in practice. Now, if we are dealing not just with preferences over separate policy options, but with judgments over an entire set of interconnected propositions, then substantive agreement may become even harder to attain.

There is, however, a special case of substantive agreement that may seem more practically attainable, although it is still demanding (for a more detailed discussion, see List 2004b). This special case is a version of an ‘incompletely theorized agreement’, which Sunstein originally proposed for the legal realm. Often different individuals’ judgments on fundamental moral and political issues are mutually incompatible, as they reflect genuinely different views of politics and morality, including different supporting reasons even for those judgments on which there is agreement (such as ‘killing is wrong’). But, on less fundamental and more pragmatic issues, there may be more agreement. Given someone’s overall set of moral and political judgments, this set will contain some judgments that the individual considers fundamental, perhaps even ‘axiomatic’, and others that he or she considers less fundamental, or more pragmatic. Now it is possible that different mutually incompatible fundamental judgments may nonetheless have certain implications in common at a less fundamental level. This is possible because, if there is a relation of logical entailment between someone’s fundamental judgments and his or her less fundamental ones, this relation is usually a one-way entailment. The more fundamental judgments entail the less fundamental ones, but not vice-versa. The less fundamental and more pragmatic judgments are usually logically insufficient to tell us what the underlying fundamental judgments are. The same pragmatic judgments may be consistent with more than one fundamental system of supporting reasons.

For example, consider how many different fundamental reasons one might give to support a particular social welfare policy. One might believe in social justice and equality for liberal reasons or for religious reasons. Or one might believe that the main objective of the state is to keep the streets safe and to prevent crime and social disorder, and that welfare policies are the best way to achieve this. Or consider how many different fundamental reasons one might give for an environmental protection policy. One might believe in the rights of future generations; or one might believe that the rights of non-human animals ought to be respected, or that ecosystems should be treated as ‘ends in themselves’. Or one might believe that environmental disasters would ruin the economy and that the best way to secure long-term economic stability would be to protect the environment.

An incompletely theorized agreement requires the identification of a certain set of non-fundamental or pragmatic propositions such that substantive agreement *on these propositions* is feasible, even when there is no substantive agreement on any underlying reasons. In the example of table 3, individuals 2 and 3 both hold that spending on education should not be increased, that is, they both reject proposition Q, even though they fundamentally disagree on *why* it is that spending on education should not be increased. Individual 2 holds that, although a budget deficit is affordable, education would not be a good way to spend additional funds. By contrast, individual 3 holds that education would be a good way to spend additional funds if a budget deficit were affordable, but a deficit simply cannot be afforded. So individuals 2 and 3 are in substantive agreement on proposition Q (to be precise, on its rejection), but not on any of the other propositions; so their substantive agreement on proposition Q is an incompletely theorized one.

It is an open question whether, and under what conditions, political deliberation can bring about substantive agreement – albeit perhaps an incompletely theorized one – on a sufficiently broad range of issues. In this brief discussion, however, two points should have become clear. First, a substantive agreement on a restricted range of issues is easier to achieve than a substantive agreement on all issues. Second, the idea of an incompletely theorized agreement is clearly different from that of a meta-agreement. In an incompletely theorized agreement, 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 questions are. Meta-agreement, by contrast, requires agreement on questions, but not necessarily on answers.

7. Meta-agreement revisited

Meta-agreement is defined as agreement on a common issue dimension in terms of which a given decision problem is to be conceptualized – and in terms of which preferences or judgments are to be rationalized. In a preference aggregation context, I have argued that single-peakedness may be an implication of meta-agreement. Specifically, if the individuals rationalize their preferences in terms of a common issue dimension, then the resulting combination of individual preferences will satisfy single-peakedness, provided the common issue dimension also corresponds to a common structuring dimension.

While Black's structure condition of single-peakedness itself is not applicable to a judgment aggregation context, I will now show that an alternative structure condition can be devised for the latter context too (List 2003). Moreover, the new structure condition, like single-peakedness, can be interpreted as an implication of meta-agreement.

Again, the question is not whether two or more individuals hold the same judgments, but now it is whether there exists a single alignment of the individuals (as opposed to options in the preference aggregation context) from left-most to right-most such that, for every proposition under consideration, the individuals accepting the proposition are either all to the left, or all the right, of those rejecting it. If there exists a left-right alignment of the individuals with this property, then the given combination of judgments across individuals satisfies 'unidimensional alignment'. Once again, a left-right alignment of the individuals with this property is called a 'structuring dimension'.

It is easy to see that the judgments in table 3 above violate unidimensional alignment. No matter how the individuals are aligned from left to right, it is impossible to get the pattern of acceptance and rejection required for unidimensional alignment. By contrast, the judgments of the five individuals in table 4 below satisfy unidimensional alignment: there exists a single left-right alignment of the five individuals – namely, 1, 2, 3, 4, 5 – with respect to which, for every proposition, the individuals accepting the proposition are either all to the left, or all the right, of those rejecting it.

Table 4. Unidimensionally aligned judgments

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5
P	No	No	No	No	Yes
If P then Q	Yes	Yes	Yes	Yes	No
Q	Yes	Yes	No	No	No

Now a result similar to Black's result on single-peakedness can be proved (List 2003). Given a combination of judgments across individuals satisfying unidimensional alignment, align the individuals from left-most to right-most on the corresponding structuring dimension. As before, the 'median individual' with respect to this alignment is the one who has an equal number of individuals to the left and to the right (again assuming an odd number of individuals, for simplicity). Then the judgments held by that median individual – the judgments of individual 3 in table 4 – will be accepted in proposition-by-proposition majority voting. And provided the median individual's judgments are internally consistent, so are the resulting collective judgments. Again, a simple corollary is that, if the domain of individual input to a judgment aggregation procedure consists only of judgment combinations satisfying unidimensional alignment, then proposition-by-proposition majority voting is guaranteed to generate collective judgments in accordance with the minimal conditions on judgment aggregation introduced above (except of course 'universal domain').

The claim that unidimensional alignment is an implication of meta-agreement may seem less straightforward than the analogous claim for single-peakedness, but here is a way of making it plausible. Suppose, first, that there is a 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, let me call it a left-right dimension, but again a range of interpretations is possible. And suppose, second, that, for each proposition, the extreme positions on that left-right dimension correspond to either clear acceptance or clear rejection of the proposition; and, further, 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 substantive agreement at all. In the case of table 4, the left-most and right-most individuals disagree about *every* proposition; yet their judgments are unidimensionally aligned. Unidimensional alignment requires only a common left-right alignment of the individuals that systematically structures their pattern of acceptance and rejection over the various propositions.

Once again, my claim is only that unidimensional alignment may be an *implication* of meta-agreement. Like single-peakedness, unidimensional alignment is logically weaker than meta-agreement. A combination of judgments across individuals may accidentally have the right formal structure for unidimensional alignment without the individuals conceptualizing all propositions in terms of a common issue dimension.

Moreover, like single-peakedness, unidimensional alignment is not merely a consistency condition on individual judgments. The judgments of a single individual always vacuously satisfy unidimensional alignment, though uninformatively so. Like single-peakedness, unidimensional alignment becomes non-vacuous only when we apply the condition to several individuals’ judgments. In this sense unidimensional alignment also captures an implication of *agreement*, albeit again at a meta-level.

Unlike in the case of single-peakedness, however, no empirical research has been done on whether group deliberation can induce unidimensional alignment or on whether there are any non-trivial real-world situations in which the judgments across different individuals satisfy unidimensional alignment. But the mere observation that substantive agreement is often hard to attain, while unidimensional alignment is a less demanding condition, should lead us to give more attention to that condition.

8. Empirical contingencies and the design of democratic procedures

A critic might be unconvinced that deliberation-induced meta-agreement and corresponding structure conditions such as single-peakedness or unidimensional alignment open up attractive escape-routes from the paradoxes and impossibility results of social choice theory. The critic might argue as follows. Let us grant that, *if* empirical circumstances are such that individual preferences or judgments satisfy (or approximate) the relevant conditions, *then* familiar democratic procedures will indeed generate consistent collective outcomes. But, as soon as empirical circumstances are different, the very same procedures will fail to do so. Social choice theorists can even predict when such collective inconsistencies will occur. The procedures work well in some empirical circumstances (such as for certain combinations of preferences or

judgments) but not in others. Further, the impossibility theorems tell us that this problem is not just an artefact of specific majoritarian procedures, but that it is a much more general problem. No democratic procedure will avoid that problem, unless we are willing to sacrifice some seemingly attractive minimal conditions.

Consider an analogy from engineering. One would not like to design a house merely on the basis that there are some empirical circumstances in which the house would be stable, while there are others in which it would collapse. Rather, one would seek to design a house on the basis of careful physical calculations confirming its stability. Analogously, the critic might argue, it is a risky strategy to use familiar majoritarian decision procedures and to rely on the observation that for some empirical circumstances (such as situations of meta-agreement) these procedures will work well, while ignoring the fact that for others they will not. Like a house, a democratic procedure should be designed so as to work well in *all* relevant circumstances (that is, for all possible combinations of preferences or judgments). A procedural designer should not rely on the hope that problematic empirical circumstances will not arise. Rather, the designer should make sure that the procedure guarantees consistent outcomes whatever the circumstances are. So far the critic's objection.

The objection raises an important question. Should democratic procedures be designed in such a way as to work robustly under *all* possible empirical circumstances or is it acceptable for such procedures to rely on specific empirical contingencies that are exogenous to, and not guaranteed by, them? At first sight, the critic's objection has some force. After all, democratic procedures that rely on specific contingencies may seem prone to erratic behaviour, just as a house that is stable only under specific conditions may seem unsafe. But, on closer inspection, the objection loses some of its force. Even the most well-designed house will collapse under *some* circumstances, for example if there is a sufficiently strong earthquake. It is simply not true that a well-designed house will be stable in *all* circumstances. The critic might respond that official building standards take that problem into account. In earthquake zones like California houses are required to meet more demanding building standards than in earthquake-free zones like Britain. But, if an exceptionally strong earthquake were to occur, even a house built according to the most demanding standards might collapse.

We can make the following observation from these points. Whenever something is to be designed, be it a house or a democratic procedure, there is a domain of possible empirical circumstances that might arise. In some of these circumstances (call them type 1) the house will be stable and the procedure will work well, while in others (call them type 2) the house will collapse and the procedure will generate inconsistent outcomes. Rather than trying to design a house or procedure for which there are no type 2 circumstances, which may be difficult if not impossible, a more reasonable strategy would be the following. First, consider the probability distribution over the relevant domain of circumstances, and then ask whether the probability of type 2 circumstances is sufficiently low to make the risk of their occurrence bearable. For example, even in Britain the probability of a massive earthquake is presumably non-zero. 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 the stability of a house or procedure across *all* possible circumstances cannot be achieved, the building standards for houses or procedures would have to be adjusted

to how probable the various possible circumstances are. The idea would be to design a house or procedure so as to ensure a high probability of type 1 circumstances and a low probability of type 2 circumstances.

Suppose the democratic culture in a group or society is such that the achievement of (sufficient) meta-agreement is highly probable (with single-peakedness or unidimensional alignment as an implication), while the occurrence of preference or judgment combinations leading to majority inconsistencies is highly improbable. Then the use of familiar majoritarian procedures for democratic decision making is as defensible as building houses in Britain that do not meet Californian building standards.⁴ But suppose the democratic culture is different and combinations of preferences or judgments leading to majority inconsistencies are quite probable. Then the use of such procedures is much less defensible. The defensibility of a democratic procedure is therefore sensitive to the relevant probability distribution over the set of all possible inputs that might come up. This probability distribution, in turn, depends on a range of empirical features of the relevant group or society, their values, beliefs, ideological attitudes, psychological dispositions and so on. Of course, deliberative democrats would not simply take this probability distribution as given. Rather, they would encourage deliberative arrangements specifically with the aim of transforming it in such a way as to increase the probability of type 1 circumstances and decrease that of type 2 circumstances. Whether, and how, this can be achieved is of course a difficult empirical question, on which much further research is required.

If we take these considerations seriously, then the defensibility of a particular democratic procedure is no longer an *a priori* matter, as sometimes thought in social choice theory, but dependent on empirical contingencies, which may differ from context to context.

9. Concluding remarks

The distinction between substantive agreement and meta-agreement was motivated by Duncan Black's insight into how Condorcet's paradox can be avoided. And, indeed, in the context of preference aggregation – the focus of Black's own work – there are both theoretical and empirical results in support of the view that deliberation-induced meta-agreement can facilitate democratic decision making in the face of conflicting individual preferences.

In the context of judgment aggregation, on the other hand, my conclusions must remain much more tentative. Again, the two types of agreement can be identified, and their logical status, even in terms of avoiding majority inconsistencies, is very similar to that of their more well-known counterparts in the context of preference aggregation. But whether deliberation-induced meta-agreement is practically attainable in this new context remains an open question for democratic theory and a challenge for democratic practice.

⁴ Under arguably undemanding conditions on the probability distribution over combinations of preferences, it can be shown that the probability of cyclical majority preferences as in Condorcet's paradox converges to 0 with an increasing number of individuals (see the appendix on the probability of cycles in List and Goodin 2001). Unfortunately, the conditions required for avoiding majority inconsistencies in judgment aggregation are more demanding (List 2005).

In conclusion, many important questions remain open. Most importantly, we will need to tackle the question of whether a certain minimal level of cohesion among individual preferences or judgments is necessary for collective decisions to be both sufficiently democratic and sufficiently consistent, or whether democratic decision procedures can be made so robust as to cope with even the most extreme cases of disagreement. The impossibility results of social choice theory would certainly seem to apply in the latter cases.

References

- 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. (2001) 'Collective Coherence?' *International Review of Law and Economics* 21: 197-211
- Chapman, B. (1998) 'More Easily Done than Said: Rules, Reason and Rational Social Choice,' *Oxford Journal of Legal Studies* 18: 293-3
- Chapman, B. (2002) 'Rational Aggregation,' *Politics, Philosophy and Economics* 1: 337-354
- Dietrich, F. (2005) 'Judgment aggregation: (im)possibility theorems,' *Journal of Economic Theory*, forthcoming
- Dryzek, J., and C. List (2003) 'Social Choice Theory and Deliberative Democracy: A Reconciliation,' *British Journal of Political Science* 33: 1-28
- Dryzek, J., and C. List (2004) 'Social Choice Theory and Deliberative Democracy: A Response to Aldred,' *British Journal of Political Science* 34: 752-758
- Elster, J. (1986) 'The Market and the Forum,' in J. Elster and A. Hylland (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
- Gehrlein, W. (2004) 'Probabilities of Election Outcomes with Two Parameters: The Relative Impact of Unifying and Polarizing Candidates,' working paper, University of Delaware
- Knight, J., and Johnson J. (1994) 'Aggregation and Deliberation: On the Possibility of Democratic Legitimacy,' *Political Theory* 22: 277-296
- List, C. (2002) 'Two Concepts of Agreement,' *The Good Society* 11: 72-79, reprinted in an extended form as 'Substantive and Meta-Agreement,' in A. van Aaken, C. List and C. Luetge (eds.) (2003) *Deliberation and Decision*, Aldershot (Ashgate)
- List, C. (2003) 'A Possibility Theorem on Aggregation over Multiple Interconnected Propositions,' *Mathematical Social Sciences* 45: 1-13
- List, C. (2004a) 'A Model of Path-Dependence in Decisions over Multiple Propositions,' *American Political Science Review* 98: 495-513
- List, C. (2004b) 'The Discursive Dilemma and Public Reason,' working paper, London School of Economics
- List, C. (2005) 'The probability of inconsistencies in complex collective decisions,' *Social Choice and Welfare*, forthcoming

- List, C., and R. E. Goodin (2001) 'Epistemic Democracy: Generalizing the Condorcet Jury Theorem,' *Journal of Political Philosophy* 9: 277-306
- List, C., R. C. Luskin, J. S. Fishkin, and I. McLean (2000/5) 'Can Deliberation Induce Single-Peakedness: Evidence from Deliberative Polls,' paper presented at the 2000 conference of the American Political Science Association
- List, C., and P. Pettit (2002) 'Aggregating Sets of Judgments: An Impossibility Result,' *Economics and Philosophy* 18: 89-110
- List, C., and P. Pettit (2004) 'Aggregating Sets of Judgments: Two Impossibility Results Compared,' *Synthese* 140: 207-235
- 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
- Pauly, M., and M. van Hees (2005) 'Logical constraints on judgment aggregation,' *Journal of Philosophical Logic*, forthcoming
- Pettit, P. (2001) 'Deliberative Democracy and the Discursive Dilemma,' *Philosophical Issues* 11: 268-299
- Sunstein, C. (1994) *Political Conflict and Legal Agreement*, Tanner Lectures on Human Values, Harvard University
- Sunstein, C. R. (2002) 'The Law of Group Polarization,' *Journal of Political Philosophy* 10: 175-195
- Tullock, G., and C. D. Campbell (1970) 'Computer Simulation of a Small Voting System,' *Economics Journal* 80: 97-104