Social Choice Theory and Deliberative Democracy: A Reconciliation*

version 27 February 2002
forthcoming in *British Journal of Political Science*

John S. Dryzek
Social and Political Theory Program
Research School of Social Sciences
Australian National University
Canberra, ACT 0200
Australia
jdryzek@coombs.anu.edu.au

Christian List
Nuffield College
Oxford
OX1 1NF
U.K.
christian.list@nuffield.oxford.ac.uk

---

* Previous versions of this paper were presented at the Joint Sessions of the European Consortium for Political Research in Mannheim, March 1999, and at the 2000 Annual Meeting of the American Political Science Association. The authors did much of the initial work while Dryzek was a visitor at Nuffield College in 1998, and Dryzek thanks the College for its hospitality. For helpful comments, we thank Keith Dowding, James Fishkin, Natalie Gold, Robert Goodin, Iain McLean, Gerry Mackie, David Miller, Claus Offe, Anne Sliwka, the editor, and the anonymous reviewers of this paper.

Social Choice Theory and Deliberative Democracy:
A Reconciliation

Abstract

The two most influential traditions of contemporary theorizing about democracy, social choice theory and deliberative democracy, are generally thought to be at loggerheads, in that the former demonstrates the impossibility, instability or meaninglessness of the rational collective outcomes sought by the latter. We argue that the two traditions can be reconciled. After expounding the central Arrow and Gibbard-Satterthwaite impossibility results, we reassess their implications, identifying the conditions under which meaningful democratic decision making is possible. We argue that deliberation can promote these conditions, and hence that social choice theory suggests not that democratic decision making is impossible, but rather that democracy must have a deliberative aspect.

1. Two Traditions of Democratic Theory

In the past decade the theory of democracy has been dominated by two very different approaches. Within democratic theory as conventionally defined the strongest current is now deliberative.[1] For deliberative democrats, the essence of democratic legitimacy is the capacity of those affected by a collective decision to deliberate in the production of that decision. Deliberation involves discussion in which individuals are amenable to scrutinizing and changing their preferences in light of persuasion (but not manipulation, deception, or coercion) from other participants. Claims for and against courses of action must be justified to others in terms they can accept. Jürgen Habermas and John Rawls, respectively the most influential continental and Anglo-American political philosophers of the late 20th century, have both identified themselves as deliberative democrats.[2] Deliberative democrats are uniformly optimistic that deliberation yields rational collective outcomes.

The main competing tradition is social choice theory, whose proponents generally deduce far less optimistic results. To social choice theorists, the democratic problem involves aggregation of views, interests, or preferences across individuals, not deliberation over their content. From the seminal work of Kenneth Arrow on, it has been argued that such aggregation is bedeviled by impossibility, instability and arbitrariness.[3] Arrow proved the non-existence of any aggregation mechanism satisfying a set of seemingly innocuous conditions. This critique of democracy was radicalized by William Riker, who argued that any notion of a popular will independent of the mechanism used to aggregate preferences was untenable.[4] Given that there is no good reason to choose any particular mechanism over any other, supposedly democratic collective choices are arbitrary, and democracy is emptied of meaning. As Hardin puts it, social choice theory has exposed "flaws – grievous foundational flaws – in democratic thought and practice."[5]

Riker's radicalization of the social-choice-theoretic critique created a chasm between the two traditions that might seem impossible to bridge.[6] We argue that the two traditions can in fact be reconciled. Though social choice practitioners may be unaware of it, some even arguing the opposite, we argue that their theory points to the functions deliberation can perform in making collective decisions both tractable and meaningful, thus providing a crucial service to deliberative democracy. The structure of this paper follows the results of social choice theory, for it is these that both pose the challenges to democracy and pinpoint the locations at which deliberative responses must be sought.

Methodologically, our arguments consist of a logical component, a normative component and an empirical-hypothetical component. The logical component takes an "if-then"-form: *If* condition X obtains, *then*, by the logic of social choice theory, meaningful collective decisions are possible. The normative component defends the claim that the constraints required for bringing about condition X are inherent in or consistent with core elements of deliberative democracy. The empirical-hypothetical component, finally, seeks to render plausible the empirical hypothesis that deliberation facilitates the

emergence of condition X. While we provide empirical illustrations, more systematic testing is beyond our scope here.

2. The Social-Choice-Theoretic Challenge

Before we explicate some of the impossibility results at the centre of the social-choice-theoretic critique of democracy (2.2) and describe how they are usually invoked (2.3), we briefly sketch what we take to be the essence of social choice theory (2.1).

2.1 Disentangling Social Choice Theory from Rational Choice Theory

Social choice theory is a mathematical theory of group decision making. Its concern is not so much the empirical question of how groups actually do make decisions, rather the normative and logical questions of how they should, and could, aggregate information about the views, interests, or preferences of individuals into group decisions. The normative aspect is the specification of minimal conditions an acceptable aggregation mechanism must satisfy. The logical aspect is the identification of the class of logically possible aggregation mechanisms satisfying a given set of conditions. But the purely logical side of social choice theory will not favour one such set of conditions over another. Hence Riker's argument that there is no popular will independent of a particular aggregation mechanism is not in itself a decisive challenge to democracy: without normative input in the form of a favoured set of conditions, it is impossible to design a social choice mechanism. One of the roles of deliberation is to seek agreement on such a set of conditions.

While social choice theorists often model individuals as self-interested utility maximizers,[7] there is no reason why the purely logical and normative analysis of aggregation mechanisms should presuppose any specific behavioural assumption. Social choice theory is, then, distinct from, and not committed to the premises of, rational choice theory.

Nonetheless, an empirical account of individual behaviour may (implicitly) affect the normative choice of minimal conditions on aggregation. For instance, if, empirically, human beings tend to distort information when expedient, we may require that aggregation mechanisms minimize people's incentives to manipulate information (an example of such a condition is strategy-proofness as discussed below). Such rational-choice-theoretic premises may or may not turn out to be empirically adequate. Our point is simply that social choice theory and rational choice theory should not be conflated.

2.2 Arrow's Theorem and the Gibbard-Satterthwaite Theorem

To introduce Arrow's theorem, we recall Condorcet's paradox of cyclical majority preferences. Suppose there are three individuals, labelled 1, 2 and 3, and three alternatives, labelled *x*, *y* and *z*, with the following preferences:

individual 1: $x > y > z$
individual 2: $y > z > x$
individual 3: $z > x >$ y

Then there are majorities of 2 out of 3 individuals for $x > y$, for $y > z$ and for $z > x$. The resulting majority preference ordering is cyclical: $x > y > z > x$.

Arrow's theorem generalizes Condorcet's insight.[8] We consider a set of individuals (e.g. voters, decision-makers, committee members), labelled $N = \{1, 2, ..., n\}$, and a set of alternatives (e.g. policy options, election candidates), labelled $X = \{x, y, z, ...\}$.[9] To each individual, *i*, there corresponds a *personal preference ordering*, R_i, over the alternatives in X.[10] A *profile of personal preference orderings* is an assignment of one such ordering to each individual, i.e. an *n*-tuple $\{R_i\}_{i \in N}$. This formalism allows different interpretations: R_i could represent individual *i*'s explicitly expressed views or judgements or, alternatively, *i*'s interests as assessed by some external evaluation standard. Abbreviated, xR_iy will be interpreted as "from individual *i*'s perspective, *x is at least as good as y*". As a further abbreviation, we write xP_iy if xR_iy and not yR_ix.

A *social welfare function* (*SWF*) is an aggregation function F whose input is a profile of personal preference orderings and whose output is a *social ordering* R over the alternatives in X.[11] xRy will be interpreted as "from the perspective of the group *N*, *x is at least as good as y*". We write xPy if xRy and not yRx.

Pairwise majority voting is an *example* of a SWF. For each profile of personal preference orderings, the *Condorcet ordering* is the social ordering defined as follows: xRy if and only if the number of individuals with the preference xR_iy is at least as great as the number of individuals with the preference yR_ix. Now pairwise majority voting is the function F mapping each $\{R_i\}_{i \in N}$ to the corresponding Condorcet ordering R. What Condorcet's paradox shows is that pairwise majority voting may fail to generate transitive social orderings. Arrow's theorem abstracts from pairwise majority voting, showing that there exists no other SWF, F, satisfying certain minimal conditions that generates transitive social preference orderings. Arrow's conditions are as follows. Let $R = F(\{R_i\}_{i \in N})$.

Universal Domain (U). The domain of F is the set of all logically possible profiles of personal preference orderings.

Weak Pareto Principle (P). If, for all *i* in *N*, xP_iy, then xPy.

Independence of Irrelevant Alternatives (I). The position of x relative to y in the social ordering R depends exclusively on the position of x relative to y in each of the personal preference orderings in $\{R_i\}_{i\in N}$.

Non-Dictatorship (D). F is not dictatorial: there does not exist an i in N such that, for all $\{R_i\}_{i\in N}$ in the domain of F and all x and y in X, xP_iy implies xPy.

Theorem 1. There exists no SWF F (generating transitive social orderings) which satisfies (U), (P), (I) and (D).[12]

By Arrow's theorem, any SWF F will *of logical necessity* violate at least one of (U), (P), (I) and (D). Any democratic decision mechanism thus exhibits at least one of the following flaws: a failure to generate a determinate social ordering for certain profiles of personal preference orderings (if (U) is violated), inefficiency by sometimes ranking Pareto-suboptimal alternatives above Pareto-optimal ones (if (P) is violated), manipulability by changes of the set of options X (the 'agenda') (if (I) is violated), or dictatorship (if (D) is violated). Moreover, relaxation of the requirement that social orderings be transitive does not solve the problem. The weaker requirement of *quasi-transitivity* may still give rise to a so-called oligarchy.[13]

To state the Gibbard-Satterthwaite theorem, define a *social choice function* (SCF) to be an aggregation function F whose input is a profile of personal preference orderings and whose output is a single winning alternative in X.[14] Pairwise majority voting also provides an example of a SCF. A *Condorcet winner* is defined to be a top-ranked alternative in a Condorcet ordering. Now pairwise majority voting is the function F mapping each $\{R_i\}_{i\in N}$ to a corresponding Condorcet winner. In analogy to Arrow's theorem, the Gibbard-Satterthwaite theorem is concerned with SCFs in general, abstracting from pairwise majority voting. In addition to (U) and (D)[15], two minimal conditions are imposed on a SCF F:

The Range-Constraint (R). The range of F contains at least three distinct alternatives in X.

F is *manipulable* by individual i at the profile $\{R_i\}_{i\in N}$ if the following condition holds:

Strategic Incentives (SI). If i submits a *false* preference ordering R'_i (where $R'_i \neq R_i$), then F selects an alternative y' that is strictly better from the perspective of i's *true* preference ordering than the alternative y that would be selected by F if i submitted the *true* preference ordering R_i – formally $y'P_iy$, where $y' = F(\{R_1, ..., R'_i, ..., R_n\})$ and $y = F(\{R_1, ..., R_i, ..., R_n\})$.

Strategy-Proofness (S). There does not exist a profile $\{R_i\}_{i\in N}$ at which F is manipulable by some i in N.

As indicated in section 2.1, (S) is a (normative) condition whose appeal is closely linked with an empirical assumption. If, empirically, individuals *act* on the incentives for manipulation provided by situations of type (SI), then a condition that rules out the occurrence of such situations, namely condition (S), is normatively attractive. In section 3.2, we challenge the plausibility of this empirical assumption in a deliberative setting.

Theorem 2. There exists no SCF F which satisfies (U), (R), (D) and (S).[16]

By the Gibbard-Satterthwaite theorem, any SCF will violate at least one of (U), (R), (D) and (S). Given any SCF F, at least one of the following flaws thus seems inevitable: a failure to generate a determinate winning alternative for certain profiles of personal preference orderings (if (U) is violated), insensitivity to individuals' views, interests, or preferences (if (R) is violated), dictatorship (if (D) is violated), strategic manipulability (if (S) is violated).

The problem of strategic manipulation becomes even more intractable if we allow the possibility of *counterthreats* of the form "if you submit false preferences, then so will I/we".[17] A SCF violating (S) can then create perverse incentives in the form of a prisoners' dilemma. Suppose F is manipulable by individual i at $\{R_i\}_{i\in N}$, and suppose a coalition M of other individuals states a counterthreat against i's manipulation to submit false preferences themselves. If the incentives for individual i and coalition M are as shown in table 1, then the dominant strategy for everyone is to submit false preferences, and the (unique) Nash equilibrium is the bottom right-hand box of the payoff matrix. However, everyone would be better off submitting true preferences. If such a situation can occur, the Gibbard-Satterthwaite theorem clearly poses a significant challenge to democracy.[18]

TABLE 1 ABOUT HERE

2.3 Two Ways to Deploy Impossibility Arguments

In the sporadic encounters between deliberative democracy and social choice theory, the Arrow and Gibbard-Satterthwaite theorems have been deployed to support diametrically opposed positions.

From the direction of deliberative democracy, Sunstein argues, citing Arrow, that "it is doubtful that private desires or even aspirations can be well-aggregated through the process of majority rule", thus proving the necessity for deliberation across those holding initially different

preferences.[19] But this invocation merely asserts deliberation's superiority through reference to the problems of aggregation; it does not show why deliberation avoids the same problems. Elster holds that under deliberation "there would not be any need for an aggregating mechanism, since a rational discussion would tend to produce unanimous preferences."[20] However, it is surely overly optimistic to expect that rational discussion will always produce unanimity, so even after discussion-induced preference changes aggregation of conflicting preferences may be necessary. Miller argues that deliberation produces preference profiles that satisfy single-peakedness,[21] a structure condition to be discussed below. He further suggests that if the basic cause of cycling is the combination of several normative dimensions into a single vote or choice, then deliberation can unpack the various dimensions. Deliberators can search for acceptable choices on each dimension, then aggregate these choices into a coherent collective choice. Miller's argument points in the right direction, but requires further elaboration. Below we will address his hypothesis that deliberation induces single-peakedness in greater detail, and we will argue that disaggregation helps only under specific conditions.

From the direction of social choice theory, followers of Riker have argued that Arrow problems devastate deliberation, because deliberative democrats prescribe a procedure as open as possible in allowing different viewpoints and as free as possible from constraints imposed by rules and inequalities. But these are exactly the conditions of structurelessness conducive to impossibility, instability and strategic manipulability in collective choice.[22] Regarding the Gibbard-Satterthwaite theorem, deliberation might even exacerbate strategic behaviour. For effective manipulation requires information about others' preferences, and deliberation can supply such information. In short, the prescriptions of deliberative democrats are likely only to make collective choice more intractable.

Knight and Johnson recognize that "the standard claim that deliberation aims at preference transformation appears … either too strong or beside the point. It is too strong if it requires convergent, homogeneous preferences." And they point out – along the lines of Miller's hypothesis – that "if voters can agree about the dimension over which they disagree … majority rule need not generate cyclical social orderings", where agreement about the dimension of disagreement is interpreted as single-peakedness.[23] But Knight and Johnson are less optimistic than Miller on whether such agreement can be achieved in practice, arguing that openness and freedom of communication within the forum will "unsettle, if not altogether subvert, any extant shared understanding of political conflict".[24] Against this interpretation of the conditions of deliberation as lack of structure, we argue that there are "structuration" processes endogenous to deliberation that can speak to the problems highlighted by social choice theory.

3. A Deliberative Response

Each condition of Arrow's theorem and of the Gibbard-Satterthwaite theorem points towards a potential escape-route from the impossibility problems. If *any one* of these conditions is relaxed, there exist social choice procedures satisfying all the others, and such procedures can, in principle, be employed in democratic decision making. The problem, according to social-choice-theoretic critics, is that all these conditions are so basic that weakening even one would yield undesirable consequences. If we relax (I) or (S), for example, we would solve the impossibility problems seemingly at the expense of, respectively, possible agenda manipulation or possible submission of false preferences. However, we argue in sections 3.2 to 3.4 that a relaxation of each of (S), (U) and (I) is an option in a deliberative setting. In section 3.5 we discuss an escape-route from Arrow's theorem that is – at first sight surprisingly – consistent with *all* of Arrow's conditions.

We explained above that our arguments consist of a logical "if-then" component, a normative component and an empirical-hypothetical component. The "if-then" component is given by social-choice-theoretic *possibility* results. The aim of deliberation must then be to impose certain constraints on democratic decision processes required for satisfying the antecedents of these "if-then" results. The normative component of our arguments defends the use of these constraints as consistent with, or even inherent in, deliberation. In some of our arguments, particularly those in sections 3.4 and 3.5, the recognition that the requisite constraints could be consistently implemented in the institutions of a deliberative democracy does most of the work – whether they will be implemented is of course a different matter. Other arguments, particularly those in sections 3.2 and 3.3, depend more crucially on empirical hypotheses, which require further empirical corroboration.

But first we introduce some conceptual tools to classify aggregation problems and those aspects of deliberation that can be brought to bear.

3.1 A Simple Taxonomy of Decision Processes in a Deliberative Democracy

We use two (purely formal) criteria for the classification of aggregation problems that fit our description of social choice theory:[25]

(i) What is the input of the aggregation?

- J *Views* or *judgements* of individuals, as explicitly expressed by them (e.g. by voting), or
- I *Interests* or the *welfare* of individuals, as assessed by some external standard (e.g. by an index of a person's socio-economic welfare).

(ii) What is the output of the aggregation?

- D A *decision*, or
- W A *welfare judgement*.

Of the four possible combinations of these criteria, JD and ID are most relevant to the theory of democracy. It is helpful to subdivide the category JD further by the following criteria:

(iii) Are the individuals' views and judgements subjected to

n *no deliberation*, or

d *deliberation*

prior to the actual decision?

(iv) Are the individuals' views and judgements expressed

v by (anonymous) *voting*, or

d in a group *discussion*

when the decision is finally taken?

There are now four sub-categories of JD: JD-nv, JD-nd, JD-dv, and JD-dd. In reality, criteria (iii) and (iv) admit a continuum of possibilities between n and d and between v and d, respectively. Category JD-nv represents the (rare) limiting case of pure aggregative decision making, in which no discussion is involved in the formation of people's views and judgements. Category JD-dd represents the pure deliberative case, in which people frame, revise and express their views and judgements in a reasoned discussion leading to an appropriate (consensus) decision. Deliberative democrats consider processes of type JD-dd more desirable than processes of type JD-nv, but they need not banish voting-based processes. Unlike purely aggregative democrats, however, they do not treat views and preferences as exogenously given. Category JD-dv may seem desirable if the number of decision-makers is large or if individual decision-makers are to be protected against social coercion. Fishkin's deliberative opinion polls can be interpreted as a version of JD-dv.[26] Participants first express their views by filling out confidential questionnaires, then they deliberate, and finally they express their views again in the same questionnaires. Category JD-nd, finally, seems unrealistic: it represents the limiting case in which a decision is made on the basis of views and judgements expressed in group communication, with no discussion. Perhaps group discussion in what Gambetta calls a "Claro!" culture could take this form, where highly opinionated agents are unwilling as a matter of "discursive machismo" to subject their ideas to deliberative scrutiny.[27] However, this outcome would be contingent on substantial immediate agreement among the agents on what is obviously right ("Claro!" is a conversational putdown that translates as "Obvious!").

Deliberative democrats highlight JD-dv and JD-dd – we discuss such decision problems in sections 3.2 to 3.4. Yet deliberation can focus not only on first-order decisions concerning specific outcomes, but also on second-order decisions concerning institutional arrangements, for instance concerning taxation or welfare provision. Deliberation could reach agreement on the choice of some interpersonally significant standard for assessing people's interests or welfare (e.g. Rawlsian primary goods). On the basis of such a standard, the relevant decision and distribution problems can be interpreted as type ID – section 3.5 discusses such decision problems.

How may deliberation affect people's preferences, views, judgements and social dispositions? Deliberation has informational (inf), argumentative (arg), reflective (ref) and social (soc) aspects. It can

(inf) confront people with new facts, new information or new perspectives on a given issue, as well as corroborate or falsify previously believed facts, information or perspectives;

(arg) draw people's attention to new arguments about the interdependence of issues, confirm or refute the internal consistency of such arguments, make explicit previously hidden premises and assumptions, and clarify whether controversies are about facts, methods and means, or values and ends;

(ref) induce people to reflect on their preferences, in the knowledge that these preferences have to be justified to others;

(soc) create a situation of social interaction where people talk and listen to each other, enabling each person to recognize their interrelation with a social group.

We suggest that each of these aspects plays a role in facilitating the solution of social-choice-theoretic problems.

3.2 Relaxing Strategy-Proofness

In this section we propose the empirical hypothesis that

Hypothesis 1. Group deliberation induces individuals to reveal their preferences and views truthfully.

The hypothesis responds, first, to the attempt to dismiss talk as inconsequential or even directly counterproductive in collective decision-making and, second, to the challenge of the Gibbard-Satterthwaite theorem, according to which the non-existence of SCFs satisfying (S) (together with (U), (R), (D)) implies the inherent vulnerability of democracy to strategic manipulation. *If* the hypothesis is correct, *then* a relaxation of condition (S) becomes an option: the existence of profiles which – in the formal sense of the definition of manipulability in section 2.2 – provide an incentive to submit false preferences poses no problem if individuals do not act on that incentive, i.e. if they have a disposition to be truthful. And, if we relax condition (S), then there are aggregation mechanisms satisfying other reasonable conditions (e.g. (U), (R) and (D)) *and* generating determinate winning alternatives.[28]

To defend the plausibility of the proposed hypothesis, let us briefly recall what the Gibbard-Satterthwaite theorem implies. Any SCF satisfying (U), (R) and (D) will violate (S), and hence there will exist situations with property (SI): An individual is better off from the perspective of their *true* preferences if they reveal *false* preferences than if they reveal their *true* preferences.

We argue that deliberation can have (at least) one of two effects:

Effect 1: *Explicitly changing the incentives.* In a deliberative setting, there may be risks and penalties attached to informational deception and false disclosure of preferences, i.e. untruthfulness may be costly, and hence an individual may no longer be better off if they reveal false preferences. Thus deliberation may transform situations in which, without deliberation, property (SI) would hold into ones in which property (SI) no longer holds.

Effect 2: *Inducing in individuals a greater cooperative disposition.* We have noted in section 2.2 that, when property (SI) holds, decisions about whether or not to submit true preferences may have a structure similar to a (one-shot) prisoners' dilemma, where truthfulness corresponds to cooperation and submission of false preferences corresponds to defection. Deliberation may increase the likelihood that individuals will cooperate, i.e. that they will submit true preferences, and hence that they will no longer act on the incentives provided by (SI).

We now address the two effects in greater detail. Let us first consider effect 1. From a rational-choice standpoint, Austen-Smith analyses talk in strategic terms: it can convey information, but never change preferences, which are prior and exogenous to the decision process.[29] Information conveyed in talk is selective and possibly false, "potentially influential only insofar as it alters individuals' beliefs about how actions map onto consequences".[30] Listeners are aware of the possibility of deception, and so calculate whether or not to believe speakers. If there are no punishments for being exposed as a liar, then there are no incentives for truthfulness.

However, if interaction is recurrent, there are penalties for being exposed as a liar; nobody will believe you the next time.[31] The incentives for truthfulness created by recurrent interaction parallel the incentives for cooperation in repeated (as opposed to one-shot) prisoners' dilemmas.[32]

But even in one-shot interactions a deliberative setting may create incentives for truthfulness. Austen-Smith stresses the informational (inf) aspect of deliberation and how agents will calculate what information to believe, to manipulate and to reveal. First, introducing multiple speakers helps to create incentives for truthfulness by enabling corroboration of information.[33] Second, beyond the informational (inf) aspect of deliberation, the argumentative (arg) and reflective (ref) aspects actually constrain the individuals' opportunities for manipulation.

Consider, for instance, a truly canny actor i intent on taking advantage of the deliberative revelation of preferences by others, while "hanging back" to disclose his/her own preferences selectively and possibly deceptively. Suppose i's true preference ordering is xP_iyP_iz. After listening to others reveal their preferences, i perceives that x has a better chance of beating z than beating y, so i then pretends to have an ordering xP_izP_iy. However, the reflective (ref) aspect of deliberation means that i must justify this false ordering. There is a risk that the others will not believe that i is sincere. The *best* case that i can make for revealing the preference ordering xP_izP_iy late in the day is that he/she has been persuaded of this ordering by the preceding deliberation. Yet such a lie is risky,

because the content of deliberation has actually advanced the standing of *y*, not *z* – otherwise there would be no reason for *i* to act strategically against *y* here. In short, the reflective (ref) aspect of deliberation constrains strategic and deceptive disclosure of preferences.

Now, it might be argued that in partisan settings (such as a legislature) skilful prevarication may actually be admired, at least by members of one's own side, outweighing any incentives to truthfulness. But effect 1 holds to the degree the setting has any deliberative aspects at all – prevaricators risk destruction of their stock of credibility, which in turn is necessary to make any effective *deliberative* interventions. Assemblies are only *completely* partisan in political systems on the verge of disintegration. Thus effect 1 works in "adversary" and "unitary" democracies alike, in Mansbridge's classification.[34]

In our discussion of effect 2, we will now address a different mechanism that may induce truthfulness in one-shot interactions. We argue that deliberation's social (soc) aspect may promote the individuals' cooperative disposition. A robust empirical finding in experiments on one-shot prisoners' dilemmas is that a period of discussion within the group prior to each individual choice between cooperation and defection increases the proportion of cooperative choice.[35] This result falsifies the rational-choice prediction that individuals should still defect, because the payoff structure of the choice situation is unchanged by discussion: indeed, the more confident any one individual becomes during this discussion that others will cooperate, the more that person should calculate that the payoff from defection will increase.

How do we explain discussion-induced cooperation? Discussion provides participants with opportunities for multi-lateral promise-making about the choices they will make.[36] Rational choice theory would predict that in one-shot interactions such promises will be broken, but empirical evidence suggests that social norms and/or psychological dispositions in favour of keeping promises are more powerful. Even when experimental conditions are modified to ensure that the discussion is on topics irrelevant to the choice between cooperation and defection, thus ruling out explicit promise-making, discussion still increases the proportion of cooperators. Proposed explanations range from an evolutionary-psychological or social trust mechanism, by which exposing one's vulnerability to others makes it *less* likely that they will exploit that vulnerability[37], to the idea that people are trust-responsive because there is a positive payoff in being seen by others as cooperative and trustworthy.[38]

Is evidence from stylized dilemma games relevant to the real world of deliberation, particularly its social aspect (soc)? A decision-theoretic explanation of how cooperation can be triggered even in one-shot interactions suggests a mechanism that may work in both laboratories and the real world. People's preferences are not description invariant: an agent's preferences depend not only on a decontextualized payoff matrix, but also on a *decision-frame*, "the decision-maker's conception of the acts, outcomes, and contingencies associated with a particular choice".[39] Varying an agent's frame can lead to major preference changes, up to complete reversal.[40] People express less

self-regarding preferences in games framed in social contexts than in games framed in market contexts.[41] The cooperative disposition of an agent is also connected with the question of whether the agent conceptualizes a given situation in terms of an "I"-frame of self-interest, or a "we"-frame of collective interest.[42] Experimental conditions as undemanding as asking subjects to edit a text by circling all occurrences of "we", "us" and "our" have been shown to induce use of the "we"-frame.[43] We hypothesize that deliberation can have a similar effect of making a "we"-frame of collective interest focal; and that this effect can still occur in adversarial settings, provided only that they have *some* deliberative component.

There is, then, some evidence in support of hypothesis 1: we may expect deliberative talk to be truthful rather than manipulative. And we have argued that, if hypothesis 1 is correct, then this would open up an escape-route from the Gibbard-Satterthwaite theorem.

3.3 Relaxing Universal Domain

The threat Arrow's problem poses to democratic decision-making depends on the level of diversity across different individuals' preferences. In (rare) cases of unanimity, aggregation is easy. But, as noted above, it is overly optimistic to expect deliberation to produce unanimity. Moreover, while unanimity is a *sufficient* condition for avoiding Arrow's problem, it is not a *necessary* one; *preference structuration*, exemplified by Black's condition of *single-peakedness*,[44] is already a sufficient condition. We will discuss the implications of this observation for deliberative democracy.

A profile $\{R_i\}_{i \in N}$ of personal preference orderings is *single-peaked* if there exists a single ordering of all alternatives from 'left'-most to 'right'-most such that each individual has a most preferred position on that 'left'/'right' ordering with decreasing preference for alternatives as they get increasingly distant from the most preferred position. A 'left'/'right' ordering with this property will be a called a *dimension*, labelled Ω.

Single-Peakedness. There exists a bijection $\Omega : X \rightarrow \{1, 2, ..., k\}$ such that, for every triple of alternatives x, y, z and every individual i, if ($\Omega(x) < \Omega(y) < \Omega(z)$) or ($\Omega(z) < \Omega(y) < \Omega(x)$), then xR_iy implies xP_iz.

Table 2 gives an example of two preference orderings over the alternatives x, y, z, v, w which are single-peaked with respect to the same dimension Ω (ordering the alternatives from 'left' to 'right' in the order x, z, v, y, w).

TABLE 2 ABOUT HERE

Let D_S denote the set of all profiles of personal preference orderings satisfying single-peakedness.

Theorem 3. If the number of individuals n is odd[45], there exists a Condorcet winner for each profile contained in D_S.[46]

Theorem 4. If the number of individuals n is odd, there exist SWFs on the domain D_S satisfying (P), (I) and (D), specifically pairwise majority voting.[47]

Theorem 5. If the number of individuals n is odd, there exist SCFs on the domain D^*_S satisfying (R), (D) and (S) (where D^*_S contains all profiles of strict orderings in D_S),[48] specifically pairwise majority voting.[49]

The possibility results of theorems 3, 4 and 5 still hold if D_S is replaced by D_V, where D_V is the set of all profiles of personal preference orderings satisfying the more general (and less demanding) structure condition of *value-restriction*.[50]

Further, while *full* single-peakedness is *sufficient* for the existence of SWFs satisfying (P), (I), (D) and of SCFs satisfying (R), (D), (S), it is not *necessary*. Profiles with sufficiently large single-peaked subprofiles can be included in the domain as well. Niemi has shown that, depending on the total number of individuals, transitive social orderings satisfying the conditions of Arrow's theorem (and by easy implication, the conditions of the Gibbard-Satterthwaite theorem) are likely to exist if only 75% or even fewer of the individuals have personal preference orderings that are structured by the same dimension.[51]

Single-peakedness can be related to a distinction between two different concepts of agreement: *agreement at a substantive level* and *agreement at a meta-level*.[52] Two or more individuals *agree at a substantive level* to the extent that their preferences are the same – the perfect case being unanimity. But it is also possible for two or more individuals to *disagree* on how to rank alternatives, and yet to *agree* on a common dimension in terms of which the alternatives are to be conceptualized. Such agreement is called *agreement at a meta-level*. Agreement at a meta-level may *imply* single-peakedness. If different individuals agree on a common dimension along which each individual's preferences are systematically aligned in the requisite way, then the profile of preferences orderings across these individuals satisfies single-peakedness.[53]

In section 3.3.1 we discuss mechanisms whereby deliberation might induce greater (unidimensional) single-peakedness, via agreement at a meta-level. Allowing that issue-complexity or normative disagreement may often rule out agreement on a common dimension, section 3.3.2 addresses the possibility that identifying multiple relevant issue-dimensions and determining people's

separate dimension-specific preference orderings may generate *intra-dimensional* single-peakedness, and provide a partial solution to Arrow's problem. Because the present escape-route from Arrow's problem rests on empirical hypotheses, section 3.3.3 discusses empirical evidence.[54]

3.3.1 One Dimensional Preference Structuration

In this section, we discuss the hypothesis that

Hypothesis 2. The profile of personal preference orderings $\{R_i\}_{i\in N}$ after a period of group deliberation will satisfy (or approximate) single-peakedness.[55]

It might be argued that rationality requires one determinate dimension on which preferences are single-peaked, whose identification enables a group to criticize as irrational personal preference orderings which have more than one peak on that dimension. As most individuals would eschew appearing irrational, this itself might induce single-peakedness. However, someone may have good reasons for having preferences with more than one peak on the identified dimension; this individual's preferences may be single-peaked along a different, but still reasonably explicable, dimension. Consider a preference ordering in the United States concerning the Vietnam War, preferring massive military action to withdrawal to limited action. If the relevant dimension is extent of military action, the ordering violates single-peakedness. But if the relevant dimension is expected net gain of military action, this ordering is single-peaked. Unless deliberation can produce agreement on one dimension as exclusively relevant, appeal to rationality alone cannot induce single-peakedness.

However, there are other noncoercive ways in which deliberation can narrow the domain of actually occurring preference profiles. Deliberation may rule out arguments that cannot withstand deliberative scrutiny. Deliberative theorists often stress the invocation of interests "generalizable" to deliberators, and to their society. Arguments couched in such terms are more persuasive than those couched in terms of the interests of some subgroup, which in turn are more persuasive than those couched in terms of the interests of specific individuals.[56] Preference orderings denying the personal integrity and political equality of other actual or potential deliberators are not easily sustained, for participation in deliberation has to bring to mind the interests of these others.[57] One kind of generalizable interest is the economist's idea of a public good, which can only be supplied jointly and indivisibly to all individuals, such as ecological integrity. Another kind is access to the basic needs of life (food, shelter, education etc.), formalizable in Sen's notion of functionings or in Rawls's notion of primary goods (on such 'divisible' kinds of generalizable interest, see section 3.5).

But is an appeal to a generalizable interest sufficient to induce (a greater level of) single-peakedness? We suggest that, *if*, through deliberation, (i) a particular generalizable interest becomes focal and (ii) this generalizable interest can be associated with a single dimension, *then* (a high level of) single-peakedness is a likely consequence.[58]

The mechanism can be described as follows. The reflective (ref) and social (soc) aspects of deliberation may lead people to re-frame a given decision problem in terms of a generalizable interest. Supposing (for the moment) this generalizable interest is associated with a single dimension (e.g. ecological sustainability), then the informational (inf) and argumentative (arg) aspects of deliberation may resolve factual disagreements on how alternatives are aligned on that dimension (e.g. which options least or most degrade an ecosystem). Rationality may finally lead individuals to have single-peaked preferences on the shared dimension. This mechanism requires only agreement at a meta-level (i.e. on a shared dimension), not at a substantive level (i.e. on the most preferred position on that dimension).[59]

A sceptic might counter that invocations of generalizable interests are just rationalizations for self-interest. So Riker castigates, for example, "an assertion of the general virtue of rural life on the family farm [that] justifies farm subsidies."[60] Yet rationalization can still shift debate to a single publicly-sustainable issue-dimension. On farm subsidies, a deliberator might question with regard to this dimension why subsidies should also go to large agribusinesses. Imagine three policy alternatives: x = subsidies for all farms, y = subsidies for family farms only, z = no subsidies. Assuming material self-interest, the preference ordering for agribusiness is xP_izP_iy (z is preferred to y because subsidies for family farms only would hurt the competitive position of agribusiness), for family farmers yP_ixP_iz, and for taxpayers zP_iyP_ix. If none of the three groups controls a majority but each pair of them does, then there is a cycle across x, y and z. But if deliberation induces a need for rationalization along Riker's lines, agribusiness's preference ordering cannot be sustained, and the cycle is broken. In short, deliberation-induced rationalization may narrow the domain of preference profiles.

Elster speaks of the "civilizing force of hypocrisy" accompanying deliberation.[61] Perhaps becoming civilized in Elster's terms involves 'finding one's peak' on the publicly identified issue-dimension. A more sanguine view of motivation would see individuals truly adopt positions they can sustain publicly.[62] Even if individuals privately cling to their original preference orderings, proposals are likely to be crafted only in response to publicly sustainable orderings.[63]

To illustrate, consider the preference ordering of an individual selected as one of the mandated environmental representatives on the Resource Advisory Council for Eastern Washington set up by the Federal Bureau of Land Management under the Rangeland Reform of 1995. His background, and arrest record, was in the radical environmental group Earth First!, whose slogan is "No compromise in defense of Mother Earth!". The Earth First! preference ordering violates single-peakedness with respect to the dimension "extent of wilderness preservation": members prefer wilderness preservation to desecration to compromise development, on the grounds that once it loses its pristine character, wilderness might as well be trashed to drive home the point. Participating in this deliberative forum, this Earth First!er eventually co-wrote (along with a cattle rancher) most of the Guidelines for Range Management produced by the Council.[64] No doubt preservation remained his

first choice, but compromise was now preferred to desecration. Thus this individual found his peak – and lost his pique.

A skeptic could allow that, for JD-dv situations, individuals might indeed be constrained in their public expression of preference orderings – yet still vote based on different private preference orderings. If so, our proposed mechanisms would ultimately provide no solution to Arrow's problem via single-peakedness. However, a necessary (though not sufficient) condition for this objection is that voting be secret. When voting is public, it is implausible that individuals would vote one way while simultaneously talking another way; as we argued in section 3.2, individuals who engage in such perceived deception are penalized in deliberation. As Brennan and Pettit point out, "unveiling the vote" induces individuals to vote in "discursively defensible manner".[65] But even in JD-dv situations with secret voting, the social aspect (soc) of deliberation may induce a cooperative disposition towards expressing publicly oriented rather than self-regarding preferences (see section 3.2). As we note in section 3.3.3, this claim is supported by empirical evidence from Fishkin's deliberative polls.

The present arguments are applicable *if* deliberation leads to the identification of a single shared dimension. Issue complexity may rule this out. Rational choice theorists might agree that the demanding part of our argument is *not* the claim that a focus on a single dimension would induce single-peakedness, but rather the antecedent condition: as Mueller holds, "[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".[66] Even a deliberation-induced focus on "generalizable interests" does not necessarily solve the problem, for individuals might still disagree about what is in the public interest, or whether (say) ecological integrity or economic growth should receive priority when these public interests clash.

However, as we see in section 3.3.3, empirical evidence supports hypothesis 1 – even in its above stated onedimensional form.

3.3.2 Multidimensional Preference Structuration

As we argued in section 3.3.1, *if* preferences focus on a single dimension, *then* single-peakedness is not a demanding requirement. Thus if a profile of personal preference orderings $\{R_i\}_{i \in N}$ violates single-peakedness, the reason might be that preferences are determined by more than one dimension. This leads us to propose the following hypothesis:

Hypothesis 3.

(i) Group deliberation leads to the identification of those issue-dimensions that are considered relevant to a given decision problem, represented by the numerals 1, 2, …, k (the *identification of dimensions condition*).

(ii) For each individual i in N, i's preferences can be represented by (a) a vector $<R_{i1}, R_{i2}, ..., R_{ik}>$ of dimension-specific preferences, where, for each dimension j, the ordering R_{ij} represents individual i's preferences with respect to dimension j, and (b) a specification of the relative importance individual i attaches to each of the k dimensions (the *disaggregation condition*).[67]

(iii) For each dimension j, the corresponding *dimension-specific* profile of personal preference orderings across individuals, $\{R_{ij}\}_{i \in N} = \{R_{1j}, R_{2j}, \ldots, R_{nj}\}$, after a period of group deliberation will satisfy (or approximate) single-peakedness (the *intradimensional single-peakedness condition*).[68]

Why is the hypothesis plausible? Firstly, in deliberation one must give reasons for preferences or appeal to generalizable interests, and this itself is conducive to the identification of relevant issue-dimensions. These dimensions may be uncovered by, or created through, deliberation; empirically, it is hard to distinguish between these two possibilities. Secondly, if deliberation can disaggregate a profile $\{R_i\}_{i \in N}$ into k separate profiles $\{R_{i1}\}_{i \in N}$, $\{R_{i2}\}_{i \in N}$, ..., $\{R_{ik}\}_{i \in N}$, each focusing only on a single issue-dimension, each such dimension-specific profile is more likely to be single-peaked than the original profile. Disagreements about issue-priorities are by definition 'factored out' *in each dimension-specific profile*, and hence there is less potential for the type of disagreement leading to violations of single-peakedness *in each dimension-specific profile*. The mechanisms identified in section 3.3.1 are thus likely to be effective *for each dimension-specific profile*, because their antecedent condition – namely the focus on a single dimension – is satisfied.

To illustrate, imagine a committee having difficulties deciding what level of tariffs to impose on imports. One person claims a first preference for free trade, a second preference for a blanket tariff, a third preference for a selective tariff. Upon questioning, she might say that she prefers zero to blanket to selective tariffs because a selective tariff will cause inequity across industrial sectors. Thus she reveals two dimensions: open/closed trade and equity/inequity – with a greater importance placed on the former – on both of which her preference ordering has only one peak. Other members might then rank their preferences on both these dimensions, producing a single-peaked profile on each dimension (assuming no complications from these other preference orderings). They still face the problem of aggregation across these two dimensions, but we turn to this issue below.

Given the intradimensional single-peakedness condition of hypothesis 3, theorems 4 and 5 imply that separate dimension-specific aggregation can yield a social ordering or a socially most preferred alternative (particularly, a Condorcet winner) *for each dimension* in accordance with Arrow's conditions (I), (P), (D) or the Gibbard-Satterthwaite conditions (R), (D) and (S).

There are at least four ways in which dimension-specific aggregation can help solve a collective decision problem.

Subdividing the decision. For decision problems which can be subdivided into several independent dimension-specific sub-decisions, identification of the dimensions of these sub-decisions and subsequent dimension-specific aggregation might be sufficient, provided that the relevance of

each of these dimensions is publicly accepted and that deliberation can induce sufficient single-peakedness in each dimension-specific preference profile. To illustrate, it has long been observed in negotiations that an issue can be made more tractable by the introduction of multiple dimensions. The peace process in Northern Ireland follows such a course. If confined to the single dimension of sovereignty, the issue is intractable, with effective majorities of key actors against every conceivable alternative. But once dimensions are introduced such as amnesty for politically motivated crimes, civil rights, cross-border bodies, electoral systems, guaranteed representation for particular groups, an effective super-majority for a settlement could be constituted.

Lexicographic hierarchies of dimensions. Often decisions come in 'packages' and cannot easily be subdivided into separate dimension-specific sub-decisions such as one on amnesty, one on civil rights, one on electoral systems, etc. In such cases, mere unpacking of dimensions does not solve the problem of what *overall* decision to take. While people may agree on how alternatives are aligned on the dimensions of (a) ecological sustainability, (b) employment, and (c) economic growth, they may still disagree on the relative importance of each dimension. In the most general case, the problem of aggregating k dimension-specific profiles into a single social ordering, rather than into k dimension-specific orderings, raises problems similar to Arrow's problem, even if each dimension-specific profile satisfies single-peakedness.

Theorem 6.[69] Let F be an aggregation function whose input is a vector of k dimension-specific profiles of personal preference orderings $<\{R_{i1}\}_{i\in N}, \{R_{i2}\}_{i\in N}, ..., \{R_{ik}\}_{i\in N}>$ and whose output is a *social ordering* R over the alternatives in X. Suppose F is defined for all vectors of dimension-specific profiles where each dimension-specific profile satisfies single-peakedness (i.e. the domain that would result from hypothesis 3), and suppose F satisfies (P) and (I).[70] Then F will make one dimension *dominant*: there exists a fixed dimension j such that, for all inputs in the domain of F and all x and y in X, if all individuals rank x above y in dimension j – i.e. $xP_{ij}y$ for all i –, then xPy.

The only aggregation functions satisfying the conditions of theorem 6 are (possibly lexicographic) hierarchies of dimensions. The overall social ordering is determined, first, exclusively on the basis of the dimension-specific profile $\{R_{ij}\}_{i\in N}$ corresponding to the highest-ranked dimension; only if there are ties, the dimension-specific profile corresponding to the second-highest-ranked dimension acts as a tie-breaker; only if there are still ties, the dimension-specific profile corresponding to the third-highest ranked dimension acts as a tie-breaker, and so on. A *lexicographic hierarchy of dimensions* might solve collective decision problems when deliberation can generate agreement on a lexicographic order of importance of different issue-dimensions. In particular, the weights the individuals attach to each issue-dimension could be aggregated into such an overall lexicographic ordering.[71] But this aggregation problem is not immune to the Arrow and Gibbard-Satterthwaite problems either, and democracies may feel apprehensive about solving cross-dimensional aggregation problems involving difficult matters of issue-priority by a mechanical

decision procedure. Still, one could imagine debate on the relative weights of the dimensions of (say) ecology, employment and business. Because people's views on such weights are as changeable by deliberation as their preferences on alternatives, deliberation might produce agreement on a lexicographic hierarchy.

Logrolling. In practice, problems of cross-dimensional aggregation are often solved by logrolling rather than by explicit assignment of weights to dimensions. Under logrolling, intense minorities, as potential coalition partners in a majority coalition, can get their way on particular dimensions by threatening to block overall agreement. From a manipulability perspective, logrolling provides opportunities for misrepresentation of preference intensity. From a rational-choice perspective, it is excoriated for enabling coalitions of special interests to prevail over general interests. But logrolling also provides information about preference intensity[72] and thus about what weights agents attach to different issue-dimensions. In deliberation, especially given (arg) and (ref), individuals have to justify their preferences in terms acceptable to others. One such way of justifying preference intensity is for an individual to indicate what he/she would be prepared to accept on other dimensions in return for getting his/her way on the dimension he/she cares about most. While this individual might still misrepresent this intensity, as we saw in section 3.2 lying about one's own preferences is potentially costly in deliberation.

Demonstrating the nature of the problem and crafting new alternatives. Even if there is no straightforward solution to the cross-dimensional aggregation problem, the identification of the precise nature of the problem at issue, as offered by social-choice-theoretic analysis, is much more desirable than mere recognition of inconsistency as a result of a lack of structure in individual preferences. If at this juncture a seeming impasse is reached, at least two different conclusions can be drawn. One is that the information contained in the available profile(s) of personal preference orderings (even after unpacking issue-dimensions) is insufficient for reaching a decision in accordance with Arrow's conditions[73] – see section 3.5. Alternatively, the information from unpacking dimensions could be used to craft new alternatives, not yet contained in X. In our tariff example, the committee might, for instance, think about compensatory taxes and subsidies that would restore equity across industrial sectors should a selective tariff be adopted, or contemplate ways a non-discriminatory restricted tariff could be designed.[74]

3.3.3 Empirical Evidence

Although the empirical evidence on how deliberation affects preference structuration is limited, some recent findings corroborate our hypotheses in sections 3.3.1 and 3.3.2.

On the hypothesis that deliberation can increase endorsement of "generalizable interests", Gundersen, reporting a series of "deliberative interviews" on environmental issues, finds that in every case deliberation promoted commitment to the generalizable interest of ecological integrity.[75]

On the hypothesis that deliberation induces greater single-peakedness, Pelletier et al administered Q-sorts to participants before and after a deliberative "search conference" on food supply policy in upstate New York.[76] Although this study does not measure single-peakedness as such, one result is that individuals who subscribed to two distinct positions before deliberation tended to subscribe to one or other of them (but not both) after deliberation.[77] This might suggest that deliberation leads individuals to identify more closely with their "peak". Fishkin's deliberative polls provide data for testing the hypothesis that preferences after deliberation exhibit a greater level of single-peakedness than preferences before deliberation. Fishkin's own research indicated that deliberation induces substantial preference shifts**.** The significance of these shifts for single-peakedness in the unidimensional sense is investigated by List, McLean, Fishkin and Luskin.[78] Two indices of single-peakedness are applied to individual preference data collected before and after deliberation. For several deliberative polls on Texas energy policy, deliberation substantially increased the level of single-peakedness, thus providing empirical support for hypothesis 2 above. For a deliberative poll on the 1999 proposal in Australia to replace the monarchy with a republic, the level of single-peakedness was high both before and after deliberation, though the Condorcet winner changed from a republic with a directly elected president to a republic with an appointed president. Although the effect of deliberation on the indices of single-peakedness was less marked than in the Texas polls, the Australian results remain encouraging because in a situation where substantial preference shifts occurred, and where a cycle over three options was a live possibility (the third option being the status quo), deliberation appears to have protected against a loss of structure.

Finally, the claim that unpacking multiple relevant issue-dimensions facilitates *intra-dimensional preference structuration* is compatible with a study by Budge, Robertson and Hearl, showing that the main point of disagreement and competition in politics is less what the best policy option *within a specific issue-dimension* is, more what issue-dimensions are relevant.[79]

3.4 Relaxing Independence of Irrelevant Alternatives

We now suggest:

Hypothesis 4. Group deliberation can lead individuals to reach agreement on the content of the set X of relevant alternatives.

The hypothesis responds to the challenge that the non-existence of SWFs satisfying (I) (together with (U), (P), (D)) implies the inherent vulnerability of democracy to manipulability by changes of the agenda X. *If* the hypothesis is correct, *then* a relaxation of condition (I) becomes an

option. If the individuals agree on what the set of relevant alternatives X is, then there is little scope for an agenda-setter strategically to introduce or delete alternatives from X so as to affect the relative chances of other alternatives in X, and the logical possibility of agenda manipulation (implied by violations of (I)) no longer poses a problem. And, if we relax condition (I), then there exist SWFs generating transitive social orderings and satisfying (U), (P) and (D).

Positional rules, which are sensitive to the position of an alternative in each individual's ranking R_i, are well-known such SWFs.[80] The Borda rule is the most famous example.[81] Given $\{R_i\}_{i \in N}$, the Borda rule determines a social ordering R as follows: for each alternative x in X and each individual i in N, let $B_i(x)$ be the number of alternatives y in X such that xP_iy; for every x and y in X,

xRy if and only $B_1(x)+B_2(x)+...+B_n(x) \geq B_1(y)+B_2(y)+...+B_n(y)$.

Why is hypothesis 4 plausible? The composition of the set of alternatives can itself be subjected to deliberation (or possibly to a decision by consensus or approval voting). Consider some experiments designed by Plott and Levine to demonstrate agenda manipulation, which involve subjects with induced preferences that they are told to keep secret from the others.[82] The subjects take a series of votes on how a set of five alternatives is to be partitioned; each subset is then voted on, and its winner goes to the next step, eventually producing an overall winner. The experiments generally went as expected, failing to choose the Condorcet winner and revealing ubiquitous agenda manipulation, which is why Riker says they "impress me deeply".[83] However, there was one exception: when the chair of one group allowed a straw vote at the outset, the group then converged on the Condorcet winner. This straw vote revealed that one of the five alternatives was least preferred by everyone, and so led to dropping that alternative from the set of relevant alternatives. This straw vote was functionally equivalent to the use of deliberation for demarcating the set of relevant alternatives, showing how such demarcation can limit the scope for agenda manipulation.

This process may itself seem vulnerable to strategic manipulation, though only to the extent individuals behave according to rational-choice precepts. If they do not, then deliberation can be used to decide on a procedure for distinguishing between relevant and irrelevant alternatives – conceivably by approval voting or by consensus. Once the group has decided which alternatives are relevant, positional rules such as the Borda rule may be attractive aggregation mechanisms (and so perhaps legislatures should adopt this solution). Discussion about what alternatives are relevant seems inherent in decision processes of types JD-dd and JD-dv.

The fact that deliberators *could* decide to restrict the set of relevant alternatives does not mean that they *will* do so. But there is a mechanism intrinsic to deliberation that promotes this likelihood, relating to the argumentative (arg) and reflective (ref) aspects. In deliberation, an individual cannot simply introduce or support an alternative; he/she must justify that preference in terms others can

accept. Arguing strategically for an alternative that the individual does not truly support carries two risks. The first is that the argument may persuade others, inducing the group to choose that alternative. The second is that the individual may be exposed as a liar (either in argument or by subsequent voting behavior) – potentially incurring punishment as discussed in section 3.2.

To illustrate, assume the Borda rule is being used and there are three alternatives, x, y, z, with 65 deliberators preferring x to y to z, and 35 deliberators preferring y to z to x. With Borda scores of 130 for x, 135 for y, and 35 for z, a canny individual i sees that his favored alternative x will be beaten by y, but that introducing alternative w, which the 64 other proponents of x are likely to prefer to y, could change the situation. But now i must justify his (false) preference for w over x. Individual i may be prepared to lie here, but if the lie fails to convince others then i will suffer the penalties we mentioned in section 3.2. Other deliberators may well ask i why he chose to introduce w so late in the day; this, too, needs to be justified by i, and again there are penalties for lying. Moreover, if i has to argue against x to justify w, he risks convincing others that x (his true preference) is indeed undesirable.

In conclusion, if deliberation can demarcate the set of relevant alternatives in a publicy acceptable way, a violation of condition (I) may become defensible, and positional rules may become attractive SWFs.

3.5 Introducing More Information

While the responses discussed so far involve relaxation of one of Arrow's conditions, we now discuss a solution compatible with preserving, even strengthening, all the conditions. This may seem paradoxical: for Arrow shows that these conditions are mutually inconsistent. However, Sen argued that Arrow's impossibility result is driven by the informational limitations of Arrow's ordinalist framework.[84] If we allow interpersonal comparisons of preference intensity or of a suitable individual welfare measure, then there exist aggregation mechanisms satisfying all of Arrow's conditions.

We now assign to each individual i a *personal welfare function* $W_i : X \to \mathbf{R}$, where $W_i(x)$ is interpreted as a measure of the welfare of individual i under alternative x. A *social welfare functional* (SWFL as distinct from a SWF) is a function F whose input is a *profile of personal welfare functions* $\{W_i\}_{i \in N}$ and whose output is a social ordering R.[85] Assumptions on measurability and interpersonal comparability of welfare are formalized by specifying the class of transformations with respect to which a SWFL is required to be invariant.[86] This class is also interpreted as the class of transformations up to which a profile of personal welfare functions is taken to be unique. The smaller this class of transformations, the more information is contained in a profile.

Two such assumptions are *ordinal measurability with interpersonal comparability of welfare levels* (OLC) and *cardinal measurability with interpersonal comparability of welfare units* (CUC).[87]

Informally, (OLC) is the assumption that comparisons of the form "individual *i* under alternative *x* is at least as well off as individual *j* under alternative *y*" are meaningful, and (CUC) is the assumption that comparisons of the form "if we switch from alternative *x* to *y*, the ratio of individual *i*'s welfare gain (or loss) to individual *j*'s welfare gain (or loss) equals *a*" are meaningful. Other informational assumptions have been discussed.[88]

Theorem 7. (i) There exist SWFLs satisfying (OLC), (U), (P), (I) and (D); (ii) There exist SWFLs satisfying (CUC), (U), (P), (I) and (D).[89]

Examples of SWFLs satisfying the conditions of parts (i) and (ii) of theorem 7 are, respectively, the *leximin rule* and the *utilitarian rule*.[90]

The leximin-rule is a version of Rawls's (lexicographic) difference principle: make social choices to maximize the welfare-level of the worst-off individual; if there are ties, maximize, in a lexicographic order of priority, the welfare-levels of the second worst-off, third worst-off, ..., individuals.[91] The leximin rule satisfies not only the conditions of theorem 7(i), but also more demanding conditions of anonymity, positive responsiveness, separability and minimal equity.[92]

The utilitarian rule is a version of the classical utilitarian principle: maximize the sum-total of the welfare of all individuals. The utilitarian rule satisfies not only the conditions of theorem 7(ii), but also anonymity, positive responsiveness, separability and continuity.[93]

How could this escape-route from Arrow's theorem apply in a deliberative democracy? Deliberation can concern not only first-order decisions on outcomes, but also second-order decisions on the design of institutions for solving (first-order) decision problems, such as the allocation of resources or distribution of benefits and burdens. A health care provider's decision on how much treatment to allocate to each patient, given limited resources, is a first-order decision of type ID. The provider's allocation of medical resources is based not on the expressed judgements or preferences of patients (although patients' expressions inform medical diagnoses), rather on an evaluation of different patients' needs by the (external) evaluation standard of the medical examinations of patients. The health authority's decision on what *principle* should govern such first-order health care allocation problems is the corresponding second-order decision concerning institutional design.

Institutionally soluble first-order decision problems (particularly on the allocation or distribution of resources) can be interpreted as social choice problems of type ID consisting of two components: (i) an evaluation variable for assessing the interests of the individuals, and (ii) a decision principle for aggregating these individual interests (as assessed by (i)) into a collective outcome. The evaluation according to (i) can be formalized by a profile of personal welfare functions $\{W_i\}_{i \in N}$; and the decision principle in (ii) by a SWFL *F*.

The escape-route from Arrow's theorem via introducing more information is available if the evaluation variable specified under (i) is interpersonally comparable. For certain types of evaluation variables, like the economist's 'revealed preferences' or 'utility', such comparisons may not be meaningful. Robbins famously argued that "[i]ntrospection does not enable A to measure what is going on in B's mind, nor B to measure what is going on in A's. There is no way of comparing the satisfactions of two different people".[94] But even if we concede Robbins's claim, this entails *not* that the present escape-route is unavailable, only that it is necessary to use a (normative) evaluation variable *other than utility or subjective satisfaction* for which interpersonal comparisons are meaningful.

Several such evaluation variables have been proposed, including Rawls's index of primary goods – non-mentalistic and measurable from the perspective of an external observer;[95] Sen's functionings and capabilities – again non-mentalistic and externally measurable;[96] and the United Nation's Human Development Index (HDI), combining life expectancy, educational attainment and income. Using a Rawlsian index of primary goods as the evaluation variable and the leximin rule as the decision principle will satisfy the conditions of theorem 7. Depending on what type of measurability and interpersonal comparability a chosen evaluation variable permits, several different aggregation procedures are available.

The task of deliberation can then be summarized as follows.

Hypothesis 5. Group deliberation in second-order decisions on institutional design can lead individuals to reach agreement on (i) an evaluation variable for assessing individual interests that is interpersonally comparable, and (ii) a decision principle for aggregating individual interests (as assessed by the chosen evaluation variable) into a collective outcome.

Is there any mechanism intrinsic to deliberation making it likely that suitable arguments for (i) and (ii) will be made and accepted? Rawls's original position or Habermas's ideal speech situation are *hypothetical* deliberation situations involving precisely such arguments. In these hypothetical situations, one mechanism is that the argumentative (arg) and reflective (ref) aspects of deliberation impose justification and transparency constraints on the outcomes of second-order decisions such as (1) formal generality, (2) universal applicability, (3) publicity, (4) finality and (5) ordering. In Rawls's original position,[97] the effect of these constraints is to narrow down the range of acceptable institutional arrangements. An interpersonally comparable evaluation variable – like an index of primary goods or an index of functionings – may seem more transparent and publicly justifiable than a publicly inscrutable (and interpersonally non-comparable) mentalistic evaluation variable. Similarly, a well-defined and systematic aggregation mechanism may seem more consistent with constraints (1) to (5) than a non-principled decision method based, for instance, on discretion.

Gutmann and Thompson identify reciprocity, the need for individuals to provide each other with mutually acceptable reasons for favoured decisions, as a principle resulting from the reflective

(ref) aspect of deliberation.[98] Under reciprocity, "citizens as well as theorists consider what justice requires in the case of specific laws".[99] Gutmann and Thompson argue that deliberators as well as theorists can derive substantive principles – including principles of distributive justice – from the reciprocity idea. These include, for example, choosing "basic opportunities" as an evaluation standard. Though treated by Gutmann and Thompson as subject to deliberative revision, basic opportunities resemble Rawls's primary goods in that they are foundational to all individual life projects.

Beyond these hypothetical examples, some real-world systems of government have produced results of the form of (i) and (ii). Corporatist government involves cooperation between encompassing labor and business federations, previously locked in a zero-sum distributional game. The agreements produced often constitute enduring distributive rules, with both an evaluation variable and an aggregation principle. For example, in Austria (the corporatist archetype) the Parity Commission for Wages and Prices set up in 1957 made its decisions so as to maximize growth in national income while compensating those adversely affected by the resulting structural adjustment.[100] The evaluation variable is therefore income measured in conventional terms. The aggregation principle is a type of utilitarian Pareto criterion: maximize average income and fully compensate losers. The "social corporatist" Nordic countries, in contrast, favour an evaluation variable more akin to Rawlsian primary goods, with an income-invariant distribution of equal entitlements to these goods.[101] A recently introduced British deliberative body, the National Institute for Clinical Excellence (NICE), develops principles in the form of (i) and (ii) for allocating expensive health care. NICE uses "basic opportunities" as an evaluation variable, determining whether to provide public funds for new drugs in terms of how different funding schemes for the drug would affect basic opportunities of citizens.[102]

We do not suggest that deliberative democracy should convert all collective decision problems into type ID, only that, given democratic agreement, this route *can* be taken for some allocation and distribution problems. The arguments of the previous sections apply to the many remaining decision problems of type JD (possibly including second-order decisions on how to specify (i) and (ii)).

4. Conclusion

We have argued that the seemingly conflicting approaches of deliberative democracy and social choice theory can be reconciled. Deliberation facilitates pursuit of several escape-routes from the impossibility results commonly invoked by social-choice-theoretic critics of democracy. The shift from purely voting-based decision-making to decision-making based on the informational, argumentative, reflective and social aspects of deliberation opens up many possibilities for

meaningful collective decisions. To summarize, the identified escape-routes from social-choice-theoretic impossibility problems are the following:

(i) *If* deliberation induces individuals to reveal their preferences and views truthfully (hypothesis 1), *then* strategic manipulation becomes less of a threat in deliberation, and a relaxation of condition (S) provides an acceptable escape-route from the Gibbard-Satterthwaite theorem, compatible with all other conditions of the theorem.

(ii) *If* deliberation induces preference structuration – narrowing the domain of actual preference profiles to a domain in which the Arrow and Gibbard-Satterthwaite problems do not apply – (hypothesis 2), *then* both cycling and strategic manipulation become less of a threat in deliberation, and relaxation of condition (U) provides acceptable escape-routes from Arrow's theorem and from the Gibbard-Satterthwaite theorem, compatible with all other conditions of these theorems.

(iii) *If* deliberation helps uncover or create the tacit issue-dimensions that 'cause' a lack of preference structuration, and induce greater preference structuration in each separate dimension (hypothesis 3), *then* dimension-specific aggregation in accordance with all of the conditions of Arrow's theorem or the Gibbard-Satterthwaite theorem (except condition (U)) becomes possible, and one of the following solutions to the overall decision problem *may* become available: subdividing the decision, lexicographic hierarchies of dimensions, logrolling, or demonstrating the nature of the problem and creatively crafting new alternatives.

(iv) *If* deliberation can produce agreement on what the set of relevant alternatives is (hypothesis 4), *then* agenda manipulation becomes less of a threat, and relaxation of condition (I) provides an acceptable escape-route from Arrow's theorem, compatible with all other conditions of the theorem.

(v) *If* deliberation can produce agreement on an interpersonally comparable evaluation variable for assessing individual interests and a decision principle for aggregating individual interests into a collective outcome (hypothesis 5), *then* a solution to Arrow's problem that is consistent with *all* of Arrow's conditions becomes available for a range of institutionally soluble collective allocation or distribution problems.

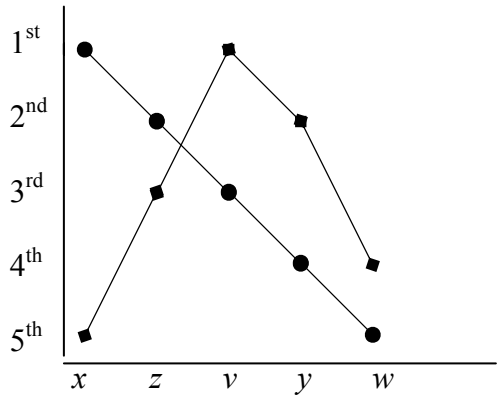
The role of deliberation is to bring about situations in which the antecedents of these "if-then" results are satisfied. The normative component of our argument is that the constraints required for bringing about such situations are either inherent in, or at least consistent with, core deliberative principles. Moreover, our arguments rest on empirical hypotheses about the effects of deliberation. Whether or not deliberation will induce each of the antecedents of the "if-then" results ultimately depends on the specifics of particular cases.

While we have adduced available empirical evidence and illustrations in support of our hypotheses, more research is necessary to investigate deliberation's strength on each of the identified escape-routes from the social-choice-theoretic impossibility problems. But for the moment we conclude that deliberative democracy and social choice theory are mutually supportive. The former is concerned with identification of the functions that deliberation ought to, and indeed can, perform in democratic decision making, and the latter is concerned with the clarification of the logical properties of available procedures for solving the aggregation aspects of democratic decision problems. Thus social choice theory shows exactly what deliberation must accomplish in order to render collective decision making tractable and meaningful, suggesting that democracy must in the end have a deliberative aspect.

Table 1: A prisoners' dilemma of manipulability

		Coalition *M*	
		submit 'true' preferences	submit 'false' preferences
individual *i*	Submit 'true' Preferences	2nd best outcome for *I* 2nd best outcome for *M*	worst outcome for *I* best outcome for *M*
	Submit 'false' Preferences	best outcome for *I* worst outcome for *M*	2nd worst outcome for *I* 2nd worst outcome for *M*

Table 2. Single-peaked personal preference orderings



Notes

[1] See, among many others, James Bohman and William Rehg, eds, *Deliberative Democracy: Essays on Reason and Politics* (Cambridge, MA: MIT Press, 1997); Joshua Cohen, 'Deliberation and Democratic Legitimacy', in Alan Hamlin and Philip Pettit, eds, *The Good Polity: Normative Analysis of the State* (Oxford: Basil Blackwell, 1989), pp. 17-34; John S. Dryzek, *Discursive Democracy: Politics, Policy and Political Science* (New York: Cambridge University Press, 1990); John S. Dryzek, *Deliberative Democracy and Beyond: Liberals, Critics, Contestations* (Oxford: Oxford University Press, 2000); Jon Elster, 'Introduction' in Jon Elster, ed, *Deliberative Democracy* (New York: Cambridge University Press, 1998), pp. 1-18; James Fishkin, *Democracy and Deliberation* (New Haven: Yale University Press, 1991); Amy Gutmann and Dennis Thompson, *Democracy and Disagreement*. (Cambridge, MA: Harvard University Press, 1996).

[2] Jürgen Habermas, *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (Cambridge, MA: MIT Press, 1996); John Rawls, 'The Idea of Public Reason Revisited', *University of Chicago Law Review,* 94 (1997), 765-807, pp. 771-2.

[3] Kenneth Arrow, *Social Choice and Individual Values* (New York: Wiley, 1951; 2nd edition 1963).

[4] William H. Riker, *Liberalism Against Populism* (San Franscisco: W. H. Freeman, 1982).

[5] Russell Hardin, 'Public Choice versus Democracy', in David Copp, Jean Hampton, and John E. Roemer, eds, *The Idea of Democracy* (Cambridge: Cambridge University Press, 1993), pp. 157-72, at p. 170.

[6] But see David Miller, 'Deliberative Democracy and Social Choice', *Political Studies*, 40 (1991) (special issue), 54-67.

[7] For example, Riker, *Liberalism Against Populism*.

[8] Arrow, *Social Choice and Individual Values*.

[9] $|N|>1$ and $|X|>2$.

[10] R_i is assumed to be reflexive, connected and transitive.

[11] R is also assumed to be reflexive, connected and, unless stated otherwise, transitive.

[12] Arrow, *Social Choice and Individual Values*.

[13] *Quasi-transitivity* is the requirement that the strong ordering P induced by R be transitive, while R itself need not be transitive. A SWF is *oligarchic* if there exists a subset M of N such that (i) xPy if, for *all* i in M, xP_iy, and (ii) xRy if, for *some* i in M, xP_iy. Any SWF generating quasi-transitive social orderings which satisfies (U), (P) and (I) is *oligarchic.* For a proof, see Amartya Sen, *Collective Choice and Social Welfare* (San Francisco: Holden-Day, 1970), pp. 76 -7.

[14] Thus the difference between SWFs and SCFs is simply that the former generate social orderings, while the latter generate single winning alternatives.

[15] To restate condition (D) for SCFs (rather than SWFs), we simply define a dictatorship as follows: F is *dictatorial* if there exists an i in N such that, for all $\{R_i\}_{i\in N}$ in the domain of F and all x in the range of F, yR_ix, where $y=F(\{R_i\}_{i\in N})$.

[16] A. Gibbard, 'Manipulability of Voting Schemes: A General Result', *Econometrica,* 41 (1973), 587-601; M. Satterthwaite, 'Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions', *Journal of Economic Theory,* 10 (1975), 187-217.

[17] Prasanta Pattanaik, *Strategy and Group Choice* (Amsterdam: North-Holland, 1978).

[18] Even under full information about the true preferences of others, the individuals may have insufficient computational power to recognize possibilities of manipulation. See J.J. Bartholdi, C. A. Tovey, and M. A. Trick, 'Voting schemes for which it can be difficult to tell who won the election', *Social Choice and Welfare* 6, (1989), 157-65 for a theoretical result; and Glenn W. Harrison and Tanga McDaniel, 'Voting Games and Computational Complexity', Economics Working Paper B-01-01, Moore School of Business, University of South Carolina (2001) for an empirical result.

[19] Cass Sunstein, *Democracy and the Problem of Free Speech* (New York: Free Press, 1993), p. 244.

[20] Jon Elster, 'The Market and the Forum', in Jon Elster and Aanund Hylland, eds, *Foundations of Social Choice Theory* (Cambridge: Cambridge University Press, 1986), pp. 103-32, at p. 112.

[21] Miller, 'Deliberative Democracy and Social Choice'.

[22] David van Mill, 'The Possibility of Rational Outcomes from Democratic Discourse and Procedures', *Journal of Politics* 58 (1996), 734-52.

[23] Jack Knight and James Johnson, 'Aggregation and Deliberation: On the Possibility of Democratic Legitimacy', *Political Theory*, 22 (1994), 277-96, at pp. 282-3.

[24] Knight and Johnson, 'Aggregation and Deliberation', p. 289.

[25] Amartya Sen, *Choice, Welfare and Measurement* (Oxford: Basil Blackwell, 1982), ch. 8.

[26] Fishkin, *Democracy and Deliberation*; James Fishkin, *The Voice of the People: Public Opinion and Democracy* (New Haven: Yale University Press, 1995).

[27] Diego Gambetta, ''Claro!': An Essay on Discursive Machismo', in Jon Elster, ed, *Deliberative Democracy* (New York: Cambridge University Press, 1998), pp. 19-43.

[28] Our claim in this section is not that deliberation leads individuals to agree on what choices should be made. Nor is it our claim that inducing individuals to submit true preferences rules out the occurrence of profiles which lead to cycles *under pairwise majority voting*. Indeed, true preferences may still lead to such cycles. Rather, our claim is only that deliberation can induce individuals to submit true preferences, and that this will make a relaxation of condition (S) more acceptable. The aggregation mechanisms that become available under a relaxation of condition (S) will be *different from pairwise majority voting* (and in particular cycling-free): they will be the kinds of mechanisms

that are often criticized on the grounds that they provide incentives for strategic misrepresentation of preferences. Examples are the Borda count (to be discussed below) or *single transferable vote* (STV).

[29] David Austen-Smith, 'Information Transmission in Debate', *American Journal of Political Science,* 34 (1990), 124-52. David Austen-Smith, 'Strategic Models of Talk in Political Decision-Making', *International Political Science Review,* 16 (1992), 45-58.

[30] Austen-Smith, 'Strategic Models of Talk', p. 57.

[31] Gerry Mackie, 'All Men are Liars: Is Deliberation Meaningless?', in Jon Elster, ed, *Deliberative Democracy* (New York: Cambridge University Press, 1998), pp. 97-122.

[32] Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984).

[33] Mackie, 'All Men Are Liars'.

[34] Jane J.Mansbridge, *Beyond Adversary Democracy* (New York: Basic Books, 1980).

[35] R. Dawes, J. McTavish, and H. Shaklee, 'Behavior, Communications, and Assumptions About Other Peoples' Behavior in a Commons Dilemma Situation', *Journal of Personality and Social Psychology*, 35 (1977), 1-11; John M. Orbell, Alphonse J.C. van de Kragt, and Robyn M. Dawes, 'Explaining Discussion-Induced Cooperation in Social Dilemmas', *Journal of Personality and Social Psychology,* 54 (1988), 811-19.

[36] John M. Orbell, Robyn M. Dawes, and Alphonse J.C. van de Kragt, 'The Limits of Multilateral Promising', *Ethics*, 100 (1990), 616-27.

[37] It might be objected that the trust mechanism identified in laboratory experiments works only in situations where the stakes are low. However, there is some evidence to suggest that a trust mechanism inducing cooperative behaviour may be effective even in high-stakes situations. In a study of the rescue of jews in Nazi Europe, Varese and Yaish showed that, after being personally addressed by Jews and asked for help, most people helped. See Federico Varese and Meir Yaish, 'The Importance of Being Asked: The Rescue of Jews in Nazi Europe', *Rationality and Society* 12 (2000), 307-34.

[38] Philip Pettit, 'The Cunning of Trust', *Philosophy and Public Affairs,* 24 (1995), 202-25.

[39] Amos Tversky and Daniel Kahnemann, 'The Framing of Decisions and the Psychology of Choice', *Science,* 211 (1981), 453-58.

[40] I. Levin, S. Schneider, and G. Gaeth, 'All Frames are not Created Equal: A Typology and Critical Analysis of Framing Effects', *Organizational Behavior and Human Decision Processes,* 76 (1998), 149-88.

[41] J. Eiser and K.-K. Bhavnani, 'The Effect of Situational Meaning on the Behavior of Subjects in the Prisoner's Dilemma Game', *European Journal of Social Psychology,* 4 (1974), 93–97.

[42] Michael Bacharach, ''We' Equilibria: A Variable Frame Theory of Cooperation', Working paper, Institute of Economics and Statistics, University of Oxford, 1997; Michael Bacharach and Michele

Bernasconi, 'The Variable Frame Theory of Focal Points: An Experimental Study', *Games and Economic Behavior*, 19 (1997), 1-45.

[43] M.B. Brewer and W. Gardner, 'Who is this 'We'? Levels of Collective Identity and Self Representation', *Journal of Personality and Social Psychology,* 71 (1996), 83–93.

[44] Duncan Black, 'On the Rationale of Group Decision Making', *Journal of Political Economy,* 56 (1948), 23-34.

[45] This requirement is a technicality; if n is even, there may be (harmless) ties, which do not undermine the basic significance of the result; see Riker, *Liberalism Against Populism*, section 5.C.

[46] Black, 'On the Rationale'; Duncan Black, *The Theory of Committees and Elections* (Cambridge: Cambridge University Press, 1958); Amartya Sen, 'A Possibility Theorem on Majority Decisions', *Econometrica,* 34 (1966), 491-99.

[47] Black, 'On the Rationale'; Black, *The Theory of Committees;* Arrow, *Social Choice and Individual Values;* Sen, 'A Possibility Theorem'.

[48] R_i is *strict* if $x \neq y$ implies that not both xR_iy and yR_ix. If $\{R_i\}_{i \in N}$ is a profile of *strict* orderings, then the Condorcet ordering R is strict too; see Riker, *Liberalism Against Populism*, section 5.C.

[49] This result is implied by the Satterthwaite correspondence theorem; a proof can be obtained from the authors on request.

[50] Sen, 'A Possibility Theorem'. An individual is *concerned* over a triple of alternatives if these alternatives are not all tied in this individual's preference ordering. A profile $\{R_i\}_{i \in N}$ of personal preference orderings is *value-restricted* if, for every triple of distinct alternatives x, y, z, the number of concerned individuals is odd, and there exist $w \in \{x, y, z\}$ and $r \in \{1, 2, 3\}$ such that no concerned individual ranks w as his or her r-th preference among x, y, z (in cases of indifference, an alternative can have more than one rank r). D_S is a proper subset of D_V. Other structure conditions that are sufficient for avoiding the impossibility results are *single-cavedness* and *separability into two groups* (see K. Inada, 'A Note on the Simple Majority Decision Rule', *Econometrica,* 32 (1964), 525-31), and *latin-square-lessness* (see B. Ward, 'Majority Voting and Alternative Forms of Public Enterprises', in J. Margolis, ed, *The Public Economy of Urban Communities* (Baltimore: Johns Hopkins University Press, 1965). See Sen, 'A Possibility Theorem' for a discussion of the logical relations between these conditions. For simplicity, we shall here consider only single-peakedness, bearing in mind that the present arguments could easily be restated in terms of the *less demanding* condition of value-restriction.

[51] Richard G. Niemi, 'Majority Decision-Making with Partial Unidimensionality', *American Political Science Review,* 63 (1969), 488-97; see also G. Tullock and C.D. Campbell, 'Computer Simulation of a Small Voting System', *Economics Journal,* 80 (1970), 97-104.

[52] Christian List, 'Two Concepts of Agreement', *The Good Society,* 11 (2002): 72-79.

[53] Single-peakedness, being only a formal structure condition, is logically less demanding than agreement at a meta-level. A profile of personal preference orderings may formally satisfy single-peakedness even when the individuals do not *semantically* conceptualize the alternatives in terms of the same common issue-dimension. But we will assume that agreement at a meta-level is one of the most plausible and common 'causes' of single-peakedness. For this reason, we will use the terms 'dimension' (in the formal sense) and 'issue-dimension' (in the semantic sense) interchangeably whenever there is no danger of confusion.

[54] We should note an alternative way of using domain restriction to circumvent some of the problems posed by the Arrow and Gibbard-Satterthwaite results. If it is acceptable to restrict expression of individual preferences to binary choices between approval or disapproval, approval voting produces meaningful social choices. Under approval voting each invidual casts a vote for all options regarded as acceptable. The winning alternative is that receiving the greatest number of approval votes (Steven J. Brams and Peter C. Fishburn, 'Approval Voting', *American Political Science Review,* 72 (1978), 831-47).

[55] The following simple index can be used for quantifying the level of single-peakedness. Let M be the largest subset of N such that the subprofile $\{R_i\}_{i\in M}$ satisfies single-peakedness. Then define the index of single-peakedness to be the size of M divided by the size of N – formally $|M|/|N|$. The index will take values in the interval from 0 to 1, with a value close to 0 representing a low level of single-peakedness and a value of 1 representing full single-peakedness.

[56] Dryzek, *Discursive Democracy*, pp. 54-5; the term "generalizable interest" comes from Habermas.

[57] Robert E. Goodin, 'Enfranchising the Earth, and its Alternatives', *Political Studies,* 44 (1996), 835-849, pp. 846-7.

[58] This is acknowledged by rational choice theorists: Riker, *Liberalism Against Populism,* p. 128 argues that "If, by reason of discussion, debate, civic education, and political socialization, voters have a common view of the political dimension (as evidenced by single-peakedness), then a transitive outcome is guaranteed." Dennis Mueller, *Public Choice II* (Cambridge: Cambridge University Press, 1989), p. 67 argues that for unidimensional issues "multipeaked preferences ... might be sufficiently unlikely so that cycling would not be much of a problem."

[59] Even less demandingly, value-restriction requires only that, for each triple of alternatives, all individuals agree that *one* of the three alternatives is not best, not medium, or not worst.

[60] Riker, *Liberalism Against Populism*, p. 205.

[61] Elster, 'Introduction', p. 12.

[62] Robert E. Goodin, *Motivating Political Morality* (Oxford: Basil Blackwell, 1992), ch. 7.

[63] James D. Fearon, 'Deliberation as Discussion', in Jon Elster, ed, *Deliberative Democracy* (New York: Cambridge University Press, 1998), pp. 44-68 at p. 54.

[64] Michael Welsh, 'Toward a Theory of Discursive Environmental Policy Making: The Case of Public Range Management', Unpublished PhD dissertation, University of Oregon, 2000.

[65] Geoffrey Brennan and Philip Pettit, 'Unveiling the Vote', *British Journal of Political Science,* 20 (1990), 311-33.

[66] Mueller, *Public Choice II*, pp. 89-90.

[67] Formally, this specification could be a weighting function $\theta_i : \{1, 2, \ldots, k\} \rightarrow \mathbf{R}$ or, less demandingly, a partial ordering Θ_i on $\{1, 2, \ldots, k\}$ (where $\{1, 2, \ldots, k\}$ is the set of issue-dimensions), thus allowing the expression of incommensurabilities.

[68] We use the term *reducible non-single-peakedness* if a non-single-peaked profile can be disaggregated into a set of single-peaked dimension-specific profiles in accordance with conditions (i), (ii) and (iii). Logically, practically any situation might seem to fit this category, particularly if we introduce as many issue-dimensions as there are individuals in *N*. However, we will use the term reducible non-single-peakedness only if an additional semantic condition is met: the identified issue-dimensions are required to have some publicly shared meaning.

[69] Christian List, 'Intradimensional Single-Peakedness and the Multidimensional Arrow Problem', *Theory and Decision*, 2002, forthcoming.

[70] Conditions (P) and (I) can be restated for the multidimensional framework as follows:

Weak Pareto Principle (P). If, for all individuals i and all dimensions j, $xP_{ij}y$, then xPy.

Independence of Irrelevant Alternatives (I). The position of x relative to y in the social ordering R depends exclusively on the position of x relative to y in each of the dimension-specific personal preference orderings in $<\{R_{i1}\}_{i\in N}, \{R_{i2}\}_{i\in N}, \ldots, \{R_{ik}\}_{i\in N}>$.

[71] See Amartya Sen, *On Economic Inequality*, enlarged edition (Oxford: Oxford University Press, 1997), A.7.3 for a related discussion on how an assignment of weights can solve problems of aggregation across multiple dimensions.

[72] Richard H. Pildes and Elizabeth Anderson, 'Slinging Arrows at Democracy: Social Choice Theory, Value Pluralism, and Democratic Politics', *Columbia Law Review,* 90 (1990), 2120-214, at p. 2191.

[73] Sen, *Collective Choice and Social Welfare*.

[74] Our discussion of multidimensional unpacking may seem similar to the dimensional median solution associated with Shepsle's concept of "structure-induced equilibrium". Kenneth A. Shepsle, 'Institutional Arrangements and Equilibrium in Multidimensional Voting Models', *American Journal of Political Science,* 23 (1979), 27-60. Shepsle observes that legislatures often disaggregate complex issues into multiple dimensions, installing constraints such as (i) a division-of-labor arrangement or *committee system*, (ii) a specialization-of-labor arrangement or *jurisdictional system*, and (iii) a monitoring arrangement or *amendment control rule*. These constraints restrict strategizing on the floor of the house such that a stable winning outcome can be produced, whose spatial position is the median on each dimension. Shepsle's critics argue that, in a multidimensional issue-space, this issue-by-issue

median position (like all positions in the issue-space) is not *in general* stable in cross-dimensional coalition formation, even if individuals' preferences are single-peaked in each separate dimension (this follows from spatial voting theory; see Richard McKelvey, 'General Conditions for Global Intransitivities in Formal Voting Models', *Econometrica,* 47 (1979), 1085-111; Norman Schofield, 'Instability of Simple Dynamic Games', *Review of Economic Studies,* 45 (1976), 575-94; Norman Schofield, *Social Choice and Democracy* (Berlin: Springer, 1986)). Shepsle is also criticized for assuming the dimensional separability of individuals' preferences. However, unlike the spatial voting framework, our framework does not identify individuals or policy alternatives with positions in a (multidimensional) issue-space; our formalism distinguishes between policy alternatives and their perceived issue-spatial positions. Individual preferences, as well as the perceived spatial position of alternatives, may change during deliberation. Secondly, we do not *in general* single out one specific, mechanically computable, alternative (or position, in Shepsle's terms), such as the dimension-by-dimension median, as the ideal outcome. In our framework, it is not *in general* the case that all positions in a multidimensional issue-space can be filled with potential options, and there may not exist an option in *X* which is the dimension-by-dimension median. The parallels between Shepsle's proposal and ours are probably closest in the case of *subdividing the decision*, which allows combining the winning outcomes in different dimensions to form an overall policy package. However, we acknowledge that often (i) the requisite separability or divisibility may not be given, (ii) people might disagree, and thus have to deliberate on, the relative weight to assign to different dimensions, so as to reach agreement on a *lexicographic hierarchy of dimensions* or on cross-dimensional trade-offs (as exemplified by *logrolling*), and (iii) the role of multidimensional unpacking is sometimes simply to analyse the precise nature of the impasse.

[75] Adolf Gundersen, *The Environmental Promise of Democratic Deliberation* (Madison, WI: University of Wisconsin Press, 1995).

[76] David Pelletier, Vivica Kraak, Christine McCallum, Ulla Uustalo, and Robert Rich, 'The Shaping of Collective Values Through Deliberative Democracy: An Empirical Study from New York's North Country', *Policy Sciences,* 32 (1999), 103-31.

[77] Pelletier et al, 'Shaping of Collective Values', p. 117.

[78] Christian List, Iain Mclean, James Fishkin, and Robert Luskin, 'Can Deliberation Induce Greater Preference Structuration?' Paper presented at the Annual Meeting of the American Political Science Association, 2000.

[79] Ian Budge, David Robertson, and Derek Hearl, eds, *Ideology, Strategy and Party Change* (Cambridge: Cambridge University Press, 1987).

[80] P. Gärdenfors, 'Positional Voting Functions', *Theory and Decision,* 4 (1973), 1–24.

[81] J.C. de Borda, 'Mémoire sur les élections au scrutin', in J.-C. de Borda, *Histoire de l'Académie Royal des Sciences*. Paris, 1781; H.P. Young, 'An Axiomatization of Borda's Rule', *Journal of Economic Theory,* 9 (1974), 43–52.

[82] Charles Plott and Michael Levine, 'A Model of Agenda Influence on Committee Decisions', *American Economic Review,* 68 (1978), 146-60.

[83] Riker, *Liberalism Against Populism*, p. 79.

[84] Sen, *Collective Choice and Social Welfare.*

[85] Arrow's conditions can be restated for SWFLs as follows: throughout conditions (U), (P), (I) and (D), simply substitute "personal welfare function" for "personal preference ordering", "$W_i(x)>W_i(y)$" for "xP_iy", "$\{W_i\}_{i\in N}$" for "$\{R_i\}_{i\in N}$"; further, in condition (I), substitute "on $W_i(x)$ and $W_i(y)$ in each of the personal welfare functions in $\{W_i\}_{i\in N}$" for "on the position of x relative to y in each of the personal preference orderings in $\{R_i\}_{i\in N}$".

[86] Sen, *Collective Choice and Social Welfare*; Sen, *Choice, Welfare and Measurement*, ch. 11.

[87] **(OLC).** For any $\{W_i\}_{i\in N}$ and $\{W^*_i\}_{i\in N}$ in the domain of F, $F(\{W_i\}_{i\in N}) = F(\{W^*_i\}_{i\in N})$ if $W^*_i = \phi(W_i)$ for some positive monotonic transformation $\phi : \mathbf{R} \to \mathbf{R}$.

(CUC). For any $\{W_i\}_{i\in N}$ and $\{W^*_i\}_{i\in N}$ in the domain of F, $F(\{W_i\}_{i\in N}) = F(\{W^*_i\}_{i\in N})$ if $W^*_i = a_i+b*W_i$, where $a_1, a_2, \ldots, a_n, b$ are real numbers and $b > 0$.

[88] For example, Sen, *Collective Choice and Social Welfare*; Sen, *Choice, Welfare and Measurement*; W. Bossert and J.A. Weymark, 'Utility in Social Choice', in S. Barberà, P.J. Hammond, and C. Seidel, eds, *Handbook of Utility Theory*, Vol. 2 (Boston: Kluwer, Boston, 1996); Christian List, 'Introducing a 'Zero-Line' of Welfare as an Escape-Route from Arrow's Theorem', *Pacific Economic Review,* 6 (2001) (special section in honour of Amartya Sen), 223-38.

[89] Sen, *Collective Choice and Social Welfare*.

[90] Given $\{R_i\}_{i\in N}$, the *leximin rule* defines a social ordering R as follows. For each alternative x in X, first define a permutation $i \mapsto [i]$ of N (depending on x) such that $W_{[1]}(x) \le W_{[2]}(x) \le \ldots \le W_{[n]}(x)$. For every x and y in X,

xPy if and only if $\quad W_{[i]}(x) > W_{[i]}(y)$ for some $i\in N$,

and $W_{[j]}(x) = W_{[j]}(y)$ for all $j<i$.

Given $\{R_i\}_{i\in N}$, the *utilitarian rule* defines a social ordering R as follows: for every x and y in X,

xRy if and only if $\quad \Sigma_{i\in N}W_i(x) \ge \Sigma_{i\in N}W_i(y)$.

[91] John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1973).

[92] Sen, *Choice, Welfare and Measurement*, ch. 11.

[93] Sen, *Choice, Welfare and Measurement*, ch. 11.

[94] Lionel Robbins, *An Essay on the Nature and Significance of Economic Science* (London: Macmillan, 1932), p. 140.

[95] Rawls, *A Theory of Justice*; John Rawls, 'Social Unity and Primary Goods', in Amartya Sen and Bernard Williams, eds, *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982).

[96] Amartya Sen, *Inequality Reexamined* (Oxford: Oxford University Press, 1992).

[97] Rawls, *A Theory of Justice*, section 23.

[98] Gutmann and Thompson, *Democracy and Disagreement*.

[99] Amy Gutmann and Dennis Thompson, 'Deliberative Democracy Beyond Process', paper presented to the Conference on Deliberating About Deliberative Democracy, Austin, TX, 4-6 February 2000.

[100] Bernd Marin, 'Austria: the Paradigm Case of Liberal Corporatism?', in Wyn Grant, ed, *The Political Economy of Corporatism* (New York: St. Martin's, 1985), pp. 89-125 at p. 116.

[101] Katri Kosonen, 'Savings and Economic Growth from a Nordic Perspective', in Jukka Pekkarin, Matti Pohjola, and Bob Rowthorn, eds, *Social Corporatism: A Superior Economic System?* (Oxford: Oxford University Press, 1992), pp. 178-209.

[102] Gutmann and Thompson, 'Deliberative Democracy Beyond Process'.