Democratic Deliberation and Social Choice: A Review

Christian List[1]

This version: June 2017

Forthcoming in the *Oxford Handbook of Deliberative Democracy*

1. Introduction

In normative political theory, it is widely accepted that democratic decision making cannot be reduced to voting alone, but that it requires reasoned and well-informed discussion by those involved in and/or subject to the decisions in question, under conditions of equality and respect. In short, democracy requires *deliberation* (e.g., Cohen 1989; Gutmann and Thompson 1996; Dryzek 2000; Fishkin 2009; Mansbridge et al. 2010). In formal political theory, by contrast, the study of democracy has focused less on deliberation, and more on the *aggregation* of individual preferences or opinions into collective decisions – *social choices* – typically through voting (e.g., Arrow 1951/1963; Riker 1982; Austen-Smith and Banks 2000, 2005; Mueller 2003). While the literature on deliberation has an optimistic flavour, the literature on social choice is more mixed. It is centred around several paradoxes and impossibility results showing that collective decision making cannot generally satisfy certain plausible desiderata. Any democratic aggregation rule that we use in practice seems, at best, a compromise.

Initially, the two literatures were largely disconnected from each other. Since the 1990s, however, there has been a growing dialogue between them (e.g., Miller 1992; Knight and Johnson 1994; van Mill 1996; Dryzek and List 2003; Landa and Meirowitz 2009). This chapter reviews the connections between the two. Deliberative democratic theory is relevant to social choice theory in that deliberation can complement aggregation and open up an escape route from some of its negative results. Social choice theory is relevant to deliberative democratic theory in that its formal models can shed light on some aspects of deliberation, such as the nature of deliberation-induced opinion change.

Sections 2 and 3 introduce the notions of social choice and deliberation. Section 4 discusses several hypotheses on the effects of deliberation on preferences and assesses their implications for social choice. Section 5 reviews some social-choice-theoretic models of deliberation and considers the mechanisms of deliberation-induced opinion change. Section 6 addresses deliberation from the perspective of judgment-aggregation theory, the branch of social choice theory that focuses on the aggregation of judgments rather than preferences.

2. The problem of social choice

Collective decision making is a key feature of social organization, in bodies such as the electorate, legislatures, committees, courts, juries, expert panels, companies, and other organizations. In social choice theory, we model collective decision making as the aggregation of individual inputs, such as votes or preferences, into collective outputs, such as collective decisions or collective preferences. In fact, *social choice theory* can be defined as the study of aggregation. It must not be confused with *rational choice theory*, the study of individually rational behaviour and its collective consequences. The study of aggregation need not be

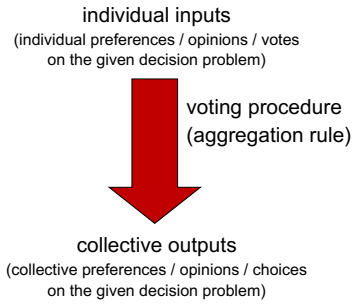
[1] I am grateful to Andre Bächtiger, John Dryzek, William Gehrlein, Dimitri Landa, Jane Mansbridge, Ashley Piggins, Daniele Porello, and Laura Valentini for helpful comments. My work was supported by a Leverhulme Major Research Fellowship.

committed to the behavioural assumptions of rational choice theory; in particular, it need not be committed to any "*homo economicus*" or "self-interest" model of decision making.

2.1 The basics

The central concept of social choice theory is that of an *aggregation rule*. This is an input-output scheme – a *function* – which takes as input the votes or preferences across the members of some group and delivers as output a collective decision or collective preference, as illustrated in Figure 1.

Figure 1: The aggregative model



For instance, a group of n individuals might have to make a choice between two options, x and y, such as two candidates, the acceptance or rejection of some proposal, or the guilt or innocence of a defendant. A combination of votes across the group is called a *profile*. It is a list of xs and ys whose i^{th} entry stands for the vote of the i^{th} individual. In a three-member group, the profile $\langle x, x, y\rangle$ represents a situation in which the first two individuals vote for x and the third votes for y. An *aggregation rule* assigns to each profile a collective decision, which could be either x, or y, or (optionally) a tie. The best-known example is the *majority rule*. Here, for each profile, the output is the option (x or y) that is supported by the most votes; it is a tie if votes are equally split. Another example is a *supermajority rule*. Here, a supermajority (such as 2/3 or 3/4 or perhaps everyone) must vote for an option in order for that option to be chosen. A third, undemocratic example is a *dictatorship of one individual*, where the output always tracks the vote of a fixed individual, regardless of others' votes.

Social choice theory is concerned not just with specific examples of aggregation rules, but with the logical space of all possible aggregation rules. A typical research question is which rules, if any, have certain desirable properties (for a survey, see List 2013).

2.2 Some arguments for the majority rule

The majority rule has long been seen as the default democratic aggregation rule. Social choice theory offers at least three formal arguments for it, in decisions between two options (cf. List 2013, sec. 2). The first is procedural. It invokes *May's theorem* (1952): the majority rule is the only aggregation rule that satisfies four procedural desiderata in a two-option choice. These are: *universal domain*, which requires that any possible profile of individual votes be admissible as input; *anonymity*, which requires equal treatment of all voters; *neutrality*, which requires equal treatment of the two options; and *positive responsiveness*, which requires the

output to be a positive function of the individual votes.[2] If we consider these desiderata indispensable, then we have a strong reason to use the majority rule.

The second argument is outcome-based. It appeals to the epistemic ("truth-tracking") qualities of the majority rule and applies when there is an independent fact of the matter as to which option is "correct". In a criminal trial, for instance, there is a fact as to whether the defendant is guilty or not. In such cases, *Condorcet's jury theorem* shows that if all voters have an equal but independent chance better than random of voting for the correct option, then the majority decision is more likely to be correct than each individual's vote, and the probability of a correct majority decision converges to 1 as the number of voters increases (e.g., Grofman, Owen and Feld 1983; List and Goodin 2001). In practice, the theorem's assumptions – the independence of voters and their better-than-random reliability – are hard to achieve, but at least in favourable conditions the majority rule seems good at reaching correct decisions.

The third argument is also outcome-based, but in a utilitarian rather than epistemic way. Suppose each voter gets a utility of 1 if the outcome of the collective decision, again among two options, matches his or her preference, and a utility of 0 otherwise. Then the majority rule selects the utility-maximizing option, a fact known as the *Rae-Taylor theorem* (e.g., Mueller 2003). More generally, the majority rule maximizes the number of voters whose preferences are satisfied. However, this utilitarian argument for the majority rule only applies if all voters have an equal stake in the decision.[3]

2.3 The paradoxes and impossibility results of social choice theory

Condorcet, who had recognized the virtues of the majority rule in two-option choices, also saw what can go wrong when there are more options: an insight known as *Condorcet's paradox* (e.g., Gehrlein 1983, 2006). Suppose there are three options, x, y, and z; a third of the voters prefer x to y to z; a second third prefer y to z to x; and the last third prefer z to x to y. Then there are majorities, of two thirds each, for x over y, for y over z, and for z over x: a *preference cycle*. Each option is beaten by another option in a majority contest. In such a situation, there is no *Condorcet-winning option*: an option that beats, or ties with, every other option in a pairwise majority vote. Furthermore, the majority preferences are *intransitive*: although x is majority-preferred to y, and y is majority-preferred to z, x is not majority-preferred to z. If an individual had those preferences, he or she would be considered irrational: if I prefer apples to bananas, and bananas to coconuts, then I am rationally required to prefer apples to coconuts as well. Condorcet's paradox challenges the notion of a majoritarian "will of people". The "collective will", if defined in a majoritarian way, may violate standard requirements of rationality, such as transitivity.

Condorcet's paradox might be considered an isolated artefact. However, a series of influential results in social choice theory, of which Arrow's impossibility theorem (1951/1963) is the most important, suggests that Condorcet's paradox is just the tip of the iceberg of a larger problem. To explain Arrow's theorem, suppose that, as before, there are n individuals, and each submits a preference ordering over the options (x, y, z, etc.), ranking them from most to least preferred. Let P_1, P_2, …, P_n denote these n individual preference orderings.[4] A *preference aggregation*

[2] [Online only] Any changes in votes for a winning option do not hurt that option, and any changes in votes for one of the options in the event of a tie break the tie in the direction of the change.
[3] On unequal stakes, see Brighouse and Fleurbaey (2010).
[4] Each P_i is a transitive, asymmetric, and connected binary relation (a *strict* ordering). Arrow employed *weak* orderings (permitting ties), but for simplicity I focus on strict orderings throughout this chapter.

rule assigns to each such list, or *profile*, of preference orderings a resulting collective preference ordering, *P*, which ranks the options from collectively most to least preferred.

Arrow suggested that an acceptable aggregation rule should satisfy at least five desiderata. *Universal domain*: any possible profile of individual preference orderings is admissible as input. *Collective rationality*: the output is a well-defined preference ordering; in particular, there are no cycles as in Condorcet's paradox. The *Pareto principle*: if everyone prefers *x* to *y*, then *x* is collectively preferred to *y*. *Independence of irrelevant alternatives*: the collective preference between any pair of options, *x* and *y*, depends only on the individual preferences between *x* and *y*, not on the individual preferences with respect to other options. *Non-dictatorship*: there is no fixed individual who always determines the collective preference. Arrow's theorem shows that the five desiderata are mutually inconsistent when there are more than two options: there exists no preference aggregation rule satisfying all of them. In the special case of two options, the majority rule satisfies all five.

In practice, then, any aggregation rule must violate at least one desideratum, and this comes at a cost. If an aggregation rule violates universal domain, it does not cope with all possible profiles of preferences and is therefore not robust to pluralism. If it violates collective rationality, it fails to rank the options in a complete order or generates preference cycles, as in Condorcet's paradox. If it violates the Pareto principle, it sometimes overrules unanimous preferences. If it violates independence of irrelevant alternatives, it has at least two potential defects. The first is a lack of transparency: the collective preference between two options, *x* and *y*, may change from one profile to another as a result of changes in preferences with respect to other options, even when everyone's preferences between *x* and *y* remain unchanged. The second defect is vulnerability to strategic voting: a failure of "strategy-proofness". Individuals may have opportunities to manipulate the outcome by voting strategically – a point established more precisely by another classic result: the Gibbard-Satterthwaite theorem (Gibbard 1973; Satterthwaite 1975). Finally, if the aggregation rule violates non-dictatorship, it is outright undemocratic.

Riker (1982) interpreted the social-choice-theoretic impossibility results as challenging the coherence of any form of democracy that relies on the notion of the "will of the people", where this is an aggregate of "individual wills" (cf. Coleman and Ferejohn 1986; Cohen 1986; Knight and Johnson 1994; McGann 2006).[5] Over the years, Riker's negative interpretation has become less widely accepted, and social choice theorists, including prominently Sen (e.g., 1998), have devoted much energy to finding escape routes from Arrow's theorem and related results. They have done so, on the one hand, by showing that there are reasons for relaxing some of Arrow's desiderata in certain contexts, and on the other hand, by reformulating the problem of social choice, permitting for instance inputs that go beyond rankings of the options. Richer inputs may take the form of cardinal utility assignments, ratings, or judgments (cf. List 2013, sec. 4-5; Balinski and Laraki 2011).

Still, one lesson is hard to contest: there does not exist a single, universally best aggregation rule. Choosing an aggregation rule requires trade-offs between different desiderata, and different solutions to those trade-offs are appropriate in different contexts.

[5] [Online only] As Coleman and Ferejohn (1986) and Knight and Johnson (1994) note, one can distinguish between the "instability" of social choices, due to majority cycling (as in Condorcet's paradox), and its "ambiguity", due to the procedure-dependence of social choices, where every procedure has certain defects (by Arrow's theorem). For an argument to the effect that the instability associated with majority cycling need not be bad for democracy, see Miller (1983).

3. Democratic deliberation

In contrast to aggregation, which is the merging of different people's opinions into a single collective output, deliberation, as noted, is the reasoned and well-informed discussion of these opinions by the people involved, under conditions of equality and respect. There are several definitions of deliberation in the literature, which differ, among other things, in whether they define deliberation as a procedure or as a behaviour, and in how idealized they are.

3.1 The procedure-behaviour distinction

Just as we distinguish between voting *procedures* and voting *behaviour*, so we can distinguish between deliberative procedures and deliberative behaviour (Landa and Meirowitz 2009).[6] *Deliberative procedures* are settings in which deliberation can take place. If we give voters an opportunity to talk before voting, this is an instance of a deliberative procedure: "first talk, then vote" (e.g., Goodin 2008, ch. 6). *Deliberative behaviours* are the ways in which people actually deliberate: how they treat each other when they communicate, what they say, whether they are truthful or manipulative, whether they change their opinions, and so on.

We can now define deliberation *either* in terms of the procedural setup, for instance as a communication procedure with equal and fair participation, *or* in terms of the behaviour that takes place, for instance as reasoned, well-informed, and respectful speech. Clearly, there is a connection between procedures and behaviours. Different procedures may lead to different behaviours. The way people deliberate may be affected, for example, by the presence of a moderator or by constraints on the timing and duration of each participant's speech. Once we have criteria for the behaviours that count as "deliberative", we can ask which procedures facilitate or promote those behaviours (cf. Landa and Meirowitz 2009).

3.2 The idealism-realism distinction

Some definitions of deliberation are more idealized (so that "deliberation" becomes harder to achieve in practice, while perhaps serving as an aspirational ideal), others more realistic (so that "deliberation" is a more common phenomenon). A definition that requires deliberation to generate an "ideal speech situation", a setting of power-free discourse (a notion associated with Habermas), would fall on the idealized side of the spectrum (cf. Bohman and Rehg 2014), as would a definition of deliberation as communication that is fully informed, rational, truthful, oriented towards the common good, egalitarian and respectful, and based on public reasons. By contrast, a definition of deliberation simply as pre-vote communication, which need not exclude negotiation and self-interest, would be more realistic (cf. Mansbridge et al. 2010).[7]

3.3 A working definition

I will here adopt a procedural definition that is relatively realistic. I will define *deliberation* as a communicative procedure, typically in the run-up to a collective decision, which is designed to promote substantive, balanced, and civil discussion. "Substantive" means that it focuses on the options and the reasons for preferring or dispreferring them (this may include narratives

---

[6] Landa and Meirowitz distinguish between *deliberative environments* and *deliberative behaviours*.

[7] [Online only] The idealism-realism distinction has to do with how empirically feasible and demanding the instantiation of deliberation is, not with its value or desirability. The question of how valuable or desirable deliberation is must be distinguished from the question of how empirically feasible and demanding it is. There can be realistic definitions under which deliberation is highly desirable. Furthermore, the value and desirability of deliberation are not settled by its definition, but depend on our background theory of political morality.

and the sharing of experiences); "balanced" means that it involves different perspectives, arguments, and views; and "civil" means that it is respectful (Fishkin and Luskin 2005; List, Luskin, Fishkin, and McLean 2013; Fishkin 2009).

We can then explore how deliberation, so defined, might relate to aggregation. According to a "pure deliberative model", deliberation replaces aggregation, and leads directly to a collective decision: a consensus or a compromise. This stands in contrast to a "mixed model", according to which deliberation complements aggregation. Here, people first deliberate and then vote. Figures 2 and 3 illustrate those two models. The comparative baseline is the "pure aggregative model", as shown in the earlier Figure 1.

Figure 2: The pure deliberative model

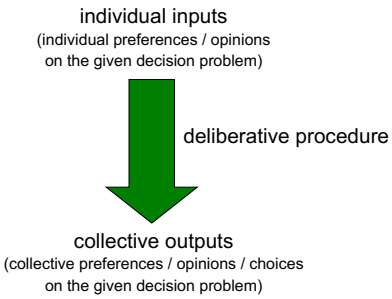
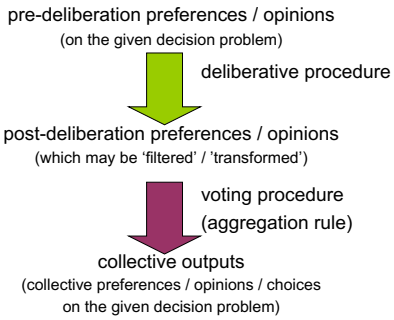


Figure 3: The mixed model



4. The effects of deliberation on individual preferences

To assess the different models of decision making, it is helpful to compare three hypotheses about the effects of deliberation on preferences. I begin with a very optimistic hypothesis.

4.1 The consensus hypothesis: Deliberation tends to generate a consensus

If true, this would make the pure deliberative model tenable. Elster (1986, p. 112) summarizes the idea as follows: "rather than aggregating or filtering preferences, the political system should be set up with a view to changing them by public debate and confrontation ... [T]here would not be any need for an aggregation mechanism, since a rational discussion would tend to produce unanimous preferences."

As most theorists recognize, however, we cannot rely on deliberation to generate a consensus. Very few real-world organizations manage to make pure consensus decisions, the Quakers being one often-cited example. Indeed, if we were to *define* deliberation as a form of consensus-generating discussion, real-world instances of deliberation would be rare. Arguably, a minimal constraint on a realistic definition of deliberation is that it is a contingent matter whether a consensus emerges, not a definitional matter. Since the consensus hypothesis is not generally true under the definition of deliberation I am using, the pure deliberative model is not realistic.

4.2 The no-helpful-change hypothesis: Deliberation does not helpfully reduce preference diversity

On a strong version of this hypothesis, people's opinions on many issues are too entrenched to be open to change in a time-limited deliberative process. Mackie (2006, p. 279) describes this as follows: "[P]ublic deliberation on a pending item seldom seems to change anyone's mind…

[D]ue to the network [structure of individual opinions], the effects of deliberative persuasion are typically latent, indirect, delayed, or disguised."

While the consensus hypothesis was too optimistic, the "no-change" hypothesis is too pessimistic. Experience suggests that deliberation sometimes changes people's minds, and there is some social-scientific evidence that deliberation promotes reflection and learning, and generates more considered preferences (e.g., Luskin, Fishkin, and Jowell 2002; Barabas 2004; Farrar et al. 2010). In particular, the experimental design of Deliberative Polling, developed by Fishkin and colleagues, allows us to compare people's opinions before and after deliberation. In a Deliberative Poll, between 130 to 350 randomly chosen members of the public are first interviewed on some policy issue; they then receive carefully balanced briefing materials and participate in a weekend of group deliberation; finally, they are interviewed again, using the same questions as before.[8] The results suggest that deliberation tends to change opinions and to make participants better informed, but also that it does not normally generate unanimity.

Even if deliberation changes opinions, this may not be enough to solve the problem of social choice. Post-deliberation preferences still need to be aggregated. As van Mill (1996, 2006) points out, if the deliberative procedure is relatively open, it might lead to post-deliberation preferences to which the social-choice-theoretic paradoxes and impossibility results continue to apply. The challenge, he says, is "to discover specific rules [of deliberation] that create stability [in the social-choice-theoretic sense] without, at the same time, completely undermining the freedom and equality so necessary for the legitimacy of outcomes. Too much participation and we do not get stability; too little and we end up with an overly constrained system" (1996, p. 749). If this challenge remains unresolved, we cannot conclude that deliberation facilitates social choice.

Interestingly, deliberation might interact with the mechanism underlying Condorcet's jury theorem. While not generating a consensus, deliberation might increase the reliability of individual opinions in cases where there is an independent fact as to which option is "correct". At the same time, it might generate interdependencies between different individuals' opinions. In short, it might helpfully affect voter reliability, but adversely affect voter independence. For a discussion of this tradeoff, see Dietrich and Spiekermann (2013). Let me now turn to a third hypothesis about deliberation's effects.

4.3 The meta-consensus hypothesis: Deliberation tends to generate a meta-consensus, which is associated with "single-peaked" preferences

To explain this, first note that, although a full consensus – a situation of unanimous preferences – is sufficient for avoiding Condorcet's paradox and Arrow's theorem,[9] it is not necessary. A certain amount of structure in the occurring profiles of individual preferences is sufficient. Black (1948) identified a relevant structure condition, called "single-peakedness".

A profile of preference orderings across a group of individuals is *single-peaked* if the options can be aligned from left to right such that each individual has an ideal point somewhere on this left-right axis, with decreasing preference for options as they get more distant, in either direction, from the ideal (cf. Arrow 1951/1963, ch. 7). For instance, a voter might most prefer a "leftist" option and prefer other options less as they are further to the right. Another voter might most prefer a "centrist" option and disprefer more extreme options. The left-right axis

[8] In some polls, there are also non-deliberating control groups.
[9] Technically, neither result applies to the domain of unanimous preferences.

relative to which preferences have this structure is called a *structuring dimension*. "Left" and "right" are geometrical notions here; they could have various meanings: "progressive" and "conservative", "urban" and "rural", "secular" and "religious", and so on. Figure 4 shows a single-peaked profile of three individual preference orderings over five options, *x*, *y*, *z*, *v*, *w* (left-right aligned in this order), and Figure 5 shows a non-single-peaked preference ordering.[10]

Figure 4: Single-peaked preferences

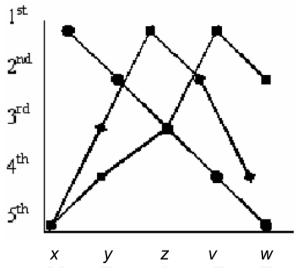
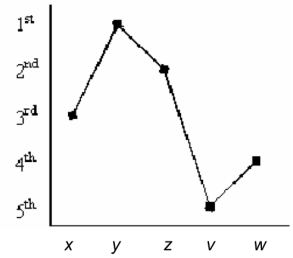


Figure 5: Non-single-peaked preferences



Single-peakedness is a sufficient condition for transitive majority preferences and for the existence of a Condorcet-winning option, namely the option that is most preferred by the median individual with respect to the relevant structuring dimension. This fact is called the *median-voter theorem*.[11]

Furthermore, suppose we replace Arrow's universal-domain desideratum with the requirement that the aggregation rule admit as input only all single-peaked preference profiles. The majority rule then satisfies all of Arrow's other desiderata. A similar point can be made with regard to the Gibbard-Satterthwaite theorem on strategy-proof aggregation. On the domain of single-peaked preference profiles, the majority rule (along with other "median voting rules") is not vulnerable to strategic voting (Moulin 1980). Moreover, as discussed later, even if a high-enough proportion of the individuals have single-peaked preferences, cyclical majority preferences are unlikely (Niemi 1969).

So, *if* deliberation generates single-peaked preferences, *then* it will open up an escape route from the paradoxes and impossibility results of social choice theory. Riker, who had emphasized the social-choice-theoretic challenge for democracy, accepted this conditional claim, writing: "[i]f, 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" (1982, p. 128). But he suggested that the effect would occur only for "issues of minor importance".

Several scholars, however, have defended the "if" part of the conditional and argued that deliberation, at least under favourable conditions, can be expected to produce single-peaked preferences (Miller 1992; Dryzek and List 2003; cf. Knight and Johnson 1994).[12] Why? A possible mechanism involves a deliberation-induced "meta-consensus". Although deliberation is unlikely to generate unanimity – a "substantive consensus" – it may generate an agreement on a common dimension in terms of which to conceptualize the issue and along which

[10] Each line represents one preference ordering: options are on the horizontal axis, their ranks on the vertical one.
[11] This assumes, for simplicity, an odd number of individuals, to rule out majority ties. [Online only] Single-peakedness, under the present Black-Arrow definition, should not be confused with *spatial single-peakedness*, common in spatial voting theory, which presupposes that options are defined as points in some exogenously given issue space. Here, I make no such assumption. The positioning of options could be endogenous. Spatial single-peakedness is sufficient for the avoidance of majority cycles only when the issue space is one-dimensional.
[12] For discussion, see also Aldred (2004) and Dryzek and List (2004).

preferences will become single-peaked: a "meta-consensus". (See List 2002 on "substantive" versus "meta-level" agreement; cf. Dryzek and List 2003.[13]) For example, a group of deliberators may fail to agree on how to rank the options, but come to agree that their disagreement concerns a trade-off between the economy and the environment. The following three-part hypothesis summarizes the proposed mechanism (quoting List 2007):[14]

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

If this mechanism works, it supports the mixed decision model of Figure 3: deliberation followed by aggregation.[15] (For a critical discussion, see Ottonelli and Porello 2013. For a formalization of the mechanism in dynamic logic, see Goldbach 2015. On the relationship between a common semantic issue dimension and single-peakedness, see Porello 2016.)

4.4 Empirical evidence and assessment

Data from Deliberative Polls support the hypothesis just discussed (List, Luskin, Fishkin, and McLean 2013; Farrar et al. 2010). Deliberative Polls enable us to compare the participants' preferences before and after deliberation. The relevant study covered nine polls, discussing thirteen issues, from energy provision in Texas to the future of the monarchy in Australia. The finding was that post-deliberation preferences tended to be closer to single-peaked than pre-deliberation preferences. Specifically, the study measured *proximity to single-peakedness* (as defined by Niemi 1969): the proportion of individuals whose preferences are single-peaked relative to a common structuring dimension.

The increases in proximity to single-peakedness were greater for low-salience issues (such as energy provision), on which people's opinions were presumably less entrenched, than for high-salience issues (such as the monarchy), on which people were presumably more opinionated. The increases were also greater for issues that more readily lend themselves to a unidimensional conceptualization. And the increases were greater for the subsamples of participants who emerged from the deliberation better informed, as judged from their answers to some factual questions.

---

[13] For further notions of meta-consensus not associated with social-choice-theoretic structure conditions, see Dryzek and Niemeyer (2006). For a technical analysis, see Gehrlein, Lepelley, and Moyouwou (2015).

[14] Variants also appeared in List (2002) and Dryzek and List (2004).

[15] A shallower mechanism generating single-peakedness might involve a tendency for deliberators to mimic the preferences of opinion leaders with single-peaked preferences (List, Luskin, Fishkin, and McLean 2013).

It remains an open question whether deliberation would also generate single-peaked preferences on issues beyond those covered in the study, and how scalable the mechanism is, i.e., whether the effect could occur in larger groups or in the electorate as a whole, for instance as a result of a nation-wide "deliberation day", as proposed by Ackerman and Fishkin (2002).

One might also wonder whether the phenomenon of deliberation-induced single-peakedness is consistent with another frequently documented effect of deliberation: group polarization. Deliberating groups sometimes take more extreme positions after deliberation than before (Sunstein 2002). A mildly conservative group might become strongly conservative; a mildly progressive group might become strongly progressive. However, group polarization has been observed especially in homogenous groups, as distinct from the more heterogeneous groups of the Deliberative Polls and other groups that are reasonably representative of the general public. Heterogeneity reduces the risk of creating an echo chamber in which prior opinions are simply reinforced.[16] That said, increases in proximity to single-peakedness are consistent with increases in polarization. Preferences can be both single-peaked and polarized.

Another common objection to the hypothesis of deliberation-induced single-peakedness is that single-peakedness is a very demanding condition, and many issues do not lend themselves to a unidimensional conceptualization (e.g., Aldred 2004). However, other less demanding structure conditions on preferences are still sufficient for avoiding majority cycles, but easier to attain than single-peakedness. An example is "triplewise value restriction" (Sen 1966), which requires that, among every triple of options, one option is never ranked top, or never ranked middle, or never ranked bottom. Unidimensionality is not required for this. A second example is "aggregate ideological consistency" (Feld and Grofman 1988), which secures a *majority* preference ordering that is single-peaked relative to a single dimension while not requiring *individual* preference orderings to be single-peaked relative to the same dimension.

As Sen (1966, p. 498) notes, "[a] comparatively limited measure of agreement seems to be sufficient to guarantee consistent majority decisions". The deliberative-democracy literature has perhaps focused more on single-peakedness than on other structure conditions because it is easier to hypothesize a mechanism by which deliberation might generate single-peakedness (given its interpretability in terms of a common issue dimension) than to hypothesize a similar mechanism for the alternative conditions.

4.5 Deliberation and the probability of majority cycles

Finally, it is worth noting that, in combinatorial terms, the avoidance of cycles via the structuration of preferences is simpler than one might think. Suppose, for the sake of argument, that individual preferences are drawn from a uniform distribution over all possible preference orderings. For example, when there are three options, there are six logically possible strict orderings, and each individual would then have a probability of $^1/_6$ of holding each of them. The resulting probability distribution over preference profiles is called an *impartial culture* (e.g., Gehrlein 1983). It is known that this distribution maximizes the probability of majority cycles (Tsetlin, Regenwetter, and Grofman 2003). Calculations further suggest that, given an impartial culture, the probability of a majority cycle increases, as the number of options

[16] Even in homogenous groups, polarization seems not to occur under sufficiently favourable deliberative conditions (Grönlund, Herne, and Setälä 2015).

increases and also as the number of voters increases (Gehrlein 1983). But one striking point is often overlooked.

Not only do slight increases in proximity to single-peakedness, compared to the baseline of an impartial culture, reduce the probability of cycles (Niemi 1969; cf. Gehrlein, Lepelley, and Moyouwou 2015). But it can also be shown that the slightest systematic deviations from the impartial-culture distribution can have the same effect in a large electorate (List and Goodin 2001, Appendix 3). In the three-option example, if each individual's probability of holding each of the six possible preference orderings is not uniformly $^1/_6$, but a little above $^1/_6$ for some orderings and a little below $^1/_6$ for suitable others, then the probability of a majority cycle converges to zero, as the number of voters increases. Thus the prediction of a sizeable probability of cycles under an impartial culture is a "knife-edge" result.

To be sure, one can also construct deviations from an impartial-culture distribution that do not have this cycle-avoiding effect, but these are arguably more contrived than the deviations that reduce the probability of cycles. Indeed, several recent studies suggest that cycles are empirically rare, and that they should be theoretically expected to be rare, even outside deliberative settings (Mackie 2003; Regenwetter et al. 2006; Gehrlein 2006; Gehrlein and Lepelley 2011). Cycles are probable only when preference diversity is very symmetrical, as it is under an impartial-culture distribution, and suitable structures in the distribution of preferences that break this symmetry can reduce their probability. It is not implausible that deliberation could have such a structuration effect.

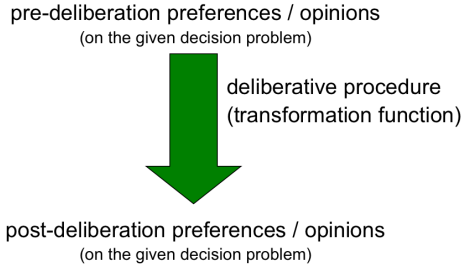
5. Formal models of deliberation

I have discussed the idea that deliberation can change individual opinions. I will now review several social-choice-theoretically inspired models of deliberation and say more about the mechanisms of deliberation-induced opinion change.

5.1 A model of deliberation as preference transformation

I begin with a simple model adapted from List (2011b). In this model, people enter the deliberative procedure with their pre-deliberation preferences and emerge from it with their post-deliberation preferences. As before, each individual's preferences take the form of a ranking of the options from most to least preferred. The model assumes *functionality*: the deliberative procedure is represented by a function which maps each profile of pre-deliberation preference orderings $\langle P_1, P_2, \ldots, P_n \rangle$ across an n-member group to a resulting profile of post-deliberation preference orderings $\langle P^*_1, P^*_2, \ldots, P^*_n \rangle$, as shown in Figure 6. Call such a function a *preference transformation function*. It captures the relationship between pre-deliberation inputs and post-deliberation outputs from a bird's eye perspective.

Figure 6: Deliberation as preference transformation



Here are some baseline desiderata that we might expect a preference transformation function to satisfy. *Universal domain*: any possible profile of individual preference orderings is admissible as input. *Post-deliberation rationality*: the output is a profile of well-defined preference orderings. *Consensus preservation*: in cases of pre-deliberation unanimity (where $P_1 = P_2 = \ldots = P_n$), the unanimous agreement is preserved after deliberation. *Minimal relevance*: the individuals do not always ignore their pre-deliberation preferences in forming their post-deliberation preferences.[17] Finally, *pairwise independence*: the post-deliberation preferences for any pair of options depend only on the pre-deliberation preferences for that pair; this ensures "strategy-proofness" of the deliberative process.

A theorem proved in List (2011b) implies that, when there are more than two options, there is only one preference transformation function that satisfies all five desiderata, namely the trivial function under which there is no preference change at all (i.e., $P^*_1 = P_1$, $P^*_2 = P_2$, $P^*_3 = P_3$, and so on). In the special case of only two options, *deference to the majority preference* is a non-trivial transformation function that satisfies all five desiderata.

As with other impossibility results, we should not over-interpret this result, but view it as telling us something about the desiderata that a deliberative procedure could or could not satisfy. For instance, pairwise independence is arguably too restrictive. Plausibly, deliberation is *holistic*: an individual's post-deliberation preference between two options may depend on what he or she learns about others' pre-deliberation preferences for third options. The downside of this holism is that deliberation may be vulnerable to strategic misrepresentation of preferences. Whether this is a serious problem will depend on a number of empirical facts, such as the participants' motivations and dispositions. Likewise, there may be situations in which a deliberative procedure need not satisfy some of the other desiderata. We may be prepared to relax universal domain in cases in which we can assume a certain level of pre-deliberation agreement. And we may be prepared to relax post-deliberation rationality if a complete ranking of the options by the deliberators is not required.

A different response is to challenge the assumption of functionality. Are post-deliberation preferences really fully determined by pre-deliberation preferences, as functionality requires? Or is deliberation an indeterministic or stochastic process? One way to relax functionality without denying that the output of the deliberation is determined by its input is to argue that post-deliberation preferences depend on a richer input, beyond pre-deliberation *preferences*. The enriched input might include the participants' *information* and/or their *reasons* for preferring or dispreferring the options. However, enriching the input of deliberation does not automatically allow us to bypass the present impossibility theorem. A variant also holds when the inputs and outputs take the form of judgments rather than preferences (List 2011b).[18]

5.2 A model of deliberation and aggregation

While the model just discussed represents the transformation of preferences without modelling their subsequent aggregation, a combined model of preference transformation and aggregation has been developed by Perote-Peña and Piggins (2015). In this model, a deliberation stage at which preferences are transformed is followed by a voting stage at which preferences are

[17] Formally, two preference profiles are called *i-variants* if they coincide for all individuals except *i*. Minimal relevance requires that, for each individual *i*, there be at least one admissible pair of *i*-variant pre-deliberation profiles for which individual *i*'s post-deliberation preference orderings differ.

[18] On the transformation of judgments in deliberation, see also Slavkovik and Jamroga (2016) and, in relation to probabilistic judgments, Lehrer and Wagner (1981).

aggregated, as in Figure 3 above. Unlike most social choice theorists, who adopt a non-cognitivist (desire or taste) interpretation of preferences, Perote-Peña and Piggins assume that each individual's preference ordering represents his or her beliefs as to how good the options are from the perspective of the group, where there exists a true objective betterness ordering. They further assume, in a deliberative spirit, that individuals reveal their preferences sincerely.

The model allows us to investigate the conditions under which the combination of deliberation and voting is *truth-revealing*, in that it leads the group to make the "correct" social choice (according to the true betterness ranking of the options), irrespective of the group's initial preference profile.[19] Perote-Peña and Piggins assume that, in deliberation, any individual who holds the correct preference over a given pair of options will not abandon this preference, but may have a chance of persuading others to adopt that preference. Others, in turn, will be open to persuasion if they perceive the given individual's preferences as sufficiently close to their own preferences, but not if they perceive them as too distant. This reflects the idea that people are more likely to be persuaded by those whose views are similar to their own views.

It then follows that the key determinant of truth revelation of the deliberation-and-aggregation process is the *persuasion cost*. This is defined as a measure of how similar any two individuals' preference orderings must be for one of the individuals to have a chance of persuading the other. If the persuasion cost is high, then truth revelation – indeed, any preference change – is not generally possible to achieve. In this case, the individuals are unwilling to be persuaded by anyone except those whose preferences are maximally similar to their own preferences. If the persuasion cost is not too high, by contrast, then truth revelation is feasible. Here, the individuals are, in effect, more willing to listen to others. Truth revelation, however, requires a carefully designed deliberation procedure, specifying who talks to whom and in which order, and a suitably chosen aggregation rule.

5.3 A game-theoretic model of deliberation as information sharing

A third model of deliberation can be motivated by reference to a game-theoretic observation about jury decisions (Feddersen and Pesendorfer 1998). Suppose a 12-member jury (initially in conditions without deliberation) has to reach a verdict in a criminal trial and uses the unanimity rule, where a "guilty" verdict is reached if and only if all jurors vote for "guilty". Suppose further that each juror has received some independent and private information, a binary signal of the form "guilty" or "not guilty", which is fallible but correlated with the truth; say, it has a 70% chance of being correct.[20] Under these assumptions, one would think it is very unlikely that an innocent person will be convicted. If all jurors vote truthfully – i.e., in line with their private information – there is only a chance of $(30\%)^{12}$ that a unanimous verdict for "guilty" will be incorrect: less than one in a million.

However, suppose that each juror is committed to the principle "convict if and only if the defendant's guilt is beyond reasonable doubt", understood as a probability of guilt above 99%. Surprisingly, the jurors may then lack an incentive to vote truthfully. Suppose I am one of the 12 jurors, and suppose, for the sake of argument, the others will vote truthfully. Should I then vote truthfully too? Note that my vote will make a difference only if everyone else votes for

[19] This is subject to one restriction: in this initial preference profile, at least one individual must hold each logically possible preference.

[20] For the purposes of the example, I assume that there is a common prior probability of ½ of guilt, and that, given guilt, each juror has a 70% chance of receiving a "guilty" signal, and given innocence, each juror has a 70% chance of receiving a "not guilty" signal.

"guilty". If some others vote for "not guilty", then the outcome will be a "not guilty" verdict, no matter how I vote. If everyone else votes for "guilty", on the other hand, then my vote will be pivotal; it will determine whether we reach unanimity. Now, if the others vote truthfully, there is only a small chance that they are all wrong: all 11 signals would have to be incorrect, a chance of $(30\%)^{11}$: less than two in a million and well below the threshold of reasonable doubt. So, whether or not my own signal supports a "guilty" verdict, I should vote for "guilty" under the present assumptions. Even if my own signal suggests "not guilty", it is more likely that this signal is wrong than that the others' signals are all wrong.

The example shows that even when all jurors have the same goal – namely to convict if and only if the defendant's guilt is beyond reasonable doubt – voting truthfully is not generally a dominant strategy. In consequence, unanimous jury decisions may fail to protect the innocent from wrongful convictions. Other aggregation rules, such as the majority rule, suffer from similar problems (Austen-Smith and Banks 1996). Crucially, however, all these problems arise under a purely aggregative model of decision making: voters have no opportunity to communicate and share their private information before voting.

What happens if we introduce deliberation? A game-theoretic body of work addresses this question (e.g., Coughlan 2000; Austen-Smith and Feddersen 2005, 2006; Meirowitz 2006; Landa and Meirowitz 2009). Deliberation is modelled thinly as an opportunity for voters to share their private information before voting. We can think of this as a straw poll, an informal round of voting in which everyone can reveal their signal to the group. We must now distinguish between two cases: the case in which there is a consensus on the underlying goal, and the case in which there is not (Austen-Smith and Feddersen 2006). An *underlying consensus* is a situation in which, if all private information were publicly revealed, voters would always agree on what the correct choice is. When there is an underlying consensus, any disagreements stem from differences in private information. In the jury example, where all jurors agree on the threshold of reasonable doubt and on the prior probability of guilt,[21] there is an underlying consensus in this sense. By contrast, when there is no underlying consensus, different voters would support different options even conditional on full disclosure of all private information. In that case, different voters are said to have different *biases*.

Clearly, if there is an underlying consensus, then deliberation is always helpful (Coughlan 2000). In the jury example, sharing all private signals among the jurors would mean that everyone will come to the same view on whether the defendant should be convicted, and the jurors would vote unanimously either for "guilty" or for "not guilty", depending on the shared information. Furthermore, the jurors would have no incentives to misrepresent their private signals during deliberation. Given the underlying consensus, it is in everyone's interest that all private information be revealed truthfully. If we think of deliberation as a game, truthful information sharing is a dominant strategy here.

By contrast, if there is no underlying consensus, the situation is more complicated. In the jury example, some jurors might be committed to a 1% threshold of reasonable doubt, others to a 10% threshold, still others to a 25% threshold, and so on. Furthermore, different jurors might assign different prior probabilities to the guilt of the defendant. We must then distinguish between two sub-cases. If the voters' biases are common knowledge, then, unfortunately, truthful information sharing may no longer be a dominant strategy in deliberation (Coughlan

[21] Also, the reliability of the jurors' signals is common knowledge.

2000). Participants may engage in untruthful cheap talk.[22] If, instead, voters are uncertain about the biases of others, then there may be some scope for deliberation to induce truth-telling, but whether it does will depend on the aggregation rule that is used after deliberation (Austen-Smith and Feddersen 2006). If it is the majority rule, then, under some conditions, it is rational for voters to reveal their signals truthfully during deliberation.[23] By contrast, under the unanimity rule, voters continue to have incentives to be untruthful in deliberation. In fact, the unanimity rule seems uniquely bad at incentivizing truth-telling in deliberation.[24]

The bottom line is that if we model deliberation as nothing more than an opportunity for voters to share private information before voting – a form of "straw polling" – then the participants' incentives for and against truth-telling depend on several factors: first, whether there is a consensus on the underlying goal (if so, deliberation induces truth-telling; if not, it doesn't generally do so); second, whether, in the absence of a consensus, there is uncertainty about voters' biases (if so, truth-telling is sometimes rational; if not, it may not be); and third, the voting rule (if it is non-unanimitarian, like the majority rule, then deliberation sometimes induces truth-telling; if it is the unanimity rule, it may not). One response to these conclusions is to argue that if we do not model deliberation as unverifiable cheap-talk but assume that deliberators can report verifiable evidence, then the unanimity rule as well as disclosure of biases may be conducive to truthful information sharing (Mathis 2011).

5.4 The mechanisms of deliberation-induced opinion change

The formal models of deliberation I have reviewed differ significantly in what they assume about the individual-level mechanisms of deliberation-induced opinion change. The preference-transformation model in Section 5.1 is most abstract, modelling deliberation simply as a process in which pre-deliberation preferences are transformed into post-deliberation preferences, without specifying any micro-mechanisms. The combined model of deliberation and aggregation in Section 5.2 suggests that deliberation can change preferences via changing the participants' beliefs about how good the options are, while presupposing that there is an objective betterness ordering. The game-theoretic model in Section 5.3 portrays deliberation as a process that can change the participants' factual-empirical beliefs, by providing them with new information, while not changing their underlying goals or preferences. The last model,

[22] [Online only] Austen-Smith and Feddersen (2006) explain the problem as follows. Consider a three-member jury. Suppose two jurors have a "high" bias: they would support a "guilty" verdict only if all jurors' private signals support guilt. And suppose the third juror has a "low" bias: he or she would support a "guilty" verdict as soon as one of the signals supports guilt. Austen-Smith and Feddersen note that "given a voting rule and given certainty regarding the [biases] of others, a juror can identify the information possessed by others under which the revelation of his or her own information is pivotal" (p. 210). In the present example, "if the rule requires all jurors to vote for conviction, then the only circumstance in which any one juror's information is relevant is when the two [high-bias jurors] have both observed the guilty signal. In this case, the low-bias juror always prefers conviction, and so does better by misreporting his signal and voting to convict in the (cheap-talk stage) straw poll" (p. 210).

[23] [Online only] As Austen-Smith and Feddersen (2006, pp. 210-211), explain: "When information is sufficiently good, jurors put more weight on the event that other jurors have observed information similar to their own and therefore (if deliberation matters) it is more likely that the other jurors share her [judgment], in which case telling the truth is in the individual's best interest".

[24] [Online only] As Austen-Smith and Feddersen (2006, p. 211) put it, each voter's ability to veto a positive decision (conviction) under the unanimity rule means that "the only pivotal event at the communication stage is when all others have information such that, by speaking in favor of conviction, a juror convinces the others to vote to convict when at least one of them would not otherwise do so; so either the individual, on the basis of what he or she learns from others' straw votes, wishes to convict and lying turns out to be in his or her interest or, given what is learned from others, the individual wishes to acquit and can ensure this simply by voting to acquit whatever he or she might have said in debate".

unlike the first two, is based on the classical rational-choice-theoretic picture of agency. I will now explain why this picture leaves relatively little room for deliberation-induced opinion change, and how we might lift its restrictions.

According to classical rational choice theory, an agent has preferences over different possible outcomes and beliefs over different possible states of the world, and acts so as to satisfy those preferences in accordance with those beliefs. Preferences are usually represented by some utility function, and beliefs by some subjective probability function. Rational choice then consists in choosing an action that maximizes expected utility, where actions are represented by functions from states of the world to outcomes (Savage 1954).[25] According to this picture, an agent's preferences over fully specified outcomes never change. Only beliefs about the state of the world are open to revision, and belief revision only happens when the agent learns new information, via Bayesian conditioning.

In this way, the agent's "surface-level" preferences over uncertain prospects may change in response to new information, while the underlying fundamental preferences remain fixed. I may change my surface-level preference between nuclear energy and coal energy when I learn how each technology affects the environment, but my fundamental preferences, such as between a clean and a polluted environment, remain the same. Classical rational choice theory further assumes full rationality: the agent has consistent beliefs and preferences, is aware of all their implications, and can calculate the best response in any choice situation. On this picture, there is no scope for deliberation-induced opinion change beyond the learning of new information.

By contrast, once we relax rational choice theory's restrictive assumptions and replace them with more realistic ones, we can identify several significant ways in which deliberation might change opinions. First, there can be forms of belief change distinct from Bayesian conditioning. Under Bayesian conditioning, belief changes are possible only when the agent acquires new information. Furthermore, this always takes the form of ruling out certain states of the world, namely those excluded by the information. The probabilities over the remaining states of the world are then reassigned in proportion to the agent's prior probabilities. A Bayesian agent cannot learn that his or her prior probabilities were incorrect. A more permissive form of conditioning is Jeffrey conditioning. Here, the agent can acquire a new probability distribution over different states of the world, without ruling out any of them. Jeffrey conditioning permits updates of the agent's prior probabilities, over and above Bayesian information acquisition. If deliberators can engage in Jeffrey conditioning, then deliberation might induce changes in their subjective prior probabilities and not just provide them with new information in the Bayesian sense (e.g., Halpern 2003; Bradley 2009; Dietrich, List, and Bradley 2016).

Second, an agent's fundamental preferences might themselves be open to change. In particular, the agent's preferences may depend on how the agent perceives or conceptualizes the options (even fully specified outcomes), and a change in perception or conceptualization might lead to a preference change (Dietrich and List 2011, 2013a,b). For example, I might initially perceive different policies solely in terms of how they affect my budget and well-being, without considering their effects on other people, let alone their effects on distant strangers and the environment. My focus on a limited set of good-making features of the options need not be due to a lack of information. It may be due to a lack of salience of other considerations. If new

[25] [Online only] One does not generally know which outcome will result from a given action if one is uncertain about the state of the world; the same action may lead to different outcomes in different states.

considerations become salient during deliberation – for example, I come to see them as relevant reasons for preferring some options over others – this might lead to a preference change. (On different forms of preference change, see also Grüne-Yanoff and Hansson 2009 and Bradley 2009.)

Third, since real people are not logically omniscient – i.e., they are not aware of all the implications of their beliefs – there is further scope for opinion change in deliberation. Deliberators may learn that their beliefs have implications they had not previously noticed. This may, in turn, induce a "reflective equilibrium" process, in which participants either come to endorse these hitherto unrecognized implications, or revise some of their beliefs in order to avoid any unwanted implications. Participants may also come to recognize some inconsistencies in their opinions and correct them. Similarly, they might notice ambiguities that need to be addressed. A hypothetical "homo economicus", by contrast, would never suffer from any limitations of rationality. Such an agent would not be able to "discover" any problematic features or unrecognized implications of his or her own prior views. For a model of deliberation as "self-discovery" under bounded rationality, see Hafer and Landa (2007).

In sum, depending on how far we depart from classical rational choice theory, deliberation may lead participants to change their opinions in at least four ways: by giving them new information; by drawing their attention to unrecognized implications, inconsistencies, or ambiguities within their beliefs and preferences; by leading participants to reflect on the considerations or reasons that are relevant and/or publicly justifiable; and by putting participants in a social situation in which they come to relate to others. These may be described as the "informational", "argumentative", "reflective", and "social aspects" of deliberation (Dryzek and List 2003).

6. Deliberation and the aggregation of judgments

A final area of social choice theory that is sometimes associated with deliberation is judgment-aggregation theory (formalized in List and Pettit 2002 and Dietrich 2007; for a survey, see List and Puppe 2009). Its focus is not on the aggregation of preference orderings, but on the aggregation of judgments, understood as "true/false" or "yes/no" verdicts on some statements or propositions, usually with logical connections between them.

Judgment-aggregation theory was inspired by the study of decision making in collegial courts (Kornhauser and Sager 1986, 1993). For example, a court may need to make judgments on propositions such as "the defendant has committed a particular act" (a), "the act was contractually prohibited" (b), and "the defendant is liable for a breach of contract" (c), where a and b are jointly necessary and sufficient for c. The task is not to arrive at a collective preference ordering, but to make collective "yes/no" judgments on the relevant propositions. For another common example, consider an expert committee faced with propositions such as "atmospheric CO_2 exceeds 400ppm" (d), "if CO_2 exceeds 400ppm, then the Greenland ice will melt" ($d \rightarrow e$), and "the Greenland ice will melt" (e). Again, the task is to arrive at collective judgments, based on the underlying individual judgments.

Judgment-aggregation theory offers a formal framework for modelling such aggregation problems and for identifying suitable aggregation rules that satisfy certain desiderata. By accommodating decisions involving "true/false" or "yes/no" judgments on logically connected propositions, it allows us to analyze a different class of decision problems than traditionally

modelled in social choice theory.[26] At first, it is not obvious why judgment-aggregation theory should be any more relevant to deliberation than traditional social choice theory is. However, its relevance lies in the fact that it can model decisions whose content goes beyond the ranking of options, and the interest in such decisions is shared by theorists of deliberative democracy. The content of a judgment-aggregation problem may be an entire web of interconnected propositions, akin to what philosophers call a "web of belief" (e.g., Quine and Ullian 1978). A court cannot simply hold the defendant liable without accepting the reasons for liability. An expert committee, similarly, cannot plausibly declare that the ice will melt if it does not also accept the premises to back up this judgment.

Once we recognize the need to arrive at collective judgments on webs of propositions, we can see some further limitations of a crude aggregative approach. Several much-discussed examples – the "doctrinal" or "discursive paradoxes" – show that propositionwise majority voting is problematic (Kornhauser and Sager 1986, 1993; Pettit 2001; List and Pettit 2002).[27] Suppose, for instance, the individual judgments in a court or in an expert committee are as shown in Tables 1 and 2.

Table 1: A profile of individual judgments in the court example

	a	*b*	*c*
Individual 1	True	True	True
Individual 2	True	False	False
Individual 3	False	True	False
Majority	True	True	False

Table 2: A profile of individual judgments in the expert-committee example

	d	$d \rightarrow e$	*e*
Individual 1	True	True	True
Individual 2	True	False	False
Individual 3	False	True	False
Majority	True	True	False

In each case, all individuals' judgments are internally consistent, and yet the majority judgments are inconsistent. If a court were to endorse the majority judgments in Table 1, it would hold that the defendant did the relevant act, that the act was prohibited, and yet that the defendant is not liable, despite accepting that the first two propositions jointly imply liability. Similarly, if the expert committee accepted the majority judgments in Table 2, it would be committed to an inconsistent view: CO_2 is above the threshold; this will cause the melting of the ice; and yet the ice will not melt.

In an attempt to avoid such inconsistencies, early discussions of judgment aggregation revolved around the contrast between "premise-based" and "conclusion-based" procedures (e.g., Kornhauser and Sager 1986, 1993; Chapman 1998; Pettit 2001). Under a premise-based procedure, the group takes majority votes only on premises – e.g., *a* and *b* or *d* and $d \rightarrow e$ – and

[26] [Online only] In fact, even preference aggregation problems can be translated into the relevant format, by re-expressing preference orderings as binary ranking judgments. For instance, a preference for *x* over *y* over *z* can be re-expressed as the set of propositions *xPy*, *yPz*, and *xPz*, where "*P*" stands for "is preferred/preferable to".

[27] [Online only] The term "doctrinal paradox" refers to the fact that the premise-based and conclusion-based procedures (discussed in the main text) may lead to opposite outcomes, while the term "discursive dilemma" (or simply: "problem of majoritarian inconsistency") refers to a broader point: namely that propositionwise majority voting on logically connected propositions may lead to inconsistent collective judgments.

then derives its judgments on all conclusions – e.g., *c* or *e* – by logical inference. In both tables, the result would be the collective acceptance of all three propositions, despite the majority against the proposition in the last column. Under a conclusion-based procedure, the group takes majority votes only on conclusions – e.g., *c* or *e* – while not making any judgments on premises. In each table, the result would be the collective rejection of the proposition in the final column, which shows that the premise-based and conclusion-based procedures may lead to opposite outcomes, even when the individual judgments are the same.

It has been argued that the premise-based procedure is more consistent with deliberative democracy's emphasis on reason-giving, while the conclusion-based procedure embodies a more minimalist conception of democracy (Pettit 2001; List 2006). Technically, both procedures are still aggregation rules and need not involve any pre-vote deliberation. Yet, the premise-based procedure captures the idea that the group *as a collective* draws some reasoned inferences from the propositions it accepts by voting, an idea that Pettit (2001) has described as the "collectivization of reason". The conclusion-based procedure, by contrast, captures the idea of a minimalist collective outcome, namely what Sunstein (1994) calls an "incompletely theorized agreement". Sometimes it may be controversial which propositions are "premises" and which are "conclusions". Different orders of priority or premisehood among the propositions may, in turn, give rise to different collective judgments, thereby generating the problem of path dependence and agenda manipulability (List 2004; Dietrich 2016).

Technical work (reviewed in List and Puppe 2009) has focused on characterizing the conditions under which consistent judgment aggregation is or is not possible, studying the properties of practicable aggregation rules, and identifying conditions that will or will not induce truth-telling. Here, it has been suggested that deliberatively motivated agents are less prone to strategic voting than purely outcome-oriented agents (Dietrich and List 2007).

As a heuristic tool, judgment-aggregation theory, like social choice theory more generally, can sharpen our understanding of the tradeoffs between different desiderata of democracy, such as robustness to pluralism, responsiveness to the majority opinion, and collective rationality. One of social choice theory's lessons is that we can achieve any two of these desiderata, but not all three (List 2011a). The challenge, in the debate about deliberation and social choice, is to decide which of those desiderata to keep and which to relax.

References

Ackerman, B., and J. S. Fishkin (2002) "Deliberation Day", *Journal of Political Philosophy* 10: 129-152.

Aldred, J. (2004) "Social Choice Theory and Deliberative Democracy: A Comment", *British Journal of Political Science* 34: 747-752.

Arrow, K. (1951/1963) *Social Choice and Individual Values*, New York: Wiley.

Austen-Smith, D., and J. S. Banks (1996) "Information Aggregation, Rationality, and the Condorcet Jury Theorem", *American Political Science Review* 90: 34-45.

Austen-Smith, D., and J. S. Banks (2000) *Positive Political Theory I: Collective Preference*, Ann Arbor: University of Michigan Press.

Austen-Smith, D., and J. S. Banks (2005) *Positive Political Theory II: Strategy and Structure*, Ann Arbor: University of Michigan Press.

Austen-Smith, D., and T. J. Feddersen (2005) "Deliberation and Voting Rules", in D. Austen-Smith and J. Duggan (eds.), *Social Choice and Strategic Decisions: Essays in Honor of Jeffrey S. Banks*, Heidelberg / New York: Springer.

Austen-Smith, D., and T. J. Feddersen (2006) "Deliberation, Preference Uncertainty, and Voting Rules", *American Political Science Review* 100: 209-217.

Balinski, M., and R. Laraki (2011) *Majority Judgment: Measuring, Ranking, and Electing*, Cambridge, MA: MIT Press.

Barabas, J. (2004) "How Deliberation Affects Policy Opinions", *American Political Science Review* 98: 687-701.

Black, D. (1948) "On the Rationale of Group Decision-Making", *Journal of Political Economy* 56: 23-34.

Bohman, J., and W. Rehg (2014) "Jürgen Habermas", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition). <http://plato.stanford.edu/archives/fall2014/entries/habermas/>

Bradley, R. (2009) "Becker's thesis and three models of preference change", *Politics, Philosophy and Economics* 8: 223-242.

Brighouse, H., and M. Fleurbaey (2010) "Democracy and Proportionality", *Journal of Political Philosophy* 18: 137-155.

Chapman, B. (1998) "More Easily Done Than Said: Rules, Reasons and Rational Social Choice", *Oxford Journal of Legal Studies* 18: 293-329.

Cohen, J. (1986) "An Epistemic Conception of Democracy", *Ethics* 97: 26-38.

Cohen, J. (1989) "Deliberation and Democratic Legitimacy", in A. Hamlin and P. Pettit (eds.), *The Good Polity: Normative Analysis of the State*, Oxford: Basil Blackwell, pp. 17-34.

Coleman, J., and J. Ferejohn (1986) "Democracy and Social Choice", *Ethics* 97: 6-25.

Coughlan, P. J. (2000) "In Defense of Unanimous Jury Verdicts: Mistrials, Communication, and Strategic Voting", *American Political Science Review* 94: 375-393.

Dietrich, F. (2007) "A generalised model of judgment aggregation", *Social Choice and Welfare* 28: 529-565.

Dietrich, F. (2016) "Judgment aggregation and agenda manipulation", *Games and Economic Behavior* 95: 113-136.

Dietrich, F., and C. List (2007) "Strategy-proof judgment aggregation", *Economics and Philosophy* 23: 269-300.

Dietrich, F., and C. List (2011) "A Model of Non-informational Preference Change", *Journal of Theoretical Politics* 23: 145-164.

Dietrich, F., and C. List (2013a) "A Reason-based Theory of Rational Choice", *Noûs* 47: 104-134.

Dietrich, F., and C. List (2013b) "Where do Preferences Come From?", *International Journal of Game Theory* 42: 613-637.

Dietrich, F., C. List, and R. Bradley (2016) "Belief revision generalized: A joint characterization of Bayes' and Jeffrey's rules", *Journal of Economic Theory* 162: 352-371.

Dietrich, F., and K. Spiekermann (2013) "Epistemic democracy with defensible premises", *Economics and Philosophy* 29: 87-120.

Dryzek, J. S. (2000) *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*, Oxford: Oxford University Press.

Dryzek, J., and C. List (2003) "Social Choice Theory and Deliberative Democracy: A Reconciliation", *British Journal of Political Science* 33: 1-28.

Dryzek, J., and C. List (2004) "Social Choice Theory and Deliberative Democracy: A Response to Aldred", *British Journal of Political Science* 34: 752-758.

Dryzek, J., and S. Niemeyer (2006) "Reconciling Pluralism and Consensus as Political Ideals", *American Journal of Political Science* 50: 634-649.

Elster, J. (1986) "The Market and the Forum", in J. Elster and A. Hylland (eds.), *Foundations of Social Choice Theory*, Cambridge: Cambridge University Press, pp. 103-132.

Farrar, C., J. S. Fishkin, D. Green, C. List, R. C. Luskin and E. L. Paluck (2010) "Disaggregating Deliberation's Effects: An Experiment within a Deliberative Poll", *British Journal of Political Science* 40: 333-347.

Feddersen, T. J. (1998) "Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting", *American Political Science Review* 92: 23-35.

Feld, S. L., and B. Grofman (1988) "Ideological Consistency as a Collective Phenomenon", *American Political Science Review* 82: 773-788.

Fishkin, J. S. (2009) *When the People Speak: Deliberative Democracy and Public Consultation*, Oxford: Oxford University Press.

Fishkin, J. S., and R. C. Luskin (2005) "Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion", *Acta Politica* 40: 284-298.

Gehrlein, W. V. (1983) "Condorcet's Paradox", *Theory and Decision* 15: 161-197.

Gehrlein, W. V. (2006) *Condorcet's Paradox*, Heidelberg / New York: Springer.

Gehrlein, W. V., and D. Lepelley (2011) *Voting Paradoxes and Group Coherence: The Condorcet Efficiency of Voting Rules*, Heidelberg / New York: Springer.

Gehrlein, W. V., D. Lepelley, and I. Moyouwou (2015) "Voters' preference diversity, concepts of agreement and Condorcet's paradox", *Quality and Quantity* 49: 2345-2368.

Goldbach, R. (2015) "Modelling Democratic Deliberation", MSc thesis, University of Amsterdam. <https://www.illc.uva.nl/Research/Publications/Reports/MoL-2015-05.text.pdf>

Goodin, R. E. (2008) *Innovating Democracy: Democratic Theory and Practice After the Deliberative Turn*, Oxford: Oxford University Press.

Gibbard, A. (1973) "Manipulation of voting schemes: a general result", *Econometrica* 41: 587-601.

Grönlund, K., K. Herne, and M. Setälä (2015) "Does Enclave Deliberation Polarize Opinions?", *Political Behavior* 37: 995-1020.

Grofman, B., G. Owen, and S. L. Feld (1983) "Thirteen theorems in search of the truth", *Theory and Decision* 15: 261-278.

Grüne-Yanoff, T., and S. O. Hansson (2009) *Preference Change: Approaches from philosophy, economics and psychology*, Dordrecht and Heidelberg: Springer.

Gutmann, A., and D. Thompson (1996) *Democracy and Disagreement*, Cambridge, MA: Harvard University Press.

Hafer, C., and D. Landa (2007) "Deliberation as Self-Discovery and Institutions for Political Speech", *Journal of Theoretical Politics* 19: 329-360.

Halpern, J. (2003) *Reasoning About Uncertainty*, Cambridge, MA: MIT Press.

Knight, J., and J. Johnson (1994) "Aggregation and Deliberation: On the Possibility of Democratic Legitimacy", *Political Theory* 22: 277-296.

Kornhauser, L. A., and L. G. Sager (1986) "Unpacking the Court", *Yale Law Journal* 96: 82-117.

Kornhauser, L. A., and L. G. Sager (1993) "The One and the Many: Adjudication in Collegial Courts", *California Law Review* 81: 1-59.

Landa, D., and A. Meirowitz (2009) "Game Theory, Information, and Deliberative Democracy", *American Journal of Political Science* 53: 427-444.

Lehrer, K., and C. Wagner (1981) *Rational Consensus in Science and Society*, Dordrecht: D. Reidel.

List, C. (2002) "Two Concepts of Agreement", *The Good Society* 11: 72-79.

List, C. (2004) "A Model of Path-Dependence in Decisions over Multiple Propositions", *American Political Science Review* 98: 495-513.

List, C. (2006) "The Discursive Dilemma and Public Reason", *Ethics* 116: 362-402.

List, C. (2008) "Deliberation and Agreement", in S. W. Rosenberg (ed.), *Deliberation, Participation and Democracy: Can the People Govern?* Basingstoke: Palgrave Macmillan, pp. 64-81.

List, C. (2011a) "The Logical Space of Democracy", *Philosophy and Public Affairs* 39: 262-297.

List, C. (2011b) "Group Communication and the Transformation of Judgments: An Impossibility Result", *Journal of Political Philosophy* 19: 1-27.

List, C. (2013) "Social Choice Theory", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition). <http://plato.stanford.edu/archives/win2013/entries/social-choice/>

List, C., and R. E. Goodin (2001) "Epistemic Democracy: Generalizing the Condorcet Jury Theorem", *Journal of Political Philosophy* 9: 277-306.

List, C., R. C. Luskin, J. S. Fishkin, and I. McLean (2013) "Deliberation, Single-Peakedness, and the Possibility of Meaningful Democracy: Evidence from Deliberative Polls", *Journal of Politics* 75: 80-95.

List, C., and P. Pettit (2002) "Aggregating Sets of Judgments: An Impossibility Result", *Economics and Philosophy* 18: 89-110.

List, C., and C. Puppe (2009) "Judgment aggregation: a survey", in P. Anand, C. Puppe and P. Pattanaik (eds.), *Oxford Handbook of Rational and Social Choice*, Oxford: Oxford University Press.

Luskin, R. C., J. S. Fishkin, and R. Jowell (2002) "Considered Opinions: Deliberative Polling in Britain", *British Journal of Political Science* 32: 455-487.

Mackie, G. (2003) *Democracy Defended*, Cambridge: Cambridge University Press.

Mackie, G. (2006) "Does democratic deliberation change minds?" *Politics, Philosophy and Economics* 5: 279-303.

Mansbridge, J., J. Bohman, S. Chambers, D. Estlund, A. Føllesdal, A. Fung, C. Lafont, B. Manin and J. L. Martí (2010) "The Place of Self-Interest and the Role of Power in Deliberative Democracy", *Journal of Political Philosophy* 18: 64-100.

Mathis, J. (2011) "Deliberation with Evidence", *American Political Science Review* 105: 516-529.

May, K. O. (1952) "A set of independent, necessary and sufficient conditions for simple majority decision", *Econometrica* 20: 680-684.

McGann, A. (2006) *The Logic of Democracy: Reconciling Equality, Deliberation, and Minority Protection*, Ann Arbor: University of Michigan Press.

Meirowitz, A. (2006) "Designing Institutions to Aggregate Preferences and Information", *Quarterly Journal of Political Science* 1: 373-392.

Miller, D. (1992) "Deliberative Democracy and Social Choice", *Political Studies* 40 (special issue): 54-67.

Miller, N. (1983) "Pluralism and Social Choice", *American Political Science Review* 77: 734-747.

Moulin, H. (1980) "On Strategy-Proofness and Single Peakedness", *Public Choice,* 35: 437-455.

Mueller, D. C. (2003) *Public Choice III*, Cambridge: Cambridge University Press.

Niemi, R. G. (1969) "Majority Decision-Making with Partial Unidimensionality", *American Political Science Review* 63: 488-497.

Ottonelli, V., and D. Porello (2013) "On the elusive notion of meta-agreement", *Politics, Philosophy and Economics* 12: 68-92.

Perote-Peña, J., and A. Piggins (2015) "A Model of Deliberative and Aggregative Democracy", *Economics and Philosophy* 31: 93-121.

Porello, D. (2016) "Single-peakedness and semantic dimensions of preferences", *Logic Journal of the IGPL* 24: 570-583.

Quine, W.V., and J. S. Ullian (1978), *The Web of Belief*, second edition, New York: Random House.

Regenwetter, M., B. Grofman, A. A. J. Marley, I. Tsetlin (2006) *Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications*, Cambridge: Cambridge University Press.

Riker, W. (1982) *Liberalism against Populism*, San Francisco: W.H. Freeman and Co.

Satterthwaite, M. (1975) "Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions", *Journal of Economic Theory* 10: 187-217.

Savage, L. (1954) *The Foundations of Statistics*, New York: Wiley.

Sen, A. K. (1966) "A Possibility Theorem on Majority Decisions", *Econometrica,* 34: 491-499.

Sen, A. K. (1998) "The Possibility of Social Choice", Nobel lecture, December 8, 1998, Stockholm. <http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1998/sen-lecture.pdf>

Slavkovik, M., and W. Jamroga (2016), "A Deliberation Procedure for Judgment Aggregation Problems", arXiv:1604.06356

Sunstein, C. R. (1994) "Political Conflict and Legal Agreement", Tanner Lectures on Human Values, Harvard. <http://tannerlectures.utah.edu/_documents/a-to-z/s/Sunstein96.pdf>

Sunstein, C. R. (2002) "The Law of Group Polarization", *Journal of Political Philosophy* 10: 175-195.

Tsetlin, I., M. Regenwetter, and B. Grofman (2003) "The impartial culture maximizes the probability of majority cycles", *Social Choice and Welfare* 21: 387-398.

van Mill, D. (1996) "The Possibility of Rational Outcomes from Democratic Discourse and Procedures", *Journal of Politics* 58: 734-752.

van Mill, D. (2006) *Deliberation, Social Choice and Absolutist Democracy*, London: Routledge.