On the elusive notion of meta-agreement

Valeria Ottonelli and Daniele Porello

Politics Philosophy Economics published online 29 March 2012
DOI: 10.1177/1470594X11433742

Article

Politics, Philosophy & Economics
1–25

DOI: 10.1177/1470594X11433742
ppe.sagepub.com
SAGE

On the elusive notion of meta-agreement

Valeria Ottonelli
University of Genova, Italy

Daniele Porello
University of Amsterdam, The Netherlands

Abstract
Public deliberation has been defended as a rational and noncoercive way to overcome paradoxical results from democratic voting, by promoting consensus on the available alternatives on the political agenda. Some critics have argued that full consensus is too demanding and inimical to pluralism and have pointed out that single-peakedness, a much less stringent condition, is sufficient to overcome voting paradoxes. According to these accounts, deliberation can induce single-peakedness through the creation of a 'meta-agreement', that is, agreement on the dimension according to which the issues at stake are 'conceptualized'. We argue here that once all the conditions needed for deliberation to bring about single-peakedness through meta-agreement are unpacked and made explicit, meta-agreement turns out to be a highly demanding condition, and one that is very inhospitable to pluralism.

Keywords
meta-agreement, deliberation, single-peakedness, pluralism, consensus

1. Introduction

According to some advocates of deliberative democracy, one of its key virtues is its tendency to promote consensus. By publicly exchanging information and arguments, the

Corresponding author:
Daniele Porello, Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Postbus 94242, Amsterdam, 1090GE, Netherlands
Email: danieleporello@gmail.com

democratic public can align its otherwise scattered and conflicting views about how to rank available policy options. Furthermore, it does this through a fully rational and non-coercive process.[1] This not only may appear as good in itself, but is also a sure remedy against the dangers of Arrovian cycles and the inconsistencies associated with democratic decision-making procedures, which have been brought to light and analyzed by social choice theory (Arrow, 1963; Riker, 1982). If there is consensus on how to rank and assess different available political options, then cycles do not occur and democracy is safeguarded against irrationality and manipulation.

However, some critics have argued that aiming at consensus as a solution to these problems might not be the right strategy. Consensus, they have pointed out, is very difficult to achieve, even in the best deliberative circumstances. Moreover, aiming at consensus as a condition for the correct functioning of democracy shows disregard for pluralism, which is a defining feature of modern democratic societies and one that (at least within certain limits) deserves respect.

In the past few years, a new and different account has emerged of how deliberation produces cycle-free democratic decisions. Specifically, what deliberation does, and should do, is not to bring about a *substantive agreement* on how to rank and assess the different options on the political agenda. Rather, it should encourage a *meta-agreement* concerning how those options should be 'conceptualized'.[2]

The basic idea behind this claim is that substantive agreement (that is, identical orderings of the available alternatives by individual voters) is not only difficult to achieve, but also unnecessary for freeing democracy from cycles and paradoxical results. As shown by Duncan Black more than 50 years ago, a sufficient condition for avoiding such results is that citizens' preference orderings be 'single-peaked'; this means that although individuals' preferences may be different, they can be represented spatially along the same left–right dimension.

Deliberation, it is argued, tends to bring about single-peakedness because it leads people to justify and represent their preference orderings according to notions and principles that others can share, in order to make their reasons available to the rest of the public. This, in turn, tends to lead people to represent the relations between the available alternatives along the same conceptual dimensions, thus making them fit the linear spatial ordering associated with single-peakedness. This process may imply that people change some of their preferences; however, it does not imply that people will need to bring their preference orderings to coincide perfectly. In other words, meta-agreement and single-peakedness are compatible with people having different views on how the available political options should be ordered.

The promises of meta-agreement, thus, seem to be very enticing. If it is true that deliberation brings about single-peakedness by inducing meta-agreement, then deliberation should be able to overcome the instability and irrationality caused by cycles. Furthermore, contrary to substantive agreement, meta-agreement allows for pluralism within society, that is, different views about how the relevant political alternatives should be ordered. Meta-agreement, then, according to proponents of this account, is both more feasible and more compatible with pluralism than substantive agreement (Dryzek and Niemeyer, 2006; Farrar et al., 2010; List, 2002, 2007; List and Koenig-Archibugi, 2010).

In what follows, we will not challenge the assumption that avoiding cycles should be a prime concern of deliberative democracy, nor will we question the idea that deliberation can really be so powerful as to change people's preferences and views. What we want to challenge, instead, is the idea that meta-agreement, as described in the relevant literature, is a desirable and feasible model of how deliberation can overcome cycles and paradoxical results. More specifically, we will address the two central claims that make meta-agreement particularly valuable and appealing in the eyes of its advocates, namely, that it is less demanding than substantive agreement and that it is respectful of pluralism. We aim to show that neither claim is true. In order to establish this, we will need to reveal and analyze various assumptions that lie behind the standard account of how deliberation can bring about single-peakedness via *meta-agreement*. Without these assumptions, such an account is unable to explain why deliberation should have this effect. Once these assumptions are made explicit, though, it is easy to see that meta-agreement might indeed be even more taxing than substantive agreement as a condition for the viability of democracy. In other words, the meta-agreement model does not seem to be the right way to explain how deliberation can work as a pluralism-friendly way to reduce inconsistency in collective decisions.

Our argument will proceed as follows. In the next section, we will review the main notions involved in our discussion: the notion of single-peakedness and its fundamental role as a remedy for social choice's impossibility results; the notion of a *semantic* common dimension (as presented in the relevant literature) which becomes focal in deliberation, thus producing meta-agreement; and the threefold process by which deliberation is supposed to bring about single-peakedness by inducing meta-agreement. In the third section, we will dispel some ambiguities in the very definitions of the notions of 'semantic dimension' and 'meta-agreement'. In the fourth section, we will expose and analyze the hidden assumptions that underlie the process that supposedly makes deliberation bring about single-peakedness through meta-agreement. In the fifth and sixth sections, we will comment on these assumptions, showing that they reveal that meta-agreement is far more demanding and inimical to pluralism than its advocates seem to assume. The seventh section concludes our argument.

2. Single-peakedness, meta-agreement, and deliberation

2.1. The challenge of social choice theory

In order to appreciate fully the relevance of the notion of meta-agreement in the debate on the normative theory of democracy, let us remind ourselves of the nature and impact of the problems that such a notion is supposed to address, notably the problem of *cycles* as a possible result of voting according to democratic procedures. This will also help us recall the simple formal notions that will be used later in our discussion of meta-agreement.

Let us say, following a consolidated definition, a preference ordering $>_i$ of an individual i is a *complete*, *transitive*, and *irreflexive* relation on the set X of alternatives. Completeness means that for every pair of distinct options a and b, individuals are able to provide a ranking such that a is better than b or b is better than a; transitivity means that

if a is considered better than b and b better than c, then a should be considered better than c; irreflexivity means that a is not better than a. Completeness, transitivity, and irreflexivity define conditions for the *rationality* of individual preference orderings. A *profile* of preference orderings is a list $>_1, \ldots, >_n$ that specifies, for each individual, her preference ordering. The aggregation of individual preferences in this model is then interpreted as a function, called a *social welfare function*, which takes preference profiles and returns a unique 'social' preference ordering. Given the definition of individual rationality just set out, Condorcet's paradox famously shows that the same rationality constraints cannot hold for a collective, or social, outcome that is obtained by pairwise comparisons of alternatives by majority voting.

Consider the following example, in which three individuals (1, 2, and 3) rank three alternative options in the following ways.

Individual 1. $a > b > c$.
Individual 2. $b > c > a$.
Individual 3. $c > a > b$.

If we aggregate those preference orderings by pairwise comparisons, we obtain that a wins over b because 1 and 3 prefer a over b and b wins over c because 1 and 2 rank b over c. Then, by transitivity, we also get $a > c$; however, we also have that c wins over a, according to the preferences of 2 and 3. So $c > a$ and $a > c$, which implies $c > c$, against the irreflexivity assumption. Moreover, the collective ranking we obtain is *cyclical* ($a > b > c > a$); among other things, this amounts to the disregarding of all information coming from the very individual preferences that our procedure was supposed to aggregate.

The well-known theorem proved by Kenneth Arrow (1963) shows that this kind of unwelcome result does not depend on the aggregation procedure, namely, pairwise comparisons; rather, this is a more challenging phenomenon because the same paradoxical outcomes may occur with *any* voting procedure that satisfies certain intuitively appealing normative constraints, which should be essential to fair and democratic decision-making.[3]

These results constitute an obvious problem for normative democratic theory. Cycles not only cause arbitrariness, inefficient decisions, and political instability, but also expose democratic decisions to manipulation. William Riker, in his *Liberalism Against Populism* (1982), provocatively argues that this means that we should abandon any hope that democracy can be an efficient way to promote the common good; its value may at best consist of the chance it offers for nonviolent (and often casual) shifts in political power. Given these dismal conclusions, it is understandable that much effort has been put into devising responses to the challenge posed by these findings of social choice theory. Some theorists have questioned the actual occurrence of cycles (see, for example, Mackie, 2003; Regenwetter et al., 2006); others have claimed that cycles are not necessarily detrimental to democracy (Miller, 1983). Some advocates of deliberative democracy have insisted that deliberation can promote consensus, whereby cycles become impossible (Sunstein, 1988). An important reaction, which is also the one that most interests us here, has come from advocates of deliberative democracy who have argued that deliberation can overcome these paradoxical results by producing 'meta-agreement', while still retaining

disagreement about how to rank the options on the agenda. In the next paragraphs, we will see how they have tried to couch this reply in the language of social choice theory.[4]

2.2. The geometrical notion of a 'dimension'

One of the procedural constraints that is responsible for Condorcet's cycles is *universal domain*. This condition states that the aggregation procedure should be open to any possible profile of preference orderings as its input. The intuitive normative justification of this condition is that there should not be any constraints on the preference orderings that individuals are allowed to have. However, assuming any possible preference profile leads, together with the other Arrovian conditions, to the paradoxical outcomes we have just sketched. A way to circumvent Arrow's impossibility result and prevent cycles from occurring is thus to find a way to restrict the domain of preference orderings.

A trivial case is the one in which all the individual preference orderings are identical. In the literature we are discussing, this phenomenon is called 'substantive agreement'. If people are unanimous about the ordering of the available alternatives, then applying majority rule will not cause cycles to occur; in this case, the social ordering will simply coincide with the identical individual orderings.

However, this solution has been discarded by the literature we are discussing. Indeed, perfect unanimity in preference orderings appears to be too rare a condition to be worth considering as a solution to Arrovian paradoxes. Furthermore, it has been argued that it is unreasonably demanding because it requires every individual to rank the alternatives identically, therefore suppressing every possible expression of pluralism (Dryzek and Niemeyer, 2006). A natural question to ask is whether there are less drastic restrictions of the domain of individual orderings that can still guarantee, as substantive agreement does, that no cycles will occur.

Among the various solutions proposed in the literature, the one that is relevant for our discussion was suggested by Duncan Black (1958). The formal content of Black's condition, known as *single-peakedness*, can be presented as follows. A *geometric dimension* is a linear order $>_\Omega$ of the alternatives: we say that z is between x and y according to $>_\Omega$ if $x >_\Omega z >_\Omega y$ or $y >_\Omega z >_\Omega x$. Given an individual preference ordering $>_i$, we denote $peak_i$ as i's best option on the geometric dimension. A profile of preference orderings is *single-peaked* if, and only if, there exists a single dimension $>_\Omega$ such that for every individual i, given an alternative $y \neq peak_i$, i prefers any option between $peak_i$ and y to y. Intuitively, the condition states that there exists a single shared ordering of the alternatives ($>_\Omega$) such that all individual preferences are compatible with it in the following sense: first individuals choose their best option on $>_\Omega$, then they order the alternatives according to the proximity of such alternatives to their best option with respect to $>_\Omega$.

For example, assume that voters agree on the ranking of five candidates along the left–right dimension, say, $a >_\Omega b >_\Omega c >_\Omega d >_\Omega e$, where a is the leftmost. We can represent this linear ordering as an oriented axis (see Figure 1). Individuals choose their preferred candidate $peak_i$ on the left–right dimension; then single-peakedness entails that individuals have preferences of the following types. They may prefer the leftmost candidate, then the second leftmost, and so on ($a > b > c > d > e$, which is represented in Figure 1 as a solid line); in this case, the individual ranking coincides with the dimension.

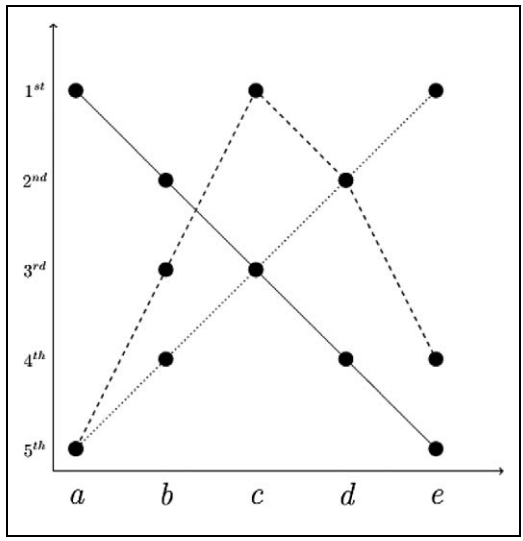

Figure 1.

They may prefer the rightmost candidate and order the others accordingly ($e > d > c > b > a$, which is represented in Figure 1 as a dotted line). Moreover, they may prefer a candidate positioned at the center and rank the others according to the distance of those others from that center along the left–right order ($c > d > b > e > a$, which is shown as a dashed line in Figure 1). However, single-peakedness is violated, for example, by a preference ordering such as $a > b > d > c > e$ (see Figure 2); in this case, the peak is a, the leftmost candidate, whereas the individual ranks $d > c$ even if c is closer to the best option on the left–right ordering.

Note that if we consider any single individual in isolation, single-peakedness is trivially satisfied because we can always choose as the geometrical dimension $>_\Omega$ a linear order that is identical to that individual's preference ordering. In this sense, every preference ordering is compatible with the dimension defined by the ranking of the alternatives provided by that very preference ordering: so every preference ordering can be considered a formal dimension,[5] the dimension according to which an individual orders the alternatives.

Moreover, if the preference ordering of an individual is compatible with a dimension $>_\Omega$, then it will also be compatible with the *opposite* dimension (call it $>_\Omega$') obtained by reversing the order $>_\Omega$. For example, if a preference ordering is compatible with the left–right dimension, it will also be compatible with the right–left dimension.

In the case of several individuals, the notion of single-peakedness becomes more demanding. The definition states that there must be *a single dimension* that is compatible with all the individual preference orderings. Still, single-peakedness is compatible with very different ways of ordering the alternatives by individuals, as the previous example shows. In this case, too, if a profile is compatible with a certain dimension $>_\Omega$, then it will also be compatible with the opposite dimension $>_\Omega$'.[6]

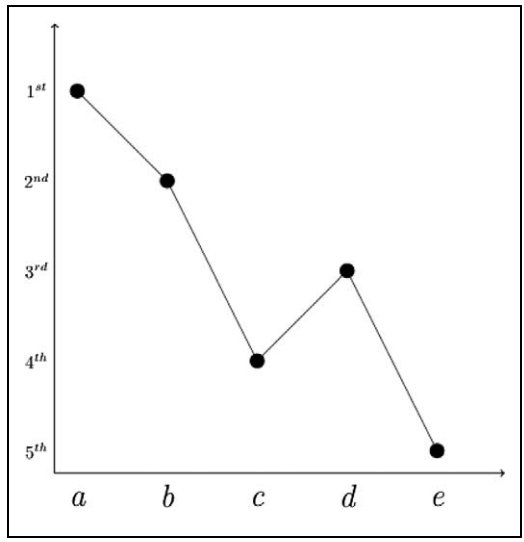

Figure 2.

Given a single-peaked profile, *Black's theorem* states that there will always be a winner:[7] on those restricted preference profiles, there are procedures that meet all the Arrovian procedural desiderata and still issue a noncyclical result.

Single-peakedness has understandably been considered a very promising property by those engaged in the project of circumventing the daunting impossibility results of social choice theory. Its main virtue, which has made it more popular than other solutions to Arrovian cycles, is that it seems to offer a *principled* way to restrict the domain of individual preference orderings, that is, it is a restriction that conforms to an independent rationale which is different from the mere need to avoid cycles. This feature of single-peakedness also appears to be particularly congenial to its treatment in the context of the theory of deliberative democracy, through its seemingly natural association with the *semantic* notion of a 'dimension', to which we will now turn.

2.3. The semantic notion of a 'dimension'

So far we have discussed the property of single-peakedness by using the notion of 'dimension' in its *formal* or *geometrical* sense. It is important to stress that the linear ordering $>_\Omega$ has no particular *meaning*; it is just a way to order the alternatives that is compatible with certain individual rankings. As Dryzek and List (2004) put it, a formal dimension 'does not carry any particular interpretation'.

Another notion of 'dimension' has been discussed in the literature:[8] the notion of *issue dimension* or *semantic dimension*. A dimension in its *semantic* sense is 'an attribute in terms of which the decision alternatives are conceptualized by decision makers' (Dryzek and List, 2004), a way to 'conceptualize the decision problem at stake' (List, 2007).

For example, I might prefer a, a ban on religious schools, to b, the establishment of private religious schools, to c, public funding of religious schools (that is, $a > b > c$). I may conceptualize the rationale of my preference ordering by referring to a semantic dimension according to which the alternatives at stake can be ranked, which in this case might be, for example, the 'availability of religious education': the three options are ordered and conceptualized as making religious education available to different degrees. My opponents in the debate might order the three options in the opposite way (that is, $c > b > a$), because they believe that more religious education is better than less. But notwithstanding our opposition on how to order the available options, we are referring to the same *semantic dimension* in order to conceptualize what is at stake in this decision.

The notion of a common issue or semantic dimension is central for the definition of *meta-agreement*: 'two or more individuals are in meta-agreement', according to List's definition (2007), 'to the extent that they agree on a common issue dimension in terms of which a given decision problem is to be conceptualized – and in terms of which preferences are to be rationalized. They may reach perfect meta-agreement while at the same time disagreeing substantively on what the most preferred option is.'[9]

In other words, meta-agreement is agreement on a common semantic dimension through which the parties conceptualize the differences and relations of the available alternatives. We are now ready to present the main target of our discussion, that is, the argument concerning the relationship between deliberation, meta-agreement, and single-peakedness.

2.4. Meta-agreement and single-peakedness

The main idea behind the claim that deliberation can bring about single-peakedness via meta-agreement is that deliberation induces people to justify their preference orderings in terms of reasons that can be understood by others. This process should naturally lead them to appeal to common semantic dimensions through which they frame the choices at stake. According to Dryzek and List (2003), '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'. In turn, convergence on the same semantic dimensions may imply single-peakedness: 'Single-peakedness may be an *implication* of meta-agreement. If the individuals agree on a common issue dimension and rationalize their preferences in terms of that dimension, then the resulting combination of individual preferences will satisfy single-peakedness, provided the common (semantic) issue dimension translates into a common (geometrical) structuring dimension' (List, 2007). The relation between convergence on a single semantic dimension and single-peakedness seems to rely on the fundamental intuition that once we have 'conceptualized' the different options according to a single semantic dimension (for instance, the various degrees of 'availability of religious education' in the example of religious schools), public deliberation will also lead the parties to align the options along one and the same geometrical dimension corresponding to the semantic dimension, and the parties will then choose their preferred options along that geometrical dimension, thus generating single-peaked preference orderings.

According to List (2002, 2007) and Dryzek and List (2003), the hypothesis that deliberation promotes single-peakedness via *meta-agreement* can thus be broken down into three related hypotheses, which we rephrase as follows.

D1. Deliberation leads people to identify a common *semantic* dimension in terms of which to conceptualize the decision problem.
D2. Deliberation leads people to agree on how the options are aligned from left to right on that issue dimension. So, deliberation leads people to determine which *geometric* dimension best represents the given semantic dimension.
D3. Given the semantic dimension and the geometric dimension, deliberation leads each individual to choose a most-preferred position (his peak) on that dimension, with decreasing preference as options that are increasingly distant from the most-preferred position.

Each of these three steps has a different status. The issue raised in D1, regarding which semantic dimension to choose, is considered to be a *normative* question (Dryzek and Niemeyer, 2006; List, 2007), as it requires individuals to agree on the common conceptualization of the choice in terms of some evaluative dimension.

Step D2 involves what Dryzek and List call a *factual* issue: once a relevant, shared semantic dimension has been identified, individuals have to agree on how the available alternatives satisfy the property expressed by the semantic dimension. For example, suppose that 'availability of religious education' is the relevant semantic dimension; it is a factual issue, according to this model, to determine which policy promotes such availability and to exactly what degree.

The third step (D3), according to the literature we are discussing, relates to an issue of *rationality*: once individuals agree that a given semantic dimension expresses the relevant terms of the decision at stake and once they agree on the corresponding geometric dimension, then they will identify their most-preferred option and they will order the other alternatives according to the distance of those alternatives from their best option with respect to the selected geometric dimension. In other words, their preferences will be *single-peaked*. Those who do not have single-peaked preferences at this point, according to this model, can be charged with irrationality. For example, imagine that the parties agree on the 'availability of religious education' as a shared semantic dimension and assume that there is agreement on the ranking of the policies according to that dimension, an individual who gives his preference to the option that best promotes the availability of religious education (this is his 'peak') on the (normative) ground that 'the more religious education, the better' would be charged with irrationality if he were to choose as his second-best preference the option that most restricts such availability, that is, the preference most distant from his peak according to the aforesaid dimension.

We can now see the appeal of this model. Meta-agreement gives content and plausibility to the idea that the domain of individual preference orderings can be restricted and disciplined so as to produce single-peaked configurations that are immune to cycles and, most importantly, that this can be done through a principled and noncoercive process guided only by reason and deliberation. Furthermore, as is claimed by its advocates, this process can occur while still leaving room for disagreement and pluralism, that is, a

variety of different ways of ranking the available alternatives on the agenda. This, it is claimed, not only makes meta-agreement less demanding than the requirement of substantive agreement (that is, of a perfect match between all individual preference orderings), but it shows due respect for difference and pluralism as essential components of democratic politics.[10]

3. The space between meta-agreement and single-peakedness

There is much ambiguity in the notion of meta-agreement as presented in the relevant literature, and it is important to explore this in more detail. A major source of ambiguity lies in the fact that the definition of meta-agreement provided in the previous section (that is, agreement on the relevant semantic dimension), in fact, corresponds only to D1, even though meta-agreement is often assumed (rightly, in our view) to be constituted by the combination of the three hypotheses D1, D2, and D3 (see, for example, List, 2007).

The relevance of this ambiguity emerges once we inquire about the exact relation between single-peakedness and agreement on some relevant semantic dimension, as described in D1. Jonathan Aldred (2004) has argued that meta-agreement (in the narrow sense of D1) is not sufficient to obtain single-peakedness for two key reasons: (1) there could be semantic dimensions that do not correspond to linear orders, such as some qualitative semantic dimension that does not allow for a linear ranking of the options (for example, 'hair color'); (2) there could be a semantic dimension corresponding to a given formal dimension, but preferences might not be aligned according to that dimension (for example, 'hair length' if I prefer extremely short and extremely long hair to medium-length hair). We can rephrase Aldred's arguments by saying that these counterexamples show that D1 alone is not sufficient for single-peakedness; we also need D2 and D3.

In order to reply to the first counterexample, we need to assume that there can be factual agreement concerning the ranking corresponding to the shared semantic dimension. This includes both the assumption that a semantic dimension can be quantified and the assumption that individuals agree on the corresponding quantitative ranking. This aspect seems to be acknowledged by List when he remarks that meta-agreement promotes single-peakedness '*provided the common (semantic) issue dimension translates into a common (geometrical) structuring dimension*' (2007, emphasis added).

The second counterexample provides an argument for the necessity of D3, the meta-agreement on rationality. The objection raised by the counterexample can be answered by pointing out that when cases like this occur, what is happening is that the *actual* dimension referred to is different from the one that is supposed to be agreed upon. For example, the individual who prefers extremely long and extremely short to medium-length hair is not really looking at length as the relevant dimension, but at the degree to which a hairdo makes one look young or 'cool'.[11] In order to have single-peakedness, we need to assume that once a relevant dimension is agreed upon, people will be committed to it and order the alternatives accordingly, rather than according to some other hidden dimension; this is implied by the 'rationality' requirement described in D3. Moreover, we need to assume that individuals will order the alternatives according to the rationality conditions on preference orderings (see Subsection 2.1.), which is again implied by the rationality assumption in D3.

Thus, Aldred's objections help us see that D2 and D3 are necessary in order for meta-agreement to induce single-peakedness. In other words, the narrower notion of meta-agreement we quoted in the last section, which actually corresponds to D1 only, is not sufficient to imply single-peakedness. This seems to be acknowledged by List (2007) when he remarks that 'The combination [of D1, D2, and D3] is in essence the hypothesis that deliberation brings about meta-agreement, which then surfaces in the form of single-peakedness.'

We propose, then, to use the following terminology, which recalls the status of the three hypotheses, D1, D2, and D3, according to Dryzek and List. We shall say that D1 concerns *normative* meta-agreement, D2 a *factual* meta-agreement, and D3 a *rationality* meta-agreement.

4. How deliberation induces single-peakedness via meta-agreement

We have recalled Aldred's remarks in order to dispel some fundamental ambiguities in the argument about the role of meta-agreement in bringing about single-peakedness. Unlike Aldred, however, we are not interested in challenging the conditional hypothesis that if deliberation satisfies D1, D2, and D3, then it promotes single-peakedness. What is more interesting from our point of view is to assume that the hypothesis is correct and can be made more precise along the lines just considered, and then to investigate *how* and *at what price* deliberation can bring about single-peakedness via the various forms of meta-agreement just defined.

In fact, although the clarifications presented in the last section help to establish the main structure and workings of the hypothesis that deliberation brings about single-peakedness via meta-agreement, we believe that much still needs to be made explicit in order to see what meta-agreement really implies. In this section, we will unpack the three hypotheses, D1, D2, and D3, showing what is required to achieve each partial goal within the general argument about the correlation between deliberation and single-peakedness.

4.1. Semantic dimensions: concepts or words?

In the formulation of the D1-D2-D3 hypothesis, the notion of a semantic dimension refers to some sort of psychological process of forming one's preferences. Usually, in the relevant literature this is referred to as a 'conceptualization' of the decision problem.

However, it is not clear how agents should *exchange* different conceptualizations or should communicate such conceptualizations to each other. Thus, the very idea in D1 of converging toward the same semantic dimension via discussion and deliberation is obscure. Evidently, we need to assume that this happens through some form of *verbalization* or public exchange of utterances in a deliberative context.

Therefore, we need to distinguish between a semantic dimension in its *psychological* sense and a semantic dimension in its *verbal* sense (hereafter, the latter is referred to as the 'verbal dimension'), which is a sentence uttered during the discussion. The underlying hypothesis is that verbalizations may constitute an apt way to access semantic dimensions understood as conceptualizations. This step is needed in order to introduce

into the analysis of the notion of meta-agreement the essential role of language in discursive situations, which has somehow been surprisingly overlooked in previous analyses of meta-agreement.

In order to account for the relation between the semantic dimensions and individual preference orderings, we should also assume that the 'verbal' dimensions are actually used publicly to *justify* individual preferences to others. The fact that semantic dimensions are used in public justifications means that parties that intend to use them are *committed* to those sentences. After all, this is how deliberation is supposed to *do* something: talk is not cheap and exchanging reasons in public has the power of transforming people's preferences. In Dryzek's words, 'the whole point of deliberation is to "restrict domain" as the language of social choice theory would have it, because deliberation in inducing reflection on preferences and requiring that they be defended publicly eliminates preference orderings that cannot be so defended' (2000: 43).

4.2. What D1 (normative meta-agreement) implies

Once we see that the agreement in D1 is a discursive form of agreement on the relevant *verbal* semantic dimension, we can identify a further assumption that lies behind the hypothesis that deliberation can promote D1. Which semantic dimension becomes focal, as we saw, is described as a *normative* question because at this level the words verbalizing semantic dimensions (such as 'fair', 'left', 'right', 'productive', and 'liberal') have a normative weight. This does not necessarily imply that the predicates involved are *essentially* normative; rather, the underlying idea is that if the predicates referred to in D1 must play a role in the public *justification* of claims, they must be *loaded* with normative weight in the course of the discussion. So, for example, when an option x is publicly defended as better than y on the grounds that it is more 'efficient', what is happening is that the speaker is appealing to efficiency as the relevant normative dimension according to which the available options need to be assessed.

Since the predicates (P) referred to in D1 are normatively loaded, saying that an option is 'more P' than another is not, of course, a neutral statement. For example, assume that 'productivity' becomes the focal verbal dimension. Those who wish to choose the most productive option will then justify their preferences in terms of productivity, by uttering, for example, statements such as 'We should choose x because it is the most productive option.' It is implausible that someone who wants to challenge their preference ordering and choose the least productive option will justify her own preference by using sentences such as 'We should choose the least productive option' or 'We should choose the worst option in terms of productivity.' Again, think about the normative load in a phrase such as 'the least liberal option'. It is very implausible that someone might use it to describe his most preferred alternative; he will, rather, refer to *another predicate*, which is perceived as inversely connected to 'liberal' and moves along the very opposite scale.[12]

These examples show that according to the meta-agreement model semantic dimensions need to be *polar*, and their normative poles are usually expressed by two normatively loaded predicates. Otherwise, individuals who are assumed to be willing to converge on the same verbal dimension P, but, for example, prefer the pole 'less P'

would lack their own justification words, or maybe even their own conceptualizations, and so their discursive power would be lost.[13]

Consider, indeed, what would happen (according to the rationale of the process described in D1-D2-D3) if the parties converged on a single normatively loaded predicate *P*, rather than a pair of polarized predicates: were this to occur, the upshot would be *substantive* agreement, rather than mere *meta-agreement*. Assume D1 selects just one normative predicate *P*, on whose normative salience all the individuals agree. Then, if (as the advocates of meta-agreement assume) D2 holds, the meaning of *P* will be fixed through deliberation; at this point, D3 will entail substantive consensus rather than single-peakedness because every individual will choose the same peak (namely, the 'most *P*' option) as their best, and will order all the other options accordingly, producing identical preference orderings.

Thus, if (according to the hypothesis that is central to this account of how deliberation saves democracy from cycles) deliberation must provide *meta-agreement* without *substantive agreement* (that is, *single-peaked*, but *not unanimous* profiles), then the agreement required at the meta-level is not and cannot be just agreement on *one single* normatively loaded predicate. It must be an agreement on a *pair of normative poles* as focal.[14]

To this we should also add a further assumption: there must be agreement on the *belief that these two values provide grounds for opposite justifications*. To see this, consider the following example. Assume that the normative pair that becomes focal is *productivity* and *fairness*. Agreeing to think of these as opposing values means that someone who wants to claim that *x* is better than *y* on the ground of 'productivity' may be *challenged* by some others who would claim that *y* is better than *x* on the ground of 'fairness': these parties recognize themselves as challenging each other on a common ground which is constituted by the agreement on the fact that arguments involving productivity are opposed by arguments involving fairness.[15] Those who agree that this pair is the focal one, but prefer a median option between fairness and productivity have then the possibility of justifying their preference in terms of a *balance* between productivity and fairness.[16]

In sum, the normative meta-agreement as described in D1 requires an agreement on two values, and it requires agreement on the belief that those values are *opposite*. This is the only way in which in this model we can account for discursively justified disagreement on the ranking of the alternatives while still having a unique 'semantic' dimension.

4.3. What D2 (factual meta-agreement) implies

D1 implies that deliberation can provide the verbalization of a focal pair of values, such as productivity and fairness. Step D2 requires that deliberation promotes a form of *factual* meta-agreement, that is, the mapping from a *verbal* dimension to a linear order defining a geometric dimension.

D2 can be considered as a step toward fixing the *precise meaning* of the pair of normative predicates. The idea here is that when individuals agree on statements like '*x* is more productive than *y*' they have to agree on the extension of the relation associated with this comparative use of the predicate 'productive'.

Furthermore, in order to imply single-peakedness, deliberation needs to be able to promote agreement on the meaning of the opposing predicates, say, P and P', so as to make them denote opposing *formal* dimensions. Therefore, a hidden assumption behind D2, which mirrors the one behind D1, is that, besides the factual agreement on how to rank the options in terms of the relevant predicates, the following must hold:

O. xPy if, and only if, $yP'x$

O implies that one predicate actually behaves as the negation of the other. Otherwise, individuals using P and P' might not refer to the same geometrical dimension.

Let us see why by means of an example. Imagine that the productivity-fairness pair becomes focal and that both the ranking corresponding to productivity and the ranking corresponding to fairness are determined. If O does not hold, then individuals might verbally agree on the fact that arguments appealing to productivity are challenged by arguments appealing to fairness, but then, when they actually have to order their options according to the formal dimension they think corresponds to their semantic dimension, come up with a non-single-peaked profile. The reason for this potential outcome is that the formal dimension corresponding to productiveness might not be the opposite of the one corresponding to fairness.

As a final remark on D2, let us point to a further assumption, which is in fact discussed in the relevant literature, but whose import might be reassessed in the light of what we have said so far. The assumption at issue is that for single-peakedness to be induced by the sharing of a semantic dimension, the parties are required to attach *exactly the same meaning* to the pair of values so as to make them denote opposite dimensions; this requires in its turn that they cannot disagree on the *facts* about the policy alternatives that are relevant for positioning the options along the shared dimension. In other words, this factual agreement implies that people share the same beliefs about the outcomes to be expected by each of the policies to be ranked.

4.4. What D3 (rationality meta-agreement) implies

Most of the hidden assumptions behind D3 have already been made explicit in D1 and D2. What is still missing is a clarification of the rationality assumptions that make D3 work. According to its formulation, D3 would be violated by someone who agreed on the relevant verbal dimension, agreed on the meaning and on the opposition of the two values involved, but still did not have single-peaked preferences with respect to the formal dimension obtained in D2. D3 implies that such an individual could be accused of being irrational. This charge may be interpreted in two ways. It might be interpreted as suggesting that those who do not display single-peaked preferences in these circumstances are violating *discursive rationality*, because in D1 and D2 they have *committed* to a given dimension as the relevant one and to the relevant beliefs about the facts at issue. Alternatively, on the assumption that no such violation occurs and these parties genuinely order their preferences according to a different rationale, it might be interpreted as suggesting that this is a plain violation of basic *instrumental rationality*.

However, it is important to make explicit that in order to consider D3 a matter of rationality, we need to assume a *sequential* structure in the meta-agreement argument. This means that when we get to D3, D1 and D2 must *already* hold: the normative predicates must be fixed, their meaning decided, and all the individuals must have committed to them. Otherwise, non-single-peaked preferences could be a matter of disagreement of any kind, rather than failures by the parties to stick to their discursive commitments or to fulfill instrumental rationality.

5. The demandingness of meta-agreement

Unfolding and making explicit all the assumptions behind the claim that deliberation can bring about single-peakedness via meta-agreement serves not only to improve analytical clarity and completeness. We want to argue here that it also reveals that meta-agreement might be much less appealing, as a model of how deliberation can produce single-peakedness, than its proponents seem to assume. More specifically, we want to argue that once we have made explicit all these assumptions, we realize the untenability of two claims that are central to this account of the merits of deliberation: (1) the claim that meta-agreement is less demanding and easier to achieve than substantive agreement and (2) the claim that meta-agreement better accommodates pluralism than substantive agreement.

Let us consider first the claim that meta-agreement is less demanding than substantive agreement. The apparent plausibility of this claim derives from the fact that there are many more combinations of individual orderings that satisfy single-peakedness (which is associated, according to the D1-D2-D3 hypothesis, with meta-agreement) than combinations that satisfy a perfect coincidence of individual preference orderings, that is, substantive agreement. This should appear obvious from the fact that the set of unanimous profiles is a subclass of the set of single-peaked profiles.

However, we should keep in mind that meta-agreement and single-peakedness are two different things. As we have seen, the latter is simply a geometric property satisfied by many combinations of individual orderings. Meta-agreement, instead, implies the full assumptions and conditions made explicit in the preceding sections. Once we look at these, we realize that it can be a much more difficult goal to achieve than mere single-peakedness.

First of all and very straightforwardly, we should note that meta-agreement requires *unanimity*, although of a different kind than the unanimity of preference orderings required by substantive agreement: it requires unanimity on how to 'conceptualize' the relations between the different options. The demandingness of this requirement has been pointed out by critics, and the proponents of meta-agreement themselves acknowledge it (Dryzek and List, 2003: 17). However, they argue that, in fact, if we think of how public deliberation works, the assumption that deliberation leads to a single shared semantic dimension appears to be plausible. Given the restrictions public justification imposes on the kind of reasons that can be used, people are likely to converge on the same way of framing the decision at stake, and this is conducive to sharing the same semantic dimension. The plain idea behind this assumption is that although there are many ways to

'conceptualize' the choices at stake, only some of them can support arguments that are based on 'generalizable interests' such that the other parties can accept them as reasonable.

Now, assume for a moment that public deliberation can actually bring about this result and that in a deliberative setting, due to the need to find reasons acceptable to others, people converge on using the same semantic dimension to describe the issue at stake. The problem is that even if people agree on a single dimension, we are still very far from having achieved single-peakedness. As we have shown, for single-peakedness to obtain from convergence on the same dimension as it is verbalized in the public arena, two types of unanimity must hold: (1) the participants must attach one and the same *meaning* to the words employed and acknowledge their *polar opposition* and (2) they cannot disagree on the *facts* about the policy alternatives that are relevant for positioning the options along the shared dimension. Otherwise, there is no guarantee that the use of the same words will translate into a shared geometric dimension along which to place the different options. As we have seen, if, for example, the tradeoff 'fairness-productivity' is the relevant dimension, the parties need to agree not only on the exact meanings of the words 'productivity' and 'fairness', but also on the properties of the various options that make it possible to establish that a given option *a* should be ranked as *more productive* than option *b*. This is arguably much more demanding than mere convergence on using the same political watchwords. Many political debates, indeed, revolve on factual matters about the best means to reach ends and goals that everybody shares. In other words, in many cases the locus of disagreement is not at the level of D1, as the meta-agreement model seems to assume, but at the level of D2: people agree that a certain value or goal should guide the decision-making process, but deeply disagree about how the different options on the agenda will serve such an end. The relative easiness with which the model of meta-agreement we are discussing assumes that D2 is not problematic may be seen as evidence of its neglect of the actual workings of deliberation.

To summarize, meta-agreement can appear to be a deceptively undemanding condition because it is easily confused with single-peakedness. Furthermore, even once it is clearly distinguished from single-peakedness, we can still be deceived into thinking that meta-agreement is easy to achieve because we tend to focus on agreement on the *words* used to utter reasons, while forgetting that for this agreement to bring about single-peakedness, an exact agreement at the semantic and factual levels is required.

To this it might be objected that, although meta-agreement might appear demanding, we should remember that the claim about its undemandingness, as it is phrased by its proponents, usually comes in comparative terms, as an amendment to the deeply unrealistic and 'naive'[17] idea that deliberation can bring about substantive agreement.

However, even in such comparative terms, this claim turns out to be controversial. First of all, we should notice that it might seem obvious that substantive agreement is more demanding than meta-agreement if we assume that in order to reach substantive agreement through deliberation we necessarily have to go through meta-agreement, that is, a consensus on the rationale by which identical individual preference orderings are built. However, there is no reason to think of this kind of consensus as the only way in which substantive agreement could be reached through deliberation. Deliberation is the exchange of principled claims in public, by offering reasons that the other parties may reasonably accept.[18] This process might lead to substantive agreement without

meta-agreement. For example, through deliberation the parties might become convinced that a certain ordering of the options is the best one for a number of different reasons and along different dimensions (perhaps summarized by one and the same polysemic watchword, such as 'freedom', 'fairness', or 'equality'), among which they are able find their most congenial ones. Alternatively, the parties might become more willing to accept one and the same ordering of their preferences once they have had the chance to state their distinctive position on its rationale in public, thus making explicit that their acceptance of the ordering does not express complicity with certain reasons for supporting it, which they do not share.[19] For example, pro-abortion feminists could give their preference to the ban of certain abortion practices over their legalization once they are given the opportunity to make it explicit that they do it because they believe that such practices are unsafe, distancing themselves from the anti-abortionist stance of other parties advocating the ban. Then again, consensus on a given ranking could be reached because in the course of deliberation some or all participants have modified the meaning and implications of their own way of 'conceptualizing' the options on the agenda, without converging on one and the same 'conceptualization'.[20] Needless to say, the same mechanisms could make deliberation bring about single-peakedness, rather than substantive agreement, without requiring meta-agreement.

This amounts to saying that the notion of deliberation assumed by the meta-agreement model is too limited to account for all the possible ways in which deliberation can have transformative effects on people's judgments and eventually bring about consensus. Among its other shortcomings, this representation forces a single precise meaning to be attached to individuals' claims that remains constant through deliberation; thus, it cannot account for the polysemic character of parties' watchwords nor for the complex dynamics of meanings in the political discourse.[21]

Second, even if 'substantive agreement' were necessarily taken to imply meta-agreement, it would not be obvious why (given meta-agreement) substantive agreement should be more demanding than substantive disagreement. As shown in Subsection 4.2, the case in which deliberation brings about substantive agreement through meta-agreement occurs when people agree on the same nonpolarized dimension, for example, they all agree that security is the most important dimension, and then proceed to rank the available alternatives accordingly. This case is structurally identical to the one (occurring when there is meta-agreement *without* substantive agreement) in which the dimension is polarized, the only difference being the number of 'poles' involved. Whether people will actually find it easier to converge on the same dimension when it is polarized, rather than when it is not, will presumably vary according to the circumstances and decisions at stake and the motivations of the parties.

The fact that from a structural point of view the two cases are identical (but for the number of normative poles involved) tells us something important. Those who feel that the full-blown *consensus* sought by some advocates of public deliberation is too demanding should be aware of the fact that what is most demanding about it is not the mere 'substantive agreement' it entails, but the assumption that such agreement can be based on the sharing of an underlying meta-agreement on the values at stake and the way they are instantiated by the different options on the agenda. But this is exactly what meta-agreement on a polarized dimension implies. Those who are looking at

meta-agreement as a way to release deliberation from the demandingness of consensus, in other words, are looking in the wrong place.

6. Meta-agreement and respect for pluralism

6.1. What pluralism is about

Let us consider now the claim that meta-agreement is more hospitable to pluralism than substantive agreement. Again, this claim borrows its apparent plausibility from the fact that single-peakedness, which is assumed to be induced by meta-agreement, is compatible with pluralism in a way that substantive agreement obviously is not. Substantive agreement implies that everybody is ranking the available options in exactly the same way, while single-peakedness is compatible with a wide range of combinations of individual rankings. If we identify pluralism with the existence of different individual rankings, then single-peakedness is obviously more hospitable to pluralism than substantive agreement. However, as before, we should remember that meta-agreement implies more than single-peakedness, namely, agreement on the same semantic dimension and the fulfillment of the conditions that cause this kind of agreement to translate into single-peakedness. Once we bear this in mind, we realize that maybe we should reconsider our judgment.

The relevant point here is that pluralism is not just about which preference orderings are adopted and advocated by citizens. This is an important dimension of pluralism, of course, but it is very far from exhausting the full meaning of the notion. Another fundamental dimension of pluralism, which is very dear to the liberal tradition, is pluralism about worldviews and the ways in which people conceptualize and describe their social space. This is exactly what is made impossible by meta-agreement. When meta-agreement occurs, all participants in deliberation come to conceptualize the options open to them in exactly same way. As we have seen, if this identical conceptualization is to induce single-peakedness, it must imply not only that they use the same *words* to describe the different options, but also that they attach *one and the same meaning* to those words and that they share *identical descriptions* of all the relevant facts.

This picture is not only deeply unrealistic, but also oblivious to what makes people disagree about politics. Political battles are fought not only over policies and agendas or about which order of priority should be chosen for collective goals and ends, but also and very importantly over the meaning of fundamental words (‘liberty’, ‘equality’, ‘efficiency’, ‘solidarity’, and many others) which are essential in conceptualizing and describing policy issues. Recognizing pluralism in the realm of politics means respecting this diversity and the importance of disagreement about the meaning of political watchwords. Think, for example, of the heated debates over the meaning of ‘sustainability’ in many countries worldwide. To give another example, think of the debates on the meaning of ‘equality’ that take place every time major pieces of social legislation concerning disadvantaged minority groups are passed. ‘Gender equality’, for example, means very different things depending on whether you are a differentialist feminist, a libertarian, or a conservative. In each case, this watchword is associated with a full vision of what a just society would look like, but each of which is based on very different normative,

sociological, and anthropological assumptions. Accordingly, the policy options on the agenda and their merits will receive very different descriptions from the different points of view. The very same policy (for example, mandatory maternity leave or the setting of different retirement ages for men and women) can look absolutely egalitarian or deeply inegalitarian, depending on which idea of gender equality we are endorsing. Nevertheless, all these perspectives are appealing to the same watchword; insisting that we should use 'equality' just for one of these visions, while the others should find some other words for indicating the values they cherish, would amount to a serious misrecognition of what pluralism implies.

Some advocates of meta-agreement suggest, reassuringly, that the relevant dimensions can be 'provisional and contestable' (Dryzek and Niemeyer, 2006: 647). But this is not enough. Pluralism at this fundamental level is not just about the possibility of a change in a worldview that everybody shares, but also the coexistence of different worldviews *at the same time* within the same society, as a consequence of the richness and contestability of the political vocabulary.

Here again, we can compare meta-agreement and substantive agreement, once more to the advantage of the latter. Substantive agreement is, of course, incompatible with pluralism about the way people rank the available policy options, but is surely compatible with there being pluralism in the fundamental sense just explored. Furthermore, as we have suggested, there might be mechanisms by which deliberation could bring about substantive agreement without requiring a fundamental agreement on how the different policies should be 'conceptualized' and described. To go back to the example of 'gender equality', we might think that given certain sets of options people may converge on the same rankings and publicly endorse their rankings by appealing to equality, while still retaining very different views of the full picture. Support for gender quotas, in many countries, has been given in the name of equality by many different strands of feminism and of liberal thought; each party often making clear that this was just one piece of a broader picture of a society of equals; something very different indeed from what the other parties meant by gender equality.

6.2. The polarization of dissent

Let us now consider another effect of meta-agreement on pluralism. As we have seen, to the extent that pluralism in the sense of different preference orderings is compatible with meta-agreement, then it must present itself in the form of a direct *opposition* between values, that is, in the form of a *polarization* (see Subsections 4.2 and 4.3). In many cases, of course, there will be a median position, perhaps taken by most people, which will represent a compromise, so to speak, between the two opposite values. Still, those at the two poles will be explicitly conceived of as directly opposed, and the median position itself will not represent an additional, distinct point of view, but will merely reflect a tradeoff between the two opposed values.

The upshot of these remarks is again that meta-agreement is less hospitable to pluralism than it may appear at first sight because in order to produce single-peakedness, meta-agreement must transform mere difference into opposition, not only by making explicit that the policy alternatives are positioned along a polarized axis, but also (and

more dangerously) by fixing the meaning of the underlying political values in confrontational terms. The harm that is thus done to pluralism consists not only in shrinking the meaning of the relevant political watchwords, but also, in doing so, to make pluralism a much more conflictual and oppositional matter than it needs to be.

This result should come as no surprise. Indeed, the whole notion of meta-agreement seems to be modeled after the popular idea that single-peakedness can be reached by the self-positioning of political parties along a unidimensional (typically, 'left–right') ideological axis, which provides a semantic dimension with which to conceptualize electoral choices that easily translates into a single geometrical dimension (Downs, 1957; Hinich and Munger, 1997; Hinich and Pollard, 1981; Poole and Rosenthal, 1994, 1997).

However, it is to be noted that the unidimensional ideological divide between parties is *artificially simple* and *inherently oppositional*. It is artificially simple in the sense that whenever it is true that parties position themselves along a unidimensional axis (according to this interpretation of how party politics works), the corresponding semantic dimension ('left–right' or 'liberal–conservative') is purposely created in order to reduce to a single dimension the multiple value dimensions along which electoral choices can be made (Downs, 1957). It is inherently oppositional in the sense that it is conceived of as being in a context of political competition, in which the parties need to position themselves antagonistically with respect to each other. So, this kind of polarization might be fit to represent party pluralism,[22] but it is inappropriate for representing value pluralism or pluralism regarding conceptions of worldviews. These types of pluralism are not necessarily oppositional because values and principles are not easily reducible to trade-offs between opposite poles. Even when people's preference orderings happen to be single-peaked, it might be the case that the values and principles according to which they 'conceptualize' their choices are not conceived in these simple oppositional terms.

6.3. Dissent and irrationality

Finally, meta-agreement is inhospitable to pluralism because it puts those who conceptualize the issues at stake in a 'nonconformist' way (that is, those who violate single-peakedness) in the uncomfortable position of being labeled 'irrational'. To this it might be objected that, as we have seen, the charge of irrationality only comes after the parties have agreed on the relevant semantic dimension by means of which the choice among different options should be 'conceptualized' and after they have agreed on the relevant facts pertaining to the options on the agenda. At this point, it might be argued, there is really no explanation for a preference ordering that violates single-peakedness, other than some violation of the requirements of rationality, either genuinely or for strategic reasons.

But this assumption about the *sequential* nature of the D1-D2-D3 process, which we underlined as one of the most important implications of D3, is exactly what should make advocates of pluralism suspicious. We can find such an assumption plausible, and apparently harmless, only if we forget how public deliberation, and indeed any verbal interaction, works in reality. The fact that agreement has *actually* been reached on the salience of a given dimension, and that there is agreement on a precise meaning of that dimension, can be ascertained only when people *actually* rank their preferences accordingly. Until that moment, all we know is that people are using the same *words* in

attempting to explain the rationale of their preferences. Conversely, if someone does not rank his preferences according to the same geometric dimension as the other parties in a given circumstance (even when he has explicitly agreed on *naming* a given semantic dimension as relevant and apparently agrees about the facts of the available options), the most natural conclusion is that, in fact, there was no real agreement on the precise meaning of the word referring to the dimension at stake. The idea that at this point the nonconformist individual should be labeled irrational misconstrues what is really happening in this kind of case by representing disagreement and difference as symptomatic of an inability to abide by the standards of discursive or instrumental rationality. Although irrationality may sometimes occur, often what is really happening is that disagreement is emerging only at the point at which people actually rank their preferences. The D1-D2-D3 hypothesis leads us to forget this possibility by misleadingly representing the process through which the parties discuss and revise their preference orderings as a *sequence* of binding steps.

7. Conclusion

It is not unreasonable to hope that deliberation can bring about single-peakedness, thus reducing the occurrence of cycles and creating the conditions for nonarbitrary decisions and political stability. Indeed, some relevant empirical studies seem to confirm that this might actually be the case (for example, Farrar et al., 2010; List et al., 2006). However, if the preceding analysis is correct, then we should not hope for this to happen through meta-agreement. In order for meta-agreement to be the *cause* of single-peakedness, D1-D2-D3 is not enough; a number of further assumptions must hold, which we have made explicit and discussed in Section 4 of this article. Once these assumptions are fully specified, it becomes clear that not only is meta-agreement quite a demanding condition to fulfill, but that it is incompatible with the nature and significance of pluralism in a democratic polity: it restricts pluralism at the fundamental level at which political watchwords are given their meanings, it transforms difference into polarization, and it leads to the stigmatization of nonconformist points of views as mere irrationality. No feasible and desirable model of deliberation should produce these outcomes, and we should look at other ways of modeling deliberation and its taming effects on the potential irrationality of collective decision-making. If meta-agreement were the price to be paid to overcome the risk of irrational democratic decisions in the absence of single-peakedness, we might well consider that price too high.

Notes

We are very grateful to Enrico Biale, Ian Carter, Emanuela Ceva, Ulle Endriss, and Elijah Millgram for their written comments on earlier drafts of this article. We also wish to thank the two anonymous referees of *Politics, Philosophy and Economics* for their insightful remarks and suggestions.

1. For a few examples, often quoted in the critical literature, see Bohman (1998), Cohen (1989), Elster (1986), and Habermas (1996).
2. The terms ‘meta-agreement’ and ‘agreement at a meta-level’ are of recent coinage. The idea, though, was presented in detail in Miller (1992) and Knight and Johnson (1994). For more recent statements, see Dryzek and List (2003), Farrar et al. (2010), List (2002, 2007), List and

Koenig-Archibugi (2010), and List et al. (2006). Dryzek and Niemeyer (2006) do not use the term 'meta-agreement', but talk of 'preference meta-consensus'.

3. The conditions required for Arrow's theorem can be intuitively presented as follows.

Universal domain. The domain of the social welfare function includes all possible profiles of individual preference orderings.

Weak Pareto property. If every individual prefers *a* over *b*, then the social order should respect unanimity and rank *a* over *b*.

Independence of irrelevant alternatives. The social ranking between *a* and *b* should only depend on the individual rankings of *a* and *b*, with no regard for their relationship with any other alternative *c*.

Non-dictatorship. The social welfare function does not always return the preference ordering of a certain individual.

4. For some important discussions on the relation between social choice theory and deliberative democracy, see Cohen (1989), Dryzek (2000), Miller (1992), and Van Mill (1996).
5. This holds if we assume, as the literature we are considering does, that individual preferences are linear orders.
6. We will make use of this remark concerning a dimension and its opposite when we discuss how deliberation brings about single-peakedness.
7. More precisely, there will be a Condorcet winner, that is, an alternative that wins in every pairwise comparison.
8. See, in particular, Dryzek and List (2003), Dryzek and Niemeyer (2006), and List (2002, 2007, 2011).
9. We report this definition of meta-agreement because it is the one most clearly stated in the literature. Other definitions are provided in Dryzek and List (2003: 14; 2004) and List and Koenig-Archibugi (2010: 95).
10. On this, see especially Dryzek and List (2003) and Dryzek and Niemeyer (2006). The idea that the 'benign kinds of pluralism' associated with meta-agreement can be reconciled with the need for coherent democratic decisions is also central in List and Koenig-Archibugi (2010: 95 especially).
11. This is a reply Aldred himself suggests, based on Dryzek and List (2003: 15).
12. Note that not every deliberative discussion is induced by normative divergent views on the policy choice space. We might imagine situations in which the agreement on values is given and the discussion among agents is triggered by disagreement about policy outcomes, that is, which policies best achieve the ends on which there is agreement. However, the meta-agreement model we are discussing here assumes that the main locus of disagreement is at the level of fundamental values.
13. David Miller seems to have clearly seen this in his seminal article on the topic, in which the examples he offers of a shared dimension are conceived in terms of tradeoffs between opposite values (1992: 63 especially).
14. What about those cases in which one single word and (from the extensional point of view) one single predicate is loaded with two opposite evaluations? For example, a right-libertarian might argue that we should prefer policy *a* to *b* because it is more 'laissez-fairist', while a state socialist might argue that *b* should be preferred over *a* exactly for the same reason, thus appealing to two opposite poles. These cases help make clear an interesting implication of our

claim, once we realize that they are structurally identical to cases in which two polarized predicates are involved. In these instances, if a new word were created to refer to one of the two poles (for example, if the state socialists described their preferred policies as 'interventionist' rather than 'non-laissez-fairist'), we would have two polarized predicates as just described. What is relevant here is not how many predicates are involved, but how many evaluations are attached to them; if there is meta-agreement without unanimity, there must be two opposite evaluations.

15. This form of agreement on what counts as a pair of opposing arguments can be analyzed more precisely by adopting Brandom's account of the discursive practice of giving and asking for reasons. See Brandom (2008) and Porello (forthcoming).
16. Someone who takes the middle ground between the two opposite poles might also be seen as implicitly referring to another dimension; however, this possibility is excluded by D1.
17. The word appears in List et al. (2006), referring to Elster (1986).
18. See, for example, Bohman (1998), Cohen (1989), and Freeman (2000). There is, of course, disagreement on what counts as a good definition of public deliberation; however, the idea that deliberation must necessarily imply meta-agreement as defined here (even assuming that deliberation has a transformative effect on preferences by appealing to generalizable interests) is not part of the common core of the most popular notions of public deliberation.
19. It might be objected that meta-agreement, although not *necessary* for reaching substantive agreement, is normatively and epistemically *desirable* because it implies not only that the parties agree on the same ranking of the alternatives, but also that they agree on the same *reasons* for endorsing it. It is difficult to see, though, how this could count as a reason for the model we are criticizing: when there is meta-agreement *without* substantive agreement (the alternative that advocates of meta-agreement favor over substantive agreement), there is no agreement on the reasons people have for how they rank the alternatives on the agenda because the parties must appeal to opposite values.
20. A more sophisticated mechanism, which cannot be fully accounted for within the framework we are discussing here, would be Rawls's 'overlapping consensus'. If we wanted to push an analogy between the two, Rawls's model might be seen as an instance of agreement on some core political values (Rawls, 1996: 14, 223), which then generates consensus on the ranking of the options at the level of constitutional essentials. There are at least three important differences, though, between Rawls's model, even so reconstructed, and the agreement model we are discussing: (1) the agreement on the core political values, according to Rawls (1996: 158), is not brought about by public deliberation; (2) in principle, agreement on the core political values should bring about agreement on the ranking of the options on the agenda (substantive consensus on constitutional essentials), while the meta-agreement model we are discussing is, instead, interested in accounting for meta-agreement *without* substantive consensus; and (3) there is no guarantee that the substantive consensus reached at the level of constitutional essentials will be kept at the level of ordinary legislation. Rawls's core political values constrain the possible outcomes of ordinary legislation, but such restriction does not guarantee that at this lower level no cycles will emerge.
21. Further evidence of the restrictive character of the theory of meaning presupposed by this approach is provided by the formal modeling of meta-agreement as induced by procedures for revising judgments that is presented in a recent article by Christian List (2011). Here, the individual judgments that may be revised through deliberation are represented by means of a given

set of propositions in (classical) propositional logic. Thus, every sentence is equipped with a very precise meaning which is constant throughout deliberation. With respect to a classical logic proposition, there are only two possible points of view: individuals accept it as true or reject it as false. Therefore, the only space left to model polysemy, revision, or disagreement is the swapping of the truth value.

22. Whether the language of the left–right ideological divide is actually conducive to single-peakedness (and the kind of role it has played in the history of western democracies) is matter of controversy. For a useful survey of the history of the debate, see Budge (2006).

References

Aldred J (2004) Social choice theory and deliberative democracy: A comment. *British Journal of Political Science* 34(4): 747–52.

Arrow K (1963) *Social Choice and Individual Values*. New Haven, CT: Yale University Press.

Black D (1958) *The Theory of Committees and Elections*. Cambridge: Cambridge University Press.

Bohman J (1998) The coming of age of deliberative democracy. *Journal of Political Philosophy* 6(4): 400–25.

Brandom RB (2008) *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press.

Budge I (2006) Identifying dimensions and locating parties: Methodological and conceptual problems. In: Katz RS and Crotty WJ (eds) *Handbook of Party Politics*. London: Sage, pp. 422–34.

Cohen J (1989) Deliberation and democratic legitimacy. In: Hamlin A and Pettit P (eds) *The Good Polity*. Oxford: Basil Blackwell, pp. 17–34.

Downs A (1957) *An Economic Theory of Democracy*. New York: Harper.

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

Dryzek JS and List C (2003) Social choice theory and deliberative democracy: A reconciliation. *British Journal of Political Science* 33(1): 1–28.

Dryzek JS and List C (2004) Social choice theory and deliberative democracy: A response to Aldred. *British Journal of Political Science* 34(4): 752–8.

Dryzek JS and Niemeyer S (2006) Reconciling pluralism and consensus as political ideals. *American Journal of Political Science* 50(3): 634–49.

Elster J (1986) The market and the forum. In: Elster J and Hylland A (eds) *Foundations of Social Choice Theory*. Cambridge: Cambridge University Press, pp. 103–32.

Farrar C, Fishkin JS, Green DP et al. (2010) Disaggregating deliberation's effects: An experiment within a deliberative poll. *British Journal of Political Science* 40(2): 333–47.

Freeman S (2000) Deliberative democracy: A sympathetic comment. *Philosophy and Public Affairs* 29(4): 371–18.

Habermas J (1996) *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA: MIT Press.

Hinich MJ and Munger MC (1997) *Analytical Politics*. Cambridge: Cambridge University Press.

Hinich MJ and Pollard W (1981) A new approach to the spatial theory of electoral competition. *American Journal of Political Science* 25(2): 323–41.

Knight J and Johnson J (1994) Aggregation and deliberation: On the possibility of democratic legitimacy. *Political Theory* 22(2): 277–96.

List C (2002) Two concepts of agreement. *The Good Society* 11(1): 72–9.

List C (2007) Deliberation and agreement. In: Rosenberg SW (ed.) *Deliberation, Participation and Democracy: Can the People Govern?* New York: Palgrave, pp. 64–81.

List C (2011) Group communication and the transformation of judgments: An impossibility result. *Journal of Political Philosophy* 19(1): 1–27.

List C and Koenig-Archibugi M (2010) Can there be a global demos? An agency-based approach. *Philosophy and Public Affairs* 38(1): 76–110.

List C, Luskin RC, Fishkin JS et al. (2006) Deliberation, single-peakedness, and the possibility of meaningful democracy: Evidence from deliberative polls. PSPE Working Paper No. 1. Available at: http://ssrn.com/abstract=1077752 (accessed 12 February 2011).

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

Miller D (1992) Deliberative democracy and social choice. *Political Studies*, special issue 40: 54–67.

Miller NR (1983) Pluralism and social choice. *American Political Science Review* 77(3): 735–47.

Poole KT and Rosenthal H (1994) Dimensional simplification and economic theories of legislative behavior. *Economics and Politics* 6(2): 163–72.

Poole KT and Rosenthal H (1997) *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Porello D (forthcoming) Incompatibility semantics from agreement. *Philosophia*.

Rawls J (1996) *Political Liberalism*, pbk. edn. New York: Columbia University Press.

Regenwetter M, Grofman B, Marley AAJ et al. (eds) (2006) *Behavioral Social Choice*. Cambridge: Cambridge University Press.

Riker W (1982) *Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice*. San Francisco, CA: Freeman.

Sunstein C (1988) Beyond the republican revival. *Yale Law Journal* 97(8): 1539–90.

Van Mill D (1996) The possibility of rational outcomes from democratic discourse and procedures. *Journal of Politics* 58(3): 734–52.

About the authors

Valeria Ottonelli is Lecturer in Political Philosophy at the University of Genova, Italy. Her main research interests are in immigration and global justice and in the normative theory of democracy.

Daniele Porello is Postdoctoral Research Fellow at the Institute for Logic, Language and Computation (ILLC) of the University of Amsterdam. His work lies at the intersection of logic, artificial intelligence, and social choice theory.