Meta-agreement and Rational Single-Peaked Preferences

Olivier Roy and Maher Jakob Abou Zeid

Abstract We revisit the claim that rationality requires participants in deliberation to form single-peaked preferences once they have reached meta-agreements. We provide two different arguments that cast doubts on this claim. The first points out the rationality of having non-single-peaked preferences in cases where consuming two goods together is less valuable than consuming each of them individually. The second argument fleshes out the notion of meta-agreements in terms of reasons supporting a particular structuring dimension. These arguments show that to the extent that deliberation fosters the formation of meta-agreements and the formation of single-peaked preferences, the bridge between these two notions might not be solely a matter of rational preference formation.

1 Introduction

This paper is a philosophical contribution to the question of whether deliberation helps avoid Condorcet cycles, and, more generally, incoherent social preferences. According to what has been called the *Received View* (Rafiee Rad & Roy, 2021), deliberation has this positive effect. The view goes back at least to Miller (1992), but its most standard formulation rests on the so-called *meta-agreement hypothesis*, as articulated by List (2002) and Dryzek and List (2003). In a nutshell, the hypothesis states that deliberation fosters the formation or the discovery of underlying meta-agreements, and that once such meta-agreements are in place rationality requires the participants to form single-peaked preferences.

This paper raises doubts regarding one important aspect of the meta-agreement hypothesis, namely that, in the presence of meta-agreement, rationality requires the agents to form single-peaked preferences. We do so in two ways. We first point

O. Roy (✉)
Department of Philosophy, University of Bayreuth, Bayreuth, Germany
e-mail: olivier.roy@uni-bayreuth.de

M. J. Abou Zeid
European Commission, Brussels, Belgium
e-mail: maher-jakob.abou-zeid@ec.europa.eu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_6

out the rationality of having non-single-peaked preferences in cases where consuming two goods together is less valuable than consuming each of them individually. We then flesh out the notion of meta-agreements in terms of reasons supporting a particular structuring dimension, and argue that there are cases where non-single-peaked preferences are permissible given a particular set of reasons. In Sect. 2 we provide the required theoretical background to our argument, introducing the meta-agreement hypothesis and single-peaked preference, and develop the counter-examples in Sect. 3.

2 The Meta-agreement Hypothesis

Black (1948) and later Arrow (1963) remarked that *single-peaked preferences* are sufficient to ensure coherent group preferences. More precisely, Black showed that pairwise majority voting always delivers a Condorcet winner when the input preference profile is single-peaked, and Arrow noticed that this aggregation method satisfies all of his other axioms once universal domain is narrowed to single-peaked preferences. See Gaertner (2001) and Puppe (2018) for comprehensive overviews of such domain conditions.

Single-peakedness is defined relative to a given ranking of the alternatives, often called a *dimension*. Informally, a particular strict ranking over a finite set of alternatives is single-peaked with respect to a given dimension whenever, as one moves away from the most preferred alternative to the left and the right along the given dimension, one always moves to strictly less preferred alternatives. Visually, one always goes "down" and never "up" again in the preference ranking. A profile of rankings is said to be single-peaked when there is a dimension along which all rankings in that profile are single-peaked.

When can the preferences of the voters be expected to be single-peaked? The Received View answers this question by bringing in the crucial notion of *meta-agreements*. A meta-agreement is an agreement regarding the structuring dimension of a particular decision problem. This is the dimension along which the participants conceptualize the problem. Later on we make this idea more precise in terms of the possible reasons that bear on the decision. For now, it suffices to point out that many such dimensions are possible. The classic example of a structuring dimension is the left-right alignments of political parties in an election. That alignment is, of course, not the only one possible. Parties might as well be aligned according to their authoritarian *vs* libertarian values. A group has formed a meta-agreement on a question when its members agree on the most relevant dimension to the problem at hand. In our example, the voters are said to have reached a meta-agreement when they agree that, for instance, the election's central issue is the choice between authoritarian and libertarian values.

Meta-agreements leave room for substantial forms of what might be called substantive disagreements. Even if the voters agree that a given election should be primarily framed as a choice between libertarian and authoritarian values, they might

disagree on the extent to which specific parties embody these values. And even if they agree on that, they might have opposing views regarding the best way to strike the compromise, if at all, between authoritarian and libertarian tendencies.

Single-peakedness, even if it implies the existence of a structuring dimension, is not sufficient for meta-agreements. The problem is, as List observes, that this dimension "is only a *formal* structural condition on individual preferences" (List, 2002 [our emphasis]). It might not be meaningful for the participants (Aldred, 2004) and, if it is, they might not agree that they should frame the problem at hand in its terms. In other words, the fact that the voters' preferences happen to be single-peaked does not necessarily entail a joint conceptualization of the decision at hand along a given dimension, which is central for meta-agreements. Returning to our example, two voters might completely disagree on the main issue of a specific election. One regards it as a choice between more left-wing or right-wing policies, while the other views it as a choice between authoritarian or libertarian values. If these two dimensions happen to be sufficiently correlated among the parties, these two voters might nonetheless have single-peaked preferences, or even reach consensus, without having reached a meta-agreement.

However, the question remains whether the existence of meta-agreements is conducive to the participants forming single-peaked preferences, and this is where the meta-agreement hypothesis comes in. The hypothesis is, in effect, a proposal regarding the mechanism through which meta-agreements foster the creation of single-peaked preferences. In the version outlined by List (2002) and Dryzek and List (2003), that mechanism works in three steps. The first step is the actual formation of the meta-agreement. Deliberation, they claim, helps the participants to identify or unveil the set of norms and values that constitute the relevant dimension(s) of the problem at hand, c.f. also Dryzek and Niemeyer (2006). This could be the trade-off between authoritarian and libertarian values in our running example. In a second step, deliberation helps the participants to agree on the factual question of how the alternatives compare on that dimension. That is, it helps them to rank the alternatives along that dimension. Again, in our example, deliberation is claimed to help the participants situate each party on the libertarian/authoritarian dimension. Note that the resulting ordering of the parties is neither an individual nor a social preference ranking. It only reflects how the alternatives compare to one another with respect to the structuring dimension. This leaves open which point on that dimension is optimal or best.

The third and final step is our main focus in this paper. Each participant should individually determine what point in the agreed dimension they find best and order the other alternatives relative to it. This is where deliberation should translate meta-agreement into single-peaked preferences (Dryzek & List, 2003; List, 2002). They suggest that rationality requires the participant to form single-peaked preferences with respect to the structuring dimension that they agree on. In other words, the claim is that if a participant has agreed that the dimension identified in the first step is indeed the *only* relevant one, and she also agreed on how each alternative fares on that dimension, then she must, on pain of incoherence, be able to identify the best alternative(s) along that dimension and order the others according to their distance from them. In our example, this means that once each voter has agreed that the trade-

off between libertarian and authoritarian values is the only relevant issue, and has determined how each party makes that trade-off, all that remains is to decide which trade-off is best, and order all strategies in terms of distance from that ideal point.

We should emphasize that this third step is done individually by the participants, but the result of group deliberation guides it. Taken in contrapositive, the claim is indeed that, if a participant forms non-single-peaked preferences at the third step, this must be because some other issues are important to her, after all, or because she disagrees on how the alternatives should be ranked on the unique dimension, contradicting the conclusions collectively reached at the first and second steps.

Our goal in this paper is to revisit and ultimately express doubts regarding this third step, but before presenting that, it is useful to review other existing criticisms. Ottonelli and Porello (2013) have argued that the first two steps of the mechanism require different forms of substantive agreements, and that it is debatable whether such agreements are easier to reach than fully consensual preferences. The mechanism indeed requires agreement regarding the relevant dimension and regarding the position of the alternatives on it. Both aspects are likely to involve a number of intricate or even thick, value-laden concepts. For those, it seems unlikely that the participants will reach a consensus on their meaning or their concrete realization in each of the alternatives. To start with the second aspect, even assuming in our running example that the voters agree on what are libertarian and authoritarian values, it is not implausible that deep disagreements will persist after deliberation regarding the extent to which each party embodies them. Regarding the first aspect, recall that it requires the participants to agree on the problem's relevant normative or evaluative dimension. This dimension will typically reflect a thick concept, intertwining factual with normative and evaluative questions, for instance, health, well-being, sustainability, freedom, or autonomy, to name a few. It seems rather unlikely, Ottonelli and Porello (2013) argue, that deliberation will lead the participants to agree on the meaning of such contested notions. Of course, deliberative democrats have long observed that public deliberation puts rational pressure on the participants to argue in terms of the common good (Miller, 1992), which might be conducive to an agreement on a shared dimension. But when it comes to such thick concepts, this agreement might be only superficial, involving political catchwords and thus leaving the participants using their own, possibly mutually incompatible, understandings of them. All of this does not exclude the fact that deliberation might make it more likely, in comparison with other democratic procedures, to generate single-peaked preferences from meta-agreements. The point is rather that starting from the latter puts the bar very high, especially if there appear to be other ways of reaching single-peaked preferences or of avoiding incoherent group rankings altogether.

3 Single-Peakedness Through Rationality?

The meta-agreement hypothesis, and the received view more generally, have received some empirical support, primarily reported in List et al. (2012) and Farrar et al.

(2010). They used prominence in the public sphere as a proxy for the existence of meta-agreements before deliberation. They observe stronger increases in proximity to single-peakedness in cases where the issue at hand was less prominently discussed publicly before deliberation, and interpret this data as showing that the formation or the discovery of meta-agreements through deliberation goes together with an increase in proximity to single-peakedness, has suggested by the meta-agreement hypothesis.

This evidence for the meta-agreement hypothesis is, however, indirect. While it suggests a correlation between the formation or the discovery of meta-agreements and an increase in proximity to single-peakedness, it does not directly test whether the specific mechanism postulated in the hypothesis is causally responsible for the increase. The data is silent, in particular, on whether the participant felt in any way compelled, presumably by rational pressure, to form single-peaked preferences once they have identified a structuring dimension and positioned the alternatives on it.

Rafiee Rad and Roy (2021) have, in fact, provided evidence from computational simulations that suggests that increases in proximity to single-peakedness might rather result from willingness to reach consensus than from a rational response to meta-agreements. On the one hand, they observe that rational preference change alone is insufficient to ensure an increase in proximity to single-peakedness. For decisions on three alternatives (Abou Zeid, 2021), rational deliberation sometimes insufficiently increases proximity to single-peakedness and even tends to create incoherent group preferences. Rather, the computational model suggests an alternative mechanism: the degree to which participants in deliberation are willing to reach a consensus with others might be the main driver for the increase in proximity to single-peakedness observed in deliberation—c.f. again Abou Zeid (2021) and also Rafiee Rad (2022) for additional remarks to that effect.

In this section, we want to formulate two conceptual arguments that support the search for alternative explanations of the observed correlation between meta-agreement and increases in proximity to single-peakedness. Both arguments question the claim that rationality requires the participants to form single-peaked preferences once they have reached meta-agreement. The first argument intuitively appeals to cases where the interaction of two goods naturally leads to non-convex preferences. We question why such cases should be seen as irrational once meta-agreements are reached. The second argument fleshes out this intuition by using the theory of reason-based rational choice, developed by Dietrich and List (2013). Interpreting meta-agreements as constraining the admissible reasons that can ground preference relations, we argue that it might be rational to form non-single-peaked preferences even in the presence of meta-agreements.

3.1 The Case of Non-convex Preferences

Consider a simple example of some friends deciding where to go for lunch in Munich. The options might be “Japanese”, “Bavarian”, and “Japanese-Bavarian”, and everyone agrees that the question is how exotic the food may be, “Bavarian” being the

least exotic, "Japanese" the most, "Japanese-Bavarian" in between.[1] Suppose that the friends have a fixed budget to be spent on two goods that are available in three combinations. Call the degree to which the food is exotic x, the degree to which it is conventional c. Let us furthermore assume that the goods have the same price $p_x = p_c$.

Assuming that these preferences can be represented by a *convex* utility function along the exotic/conventional dimension boils down to assuming that these preferences are single-peaked. The convexity assumption is often grounded on the fact that economics usually looks at goods that are voluntarily consumed together. However, whether this is the case is an empirical assumption and does not provide an argument to the effect that our agents are rationally compelled to have convex utility functions, i.e., single-peaked preferences.

Indeed, double-peaked/single-dipped (Barberà et al., 2012) preferences are still rational in the classical sense of being transitive and complete, and can naturally arise when consuming goods together is less valuable than consuming them individually. In our example, the participants are forced to consume x and c together, and the double-peaked preference can be modeled by the utility function $u(x, c) = (1 - x)^2(1 - c)^2$ where x and c are, again, the degrees to which the food is exotic and conventional. The indifference-curves that are implied by this are given by $I(x) = 1 - \frac{\sqrt{\bar{u}}}{1-x}$ and look exactly opposite to "usual" indifference-curves. An agent with such preferences consuming these goods in isolation rather than in combination.

Intuitively, it is not clear why it is rational for some of the friends in our examples to have convex preferences over x and c, but *not* otherwise. It is commonplace to have goods that are less preferred when consumed as bundles. A good example is pickles and jam. Someone might like either on their toast, but liking the combination of both on one toast is certainly less common. In other words, many have non-convex preferences when it comes to combinations of pickles and jam. When a group of friends argues about what kind of sandwiches they should prepare for the lunch pack, it seems that the meta-agreement that the relevant issue at hand is whether a given sandwich is sweet or sour will not necessarily result in single-peaked preferences.

So even in simple, two goods models of preferences over a unique dimension, it is not clear why rationality requires the agents to have convex preferences over their bundles. If that intuition is correct, then the claim that rationality requires to form single-peaked preferences once a meta-agreement is reached must at least be qualified to the "standard" case where the agents' preferences are convex.[2]

[1] We assume that the "Japanese-Bavarian" venue serves dishes from both traditions, not necessarily that they combine them in one dish.

[2] As pointed out by an anonymous reviewer to this paper, non-convexity is not the only plausible example of rational, non-single-peaked preferences that are compatible with meta-agreements. So-called group-separable preferences (Inada, 1964) seem to provide another plausible class of examples. We are very grateful to the anonymous reviewer for this pointer.

3.2 *Meta-agreements as Constraints on Reason-Based Preferences*

A natural explanation of our intuition regarding the rationality of having non-convex preferences over certain bundles of goods is that there is a different set of reasons grounding the agents' preferences. In the case of our friends choosing a restaurant, the person preferring both more conventional and more exotic venues to anything "in-between" might do so because she dislikes unusual combinations. So for her, the choice is between venues that are more "on the beaten paths" and those that are less so. Similarly, having non-convex preferences over bundles of pickles and jam as might be explained as a choice between conventional and less conventional bread spreads, instead of between sweet and sour ones. In both cases, the explanation boils down to denying that the agents have, in effect, reached a meta-agreement. The dimension structuring their choices is not what they had purportedly agreed on.

This suggests a natural way to flesh out the idea that rationality requires to form single-peaked preferences once meta-agreements are reached, in terms of the reasons that should ground the participants' preferences. The idea, already alluded to by Ottonelli and Porello (2013), would be that what the structuring dimension singled out by the meta-agreement does is to pinpoint a set of reasons that are viewed as most prominently bearing on the discussion at hand. Reaching a meta-agreement would then mean that these are the reasons that all participants agree each should take into account while individually forming their preferences. Crucially, the agents need not agree on a weighing of these reasons. The structuring dimension might induce such a weighing (see below), but as we have seen earlier this dimension only expresses how different alternatives embody possibly incompatible values/properties. It leaves open how the agents should weigh these combinations.

To go back to our examples, the dimension structuring the choice of restaurant would then be seen as singling out the property of being conventional or exotic as the relevant reason to consider. The alternative explanation for having non-convex preferences over bundles of exotic/conventional goods would then point to a different set of reasons for the agents' choices, namely how frequent these culinary offers are. Similarly, in the case of the pickles and jam spread, the first structuring dimensions would be seen as singling out sweetness and sourness as the relevant reasons to consider, and the alternative explanation points to conventionality instead.

This idea can be made precise in the framework of reason-based rational choice developed by Dietrich and List (2013). Here we illustrate it through our simple example. Our friends need to choose one of three alternatives, Japanese (j), Bavarian (b), and Japanese-Bavarian (jb). The structuring dimension is the ranking $j \prec jb \prec b$, from less more conventional. Dietrich and List (2013) propose to view such rankings as grounded in a set $\mathcal{M}$ of possible combinations of *motivating reasons*, together with a weighing relation $\succeq$ on them. A motivating reason is, in this framework, represented by a property that can be instantiated or not by each alternative. In this case the structuring dimension $j \prec jb \prec b$ can be naturally be seen as singling out the combination of two motivating reasons, being exotic (E) or being conventional

(C), with the extension of each following our informal description: $E = \{j, jb\}$, $C = \{jb, b\}$. The dimension then stems from the following weighing of reasons: $\{C\} \succ \{E, C\} \succ \{E\} \succ \emptyset$, in the case where both E and C are motivating.

We thus propose to view meta-agreement as constraining the set of possible combinations of motivating reasons, but not necessarily how one should weigh them. So in our simple example, the structuring dimension $j < jb < b$ expresses the fact that the agents agree that two reasons are relevant to the problem at hand, whether a venue is exotic or conventional, and that any possible combination of those reasons, but *not of other reasons*, may be taken as motivating for the agents. As we have seen, the structuring dimension $j < jb < b$ is not supposed to express a value judgment regarding the alternatives. It is rather an (empirical) ranking capturing the idea that exoticism and conventionality might be at least partially incompatible. From that point of view, it seems natural to suppose the meta-agreement constraints the agents to consider only those possible combinations of reasons, but that it does not constraint them to weigh these reasons in any particular way.

If that proposal is correct, then meta-agreements need not translate into single-peaked preferences along the structuring dimension. In our case an agent could weigh the possible combinations of reasons as follow, $\{C\} \succ \{E\} \succ \{E, C\} \succ \emptyset$, resulting in the preference $b > j > jb$, which is of course not single-peaked relative to the structuring dimension. The meta-agreement rules out alternative considerations like the one we envisioned at the beginning of this section, i.e., that the agent might instead conceive the decision as one between frequent and less frequent culinary offers.

4 Conclusion

We have presented two arguments putting into question the idea that rationality requires one to form single-peaked preferences once they have reached a meta-agreement. Our first argument points out a counter-intuitive consequence of that claim, namely that agents with non-convex preferences might be seen as irrational. Our second argument fleshed out the role of meta-agreement as pinpointing the set of possible reasons motivating a decision, but leaving open how one should weigh these. By formalizing our simple running example of a choice of restaurant, we showed that this understanding of meta-agreement leaves room for rationally holding non-single-peaked preferences.

We should emphasize that these arguments do not question the Received View as a whole, nor the first two steps of the meta-agreement hypothesis. One can still take the empirical evidence to show that deliberation helps form or discover underlying meta-agreements, that this comes together with increases in proximity to single-peakedness, and that this is achieved by way of situating the alternatives along a given structuring dimension. What our argument puts into question is that rational pressure is *the* cause of the resulting single-peaked preferences. As we mentioned, other possible explanations have been put forward recently, e.g., willingness to reach consensus

(Rafiee Rad & Roy, 2021). Our argument stresses the importance of investigating further, both theoretically and empirically, how deliberation affects preference formation and preference change once a meta-agreement is in place to adjudicate better between these different possible mechanisms.

References

Abou Zeid, M. J. (2021). *Collective rationality and deliberation over five and more alternatives*. Master's thesis, University of Bayreuth.

Aldred, J. (2004). Social choice theory and deliberative democracy: A comment. *British Journal of Political Science, 34*, 747–752.

Arrow, K. J. (1963). *Social choice and individual values*. Yale University Press.

Barberà, S., Berga, D., & Moreno, B. (2012). Domains, ranges and strategy-proofness: The case of single-dipped preferences. *Social Choice and Welfare, 39*, 335–352.

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy, 56*, 23–34.

Dietrich, F., & List, C. (2013). A reason-based theory of rational choice. *Nous, 47*, 104–134.

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

Dryzek, J. S., & Niemeyer, S. (2006). Reconciling pluralism and consensus as political ideals. *American Journal of Political Science, 50*, 634–649.

Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C., & Paluck, E. L. (2010). Disaggregating deliberation's effects: An experiment within a deliberative poll. *British Journal of Political Science, 40*, 333–347.

Gaertner, W. (2001). *Domain conditions in social choice theory*. Cambridge University Press.

Inada, K.-I. (1964). A note on the simple majority decision rule. *Econometrica, 32*, 525–531.

List, C. (2002). Two concepts of agreement. *The Good Society, 11*, 72–79.

List, C., Luskin, R. C., Fishkin, J. S., & McLean, I. (2012). Deliberation, single-peakedness, and the possibility of meaningful democracy: Evidence from deliberative polls. *The Journal of Politics, 75*, 80–95.

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

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

Puppe, C. (2018). The single-peaked domain revisited: A simple global characterization. *Journal of Economic Theory, 176*, 55–80.

Rafiee Rad, S., Braun, S. T., & Roy, O. (2022). *Anchoring and deliberation over preference rankings*. Working Paper, University of Bayreuth.

Rafiee Rad, S., & Roy, O. (2021). Deliberation, single-peakedness, and coherent aggregation. *American Political Science Review, 115*, 629–648.