The General Will, Group Decision Theory, and Indeterminacy

Mathias Risse, Cambridge / MA

1. Rousseau and Contemporary Group Decision Theory

1.1 Rousseau's *Social Contract* asks how individuals can live together without losing their freedom: Is there "a form of association which defends and protects with all common forces the person and goods of each associate," Rousseau wonders, "and by means of which each one, while uniting with all, nevertheless obeys only himself and remains as free as before" (I.6)? Rousseau wrote at the dawn of a golden age of reflection on group rationality. The last 50 years also witnessed a dramatic increase in understanding of formal aspects of group rationality. Yet curiously, we understand the mathematics of social choice better than its philosophy. But first let me continue with Rousseau for a bit.

Rousseau's solution to his puzzle is "an act of association" that "produces a moral and collective body composed of as many members as there are voices in the assembly, which receives from this same act its unity, its common self, its life and its will" (I.6). As I understand Rousseau, this *common self* belongs to each individual, but is also associated with the political community in the same way in which selves are associated with persons. So Rousseau's problem admits of a solution if that self is each individual's *true* self. For then each true self is identical to the self that belongs to the body politic. The *general will* of the community, then, is the will associated with this common self in the same way in which wills are associated with individual selves. Therefore, whatever the community does in accordance with the general will expresses the will of each true self, and so is binding on each individual while leaving him free. To complete the answer, we must show that the general will is constructed from individual beliefs and values by a process rendering it the will associated with each individual's true self. To this end, I interpret the general will as the decisions made by the community with its deliberation subject to suitable constraints.

However, even *constrained* deliberation is not guaranteed to reach unanimity. Decisions must often be made under circumstances of radical and persistent disagreement; conflicts of values that cannot be realized in single lives, single decisions of deliberating bodies, or single constitutions require decisions in spite of irresolvable disagreement. So we need an account of what to do when deliberation ends inconclusively. We might then still reach a general will, *provided* that for any set of circumstances under which a group might find itself in disagreement there is a uniquely reasonable rule which delivers that will. So we need to investigate the following thesis:

> *Uniqueness*: for any given set of circumstances, there is a uniquely reasonable decision rule the group should adopt if deliberation fails to deliver a decision.

If Uniqueness fails, there will be multiple inconsistent but reasonable methods under the same conditions: some will "lose" because one method was adopted, though they may have "won" had another, equally reasonable, rule been used. I argue that Uniqueness is false. Frequently, we cannot know the general will if deliberation ends inconclusively, and may also not have any reason to think that there actually *is* a general will in such situations.

1.2 We are now in the midst of contemporary group decision theory. Group decision theory constitutes no unified field. Instead, there are separate literatures on areas like fair division, game-theoretic scenarios, and types of aggregation: aggregation of preference rankings, expected utilities with shared probability functions, and, recently, Bayesian aggregation of utilities and subjective probabilities. My concern is with aggregation.

Thinking about aggregation leads to puzzles. Suppose a department must make a hiring decision. All members rank the candidates. How should they determine a group ranking? One way of doing so is by aggregating ordinal rankings, using proposals to be discussed in section 2. But suppose somebody suggests a 100-point system: each voter assigns from 0 to 100 points to each applicant, and then they average. This proposal uses more than ordinal information. Should we adopt this rule instead? Can we distinguish conditions under which ordinal rankings are appropriate from conditions under which other rankings are? Such questions are hard to answer and call for a theory that assesses, first, the conditions under which particular kinds of rankings are appropriate (e.g., ordinal or point systems); second, what specific rule(s) is (are) appropriate for the specific kinds of rankings; and third, what the criteria for "appropriateness" are, respectively. While we do not possess any such theory, I use this framework to refute Uniqueness. Section 2 argues that Uniqueness is false if we aggregate ordinal rankings, and section 3 argues this for Bayesian aggregation. Or, to put the claim in terms from the 18th century, we have no particular reason to think there is a general will unless deliberation ends with unanimity.[1]

2. Preference Aggregation: Borda vs. Condorcet

2.1 Suppose we must rank m candidates. The *Condorcet proposal* does so by looking at all m(m-1)/2 pairs among m candidates, selecting one or more of the m! rankings in light of these pariwise votes. We consider all pairwise votes as "data" and ask which ranking of the candidates these data support best. The Condorcet proposal selects a ranking *supported by a maximal number of votes in all pairwise elections*. For each of the m! rankings, we look at the m(m-1)/2 pairs of candidates and count the number of voters that rank the respective candidates in the same way as that pair and thus *support* the ranking. To illustrate, suppose a group of 48 must rank A, B, and C. 10 rank them (A, B, C), 12 (A, C, B), 5 (B, A, C), 7 (B, C, A), 3 (C, B, A) and 11 (C, A, B). So here we have 3! = 6 rankings and 3(3-1)/2 = 3 pairwise votes. The ranking with the highest support in pairwise elections is (A, C, B): In A vs. B, 33 people support it (33 people rank A over B), in B vs. C 26 people, and in A vs. C 27. Thus 86 votes support (A,

[1] Due to the length constraints, the argument in this paper is sketchy. Some of the arguments can be found at other places as well, although I have never brought them together in a unified argument opposing Uniqueness. Section 2 draws on my "Arrow's Theorem, Indeterminacy, and Multiplicity Reconsidered," in *Ethics* 111 (2001), pp 706-734; similarly, section 3 draws on my article "Bayesian Group Agents and Two Modes of Aggregation," forthcoming in *Synthese* (2003). That article, in turn, draws on joint work with Richard C. Jeffrey and Matthias Hild: "Preference Aggregation after Harsanyi," forthcoming in *Social Choice and Welfare* (2003).

C, B) in pairwise votes, compared to 82 for (A, B, C), 64 for (B, A, C), 58 for (B, C, A), 62 for (C, B, A), and 80 for (C, A, B). So (A, C, B) is a *ranking with maximal support*.

Consider now the *Borda count*. Again a group must rank m candidates. First each person ranks them, assigning 0 to her last-ranked, 1 to her second to the last ranked, etc., until she assigns m-1 to her first ranked. Then a sum over these numbers is formed for each candidate, which is that candidate's *Borda count*. The group ranks the candidates by decreasing Borda counts. Suppose we have three people (1, 2, 3) and four candidates (A, B, C, D). Person 1 ranks them (A, B, C, D), 2 (B, C, D, A), 3 (A, B, D, C). The social ranking is (B, A, C, D) because A obtains six points, B seven, C three and D two. Another characterization of the Borda count is that it ranks candidates by their *average position* across rankings. Like the Condorcet proposal, the Borda count uses as "input" all m(m-1)/2 pairwise votes. But Borda asks about the *support for each of the m candidates in all rankings*, whereas Condorcet asks about the *support for each of the m! rankings in all pairwise elections*.

2.2 How to decide between Condorcet and Borda? On earlier occasions I argued for three claims: first, arguments for majority rule tend to be limited to the case of two candidates; second, generalizations of these arguments support the Condorcet proposal, and thus that proposal is what we should mean in general by majoritarian decision making; and third, none of these arguments is decisive against the Borda count. I illustrate this for one claim. One argument for majority rule is

Condorcet's Jury Theorem: Supposes it makes sense to speak of being right or wrong about political decisions. Suppose n agents choose between two options; that each has a probability of p>½ of being right; and that their probabilities are independent of each other (i.e., they make up their minds for themselves). Then, as n grows, the probability of a majority's being right approaches 1.

Formulated like this, the theorem only applies to the case of two candidates. However, there is a generalization, and this generalization picks out the rankings with maximal support. The procedure is to go through all m! rankings and calculate the conditional probability of the pairwise voting results given that the respective ranking is the correct one. The ranking that bestows maximum likelihood to the voting result is chosen.[2]

According to the theorem, rankings selected by Condorcet bestow the highest likelihood on the election result. Yet the option ranked highest by Borda is the *single option* that, if the best, bestows highest probability upon the voting results, provided the voters' competence p is close to ½. Borda is concerned with ranking *options* in terms of their rightness, and Condorcet with finding the right *ranking*. For values of p close to ½, this difference carries over to the epistemic scenario, with Condorcet searching for the ranking with maximal likelihood, and Borda ranking the options in terms of their maximal likelihood. The Borda count has its counterpart to this theorem reflecting this difference and regards arguments drawing on it as non-starters. Other arguments in support of the Condorcet proposal fare similarly, and *mutatis mutandis* for Borda. As far as rules for aggregating rankings are concerned, Condorcet and Borda are on a par: neither has conclusive arguments against the other, whereas both are reasonable rules. Thus we have sketched an argument against Uniqueness if we are aggregating ordinal rankings.

3. Bayesian Aggregation: *Ex Post* vs. *Ex Ante*

3.1 Our second task is to show that Uniqueness fails for Bayesian aggregation. A Bayesian agent is an agent described by theories of expected utility theory that take probabilities to be subjective. Suppose we have a group of such agents, and suppose they would like for their group as such also to be a Bayesian agent. Moreover, they also would like for group decisions, and thus for group preferences, probabilities, and utilities, to be aggregated from the respective individual entities in such a way that at least unanimous agreements are preserved.

Mongin (1995), using the Savage framework, shows that Bayesian aggregation satisfying some reasonable condition is available only to fairly homogenous groups.[3] Mongin's result holds in *ex ante* frameworks, where restrictions on group expected utilities are formulated in terms of individual expected utilities. A typical *ex ante* rule may include a restriction of the kind "If each individual's expected utility is higher for a than for b, the group should rank a higher than b." Mongin's result fails in the *ex post* framework, which aggregates probabilities and utilities separately. Typical *ex post* aggregation rules may satisfy the restriction "If each individual's utility for *a* is higher than the utility for *b*, then the group utility for *a* should be higher than the group utility for *b*." A parallel condition for probabilities may hold.

Yet *ex post* models come with trouble of their own: They display a peculiar dependence on the level of detail used in describing the decision problem. Suppose a group of Bayesians aggregate probabilities and utilities according to some *ex post* rule. Suppose they analyze their situation in more detail without anybody changing her mind about the ranking of the actions. Then the group as a whole should not change "its" mind about that ranking when the more fine-grained utilities and probabilities are aggregated. However, such a reversal may happen under *ex post* rules. So we must choose between two modes of aggregation each of which has its problems.

3.2 On what grounds can we choose? Suppose Ante and Post defend those approaches. Both need, first, a positive argument for their positions; second, a negative argument attacking the opponent's view; and third, an argument why the aforementioned worries about their account are not troublesome. Let us sketch these positions.

Ante's positive argument is that for agents to be taken seriously as participants in the decision process means being accepted as *decision makers*, rather than probability and utility providers. Agents' decisions, however, are based on preferences, or expectations. Thus restrictions on the decision process should be formulated in terms of expectations, that is, in the *ex ante* fashion. Ante's complaint against Post is that the instability result shows that the *ex post* approach is futile. Since we never know for sure whether a more fine-grained look at the same situation would reverse group preferences, decisions based on the *ex post* approach are ill-founded. As far as Mongin's theorem is concerned, Ante may bite the bullet by saying that Mongin's theorems teach us a lesson about

[2] This paragraph draws on Peyton Young, "Condorcet's Theory of Voting", *American Political Science Review* 82 (1988): 1231-1244,

[3] Philippe Mongin, "Consistent Bayesian Aggregation." *Journal of Economic Theory* (1995); 66, 313-351

group decision making. Aristotle was right when, in the *Politics*, he thought of a political community as a community of like-minded persons. Rational decision making is unavailable to groups with little in common.

Post proceeds as follows: It is in accordance with the Bayesian credo to distinguish facts from values. If the group is the decision maker, its method must be to aggregate probabilities and utilities separately. Agents are taken seriously if their epistemic views and their values are taken seriously. This view takes account of the reality of the decision structure while assigning individuals the appropriate place as *members* of the group within that structure. Post's complaint against Ante is based on Mongin's theorem. Post thinks of this result as a *reductio* of the *ex ante* approach. For rational social choice cannot be restricted to homogeneous groups, leaving inhomogeneous groups without any advice. Post has two replies to the instability result: On the one hand, he may deny its relevance and argue that in all decision situations that we encounter in our lives, even in strongly idealized ones, there will be a point when we have reached the most fine-grained relevant refinement (or at least, the most fine-grained refinement we are capable of considering). Though it is hard to discuss this any further at the general level, Post might continue, it will be clear enough in each given scenario. On the other hand, it is also open to Post to bite the bullet and argue that groups are fragile decision makers, but that this insight, far from refuting the *ex post* approach, teaches us something about the nature of groups as decision makers. While, once again, this argument is sketchy, I claim that neither position has resources to refute the other. This establishes the falsity of Uniqueness of Bayesian aggregation.

4. Conclusion

Rousseau's general will, I suggested, should be understood as the decisions of the respective group if its deliberation is subject to suitable constraints. Since even such deliberation cannot ensure unanimity, we must ask how a group should reach a decision if deliberation ends inconclusively. In many cases, aggregation of individual views is needed. For us to have access to the general will, there would have to be a unique aggregation rule under all conditions under which aggregation is required. I have argued for two kinds of settings that there is no such uniqueness: there is no uniquely reasonable decision rule if ordinal rankings are aggregated, and there is no uniquely reasonable rule if both utilities and subjective probability are aggregated. Thus in many cases some individuals will be "losers" in the decision process although they would not have been if had another, equally reasonable rule been adopted. This, I claim, is an inescapable feature of group decision making. [4]

[4] Prepared for a special session on the work of Richard Jeffrey during the 26th International Wittgenstein Symposium in Kirchberg (Austria), August 3-9, 2003. My assignment is to address Jeffrey's group decision theory. He did not work much on those areas. Around 1970, Jeffrey worked on a book tentatively called "The New Utilitarianism," but later abandoned the project. The result were two papers on interpersonal comparisons of utility, reprinted in *The Art of Judgment*, a collection of his papers. He returned to group decision theory towards the end of his life, mostly in joint work with Matthias Hild and myself about the epistemic foundations of game theory and Bayesian aggregation. My concern in this paper is to put our work on Bayesian aggregation into the context of a larger question about group rationality.