Formal limits to democracy? The problem of judgment aggregation as a challenge for the epistemic justification of democracy

Daniel Eckert and Christian Klamler
University of Graz

October 2008

1 Introduction

The formal analysis of collective decision making in general and voting in particular in social choice theory has always disturbed political theorists by its essentially negative results, like Arrow's (1964) famous (Im)Possibility Theorem (for a survey on social choice theory see Kelly 1988). The latter states that the only aggregation rule for individual preferences which guarantees a rational outcome (i.e. a transitive preference relation) and satisfies a number of other desirable properties is - a dictatorship.

Recently, the formal analysis of collective decision making has been extended from the aggregation of preferences to judgment aggregation (List and Pettit 2002, for a survey see List and Puppe 2007). Far from providing an epistemic justification of democracy, the results in this literature show a conflict between the elementary rationality criterion of the logical consistency of the outcome and the weakest possible fairness criterion for collective decision procedures.

By using an ultrafilter proof technique for a typical impossibility result in judgmemt aggregation by Nehring and Puppe (2005), we show how the logical structure of the agenda (given by the interconnections between the propositions to be decided upon) determines an extremely asymmetric power structure (i.e. the distribution among the individuals of decisiveness over these propositions).

2 Ultrafilters for the analysis of judgment aggregation problems

Although ultrafilters[1] have long been used in the proof of Arrow's theorem (see e.g. Monjardet 1983), this proof technique has not been exploited in the recent, closely related literature on judgment aggregation[2] with the notable exception of Gaerdenfors 2006, Dietrich and Mongin 2007 (especially for the infinite case) and Daniels 2006 (the latter in the different context of the logical formalization of judgment aggregation). This is allthemore astonishing as the very first application of an ultrafilter proof strategy to Arrow's theorem can be found in an early extension of this result to the aggregation of logically interconnected propositions by Guilbaud (1952) which makes this paper the first contribution to judgment aggregation. (For the reconstruction and historical analysis of Guilbaud's result see Monjardet 2003 and 2005.)

The problem of judgment aggregation consists in the derivation of collective judgments over an agenda of logically interconnected propositions from individual judgment sets. Following Dietrich (2007), an agenda is given by a set X of propositions (sentences) from a formal language $\mathbf{L}$ which is closed under negation (i.e. if $p \in \mathbf{L}$, then $\neg p \in \mathbf{L}$) and satisfies the following consistency conditions on judgment sets:

(C0) For any proposition $p \in \mathbf{L}$, the sets $\{p\}$ and $\{\neg p\}$ are both consistent (i.e. $\mathbf{L}$ does not contain a tautology).

(C1) For any proposition $p \in \mathbf{L}$, the set $\{p, \neg p\}$ is inconsistent.

(C2) Subsets of consistent sets $S \subseteq \mathbf{L}$ are consistent.

(C3) $\varnothing$ is a consistent set, and each consistent set $S \subseteq \mathbf{L}$ can be completed to a consistent superset $T \subseteq \mathbf{L}$ containing a member of each proposition-negation pair $p, \neg p \in \mathbf{L}$.

An individual judgment set is a subset $A \subset X$. Typically judgment sets are assumed to be fully rational in the sense of being not only consistent but also complete (i.e. $p \in A$ or $\neg p \in A$ for every proposition p). For N being a set of n individuals ($n \geq 2$), a profile of judgment sets is an n-tuple $\underline{A} = (A_1, ..., A_n)$. A judgment aggregation rule is a function which assigns to each profile in a set of admissible profiles a collective judgment set. Obviously, the aggregation problem crucially depends on the properties

[1] An ultrafilter on a nonempty set S is a collection $\mathcal{F} \subset 2^{|S|}$ of subsets of S such that
(i) $S \in \mathcal{F}$ and $\emptyset \notin \mathcal{F}$,
(ii) if $X \in \mathcal{F}$ and $Y \in \mathcal{F}$, then $X \cap Y \in \mathcal{F}$,
(iii) if $X, Y \subset S$, $X \in \mathcal{F}$, and $X \subset Y$, then $Y \in \mathcal{F}$,
(iv) for every $X \subset S$, either $X \in \mathcal{F}$ or $S \backslash X \in \mathcal{F}$.

[2] See List and Puppe 2007 for a survey and the bibliography at: http://personal.lse.ac.uk/LIST/doctrinalparadox.htm

of the agenda, essentially the logical connections between the propositions in the agenda.

Following Nehring and Puppe 2002, and Dokow and Holzman 2005, the logical connections between the propositions in any agenda X are captured by a binary relation $\vdash^* \subset X \times X$ of conditional entailment between propositions.

Definition 1 *For any propositions $p, q \in X$ such that $p \neq \neg q$, $(p, q) \in \vdash^*$ if there exists a minimally inconsistent superset S of $\{p, \neg q\}$ (i.e. a set $P \subset X \backslash \{p, \neg q\}$ such that $S = P \cup \{p, \neg q\}$ is inconsistent while every proper subset of S is a consistent set of propositions).*

Thus for any contingent propositions $p, q \in X$, $(p, q) \in \vdash^*$ means that there exists a set of propositions $P \subset X \backslash \{p, \neg q\}$ conditional on which holding proposition p entails holding proposition q.

Example 2 *Consider an agenda $X = \{p, \neg p, q, \neg q, p \vee q, \neg(p \vee q), p \wedge q, \neg(p \wedge q)\}$. Then the set $S = \{\neg p, p \vee q, \neg q\}$ is a minimally inconsistent set which establishes the conditional entailment relation between e.g. $\neg p$ and q.*

Definition 3 *An agenda X will be called **totally blocked** if the transitive closure of the relation of conditional entailment $T(\vdash^*) = X \times X$.*

Total blockedness means that any contingent proposition in the agenda is related to any other one by a sequence of conditional entailments.

Example 4 *Verify that the above agenda $X = \{p, \neg p, q, \neg q, p \vee q, \neg(p \vee q), p \wedge q, \neg(p \wedge q)\}$ is totally blocked, while neither the agenda $Y = \{p, \neg p, q, \neg q, p \wedge q, \neg(p \wedge q)\}$ nor $Z = \{p, \neg p, q, \neg q, p \vee q, \neg(p \vee q)\}$ is totally blocked.*

Observe that in the case of an agenda with more than two non-equivalent proposition-negation pairs total blockedness implies that there always exists a minimally inconsistent set of propositions with cardinality strictly larger than two.

The social choice literature provides a wide range of domain conditions for aggregation rules. The following conditions are extensions of classical Arrovian conditions to the area of judgment aggregation.

Definition 5 *A judgment aggregation rule $f : D^n \to D$ satisfies **universal domain** and **collective rationality** if D is the set of fully rational judgment sets.*

Definition 6 *A judgment aggregation rule $f : D^n \to D$ satisfies **sovereignty** if $f(D^n) = D$, i.e. if it is surjective.*

Like social welfare functions, judgment aggregation rules can be analyzed with the help of decisive sets of individuals. In the following and for any profile $\underline{A} \in D^n$ and any proposition $p \in X$ denote by $\underline{A}(p) := \{i \in N : p \in A_i\}$ the set of all individuals who hold proposition p. This yields the following definition of decisiveness.

Definition 7 *A set $U \subseteq N$ of individuals is* ***decisive*** *for proposition $p \in X$ if for all profiles $\underline{A} \in D^n$,*
$U = \underline{A}(p) \Rightarrow p \in f(\underline{A})$.
The family of all decisive sets for proposition p will be denoted by $\mathcal{W}_p \subset 2^{|N|}$.

An important class of properties are independence conditions which are - in general - crucial for the derivation of impossibility results in aggregation problems. We use another but equivalent definition of the usual independence condition[3]:

Definition 8 *A judgment aggregation rule $f : D^n \to D$ is* ***independent*** *if for any proposition $p \in X$, and any profiles $\underline{A}, \underline{A}' \in D^n$,*
$p \in f(\underline{A}) \Rightarrow [\underline{A}'(p) = \underline{A}(p) \Rightarrow p \in f(\underline{A}')]$.

For any given propostion, independence excludes the use of any information about the individual support for other propositions. Thus, it also seems natural to introduce a monotonicity condition which guarantees that the social acceptance of a proposition is not reduced by increased individual support for it:

Definition 9 *A judgment aggregation rule $f : D^n \to D$ is* ***monotonic*** *if for any profiles $\underline{A}, \underline{A}' \in D^n$ and any proposition $p \in X$,*
$p \in f(\underline{A}) \wedge \underline{A}(p) \subset \underline{A}'(p) \Rightarrow p \in f(\underline{A}')$.

Observe that for any monotonic judgment aggregation rule $f : D^n \to D$ and any proposition $p \in X$, the family $\mathcal{W}_p$ of all decisive sets for proposition p is closed under supersets (i.e. $U \in \mathcal{W}_p \wedge U \subset V \Rightarrow V \in \mathcal{W}_p$).

From the combination of these properties follow two other conditions which are of importance for the proof, but also of significance on their own:

(i) Together with monotonicity, independence implies a strong form of decisiveness.

Definition 10 *A judgment aggregation rule $f : D^n \to D$ satisfies* ***strong decisiveness*** *if for any proposition $p \in X$ and any profile $\underline{A} \in D^n$,*
$p \in f(\underline{A}) \Rightarrow \underline{A}(p) \in \mathcal{W}_p$.

[3] In the standard formulation the judgment aggregation rule f is independent if for any proposition $p \in X$, and any profiles $\underline{A}, \underline{A}' \in D^n$, $[\underline{A}'(p) = \underline{A}(p)] \Rightarrow [p \in f(\underline{A}) \Leftrightarrow p \in f(\underline{A}')]$.

It should be clear that strong decisiveness excludes any form of inverse decisiveness as e.g. through an inverse dictator.

(ii) Together with monotonicity and sovereignty, independence implies the classical Pareto condition:

Definition 11 *A judgment aggregation rule $f : D^n \to D$ is* ***paretian*** *if for any proposition $p \in X$, $N \in \mathcal{W}_p$.*

From preference aggregation it is well known that Arrovian impossibility results are driven by a "contagion" property (as discussed in Kelly 1988), which propagates the decisiveness of a coalition from a given pair of alternatives to any pair. In judgment aggregation, this contagion property can be captured with the help of the following lemma:

Lemma 12 *$\mathcal{W}_p \subseteq \mathcal{W}_q$ for all propostions p, q such that $(p, q) \in \vdash^*$.*

Proof. Consider a pair of propositions $(p, q) \in \vdash^*$ and a non-empty set P of propositions such that $P \cup \{p, \neg q\}$ is a minimally inconsistent subset with cardinality strictly larger than two and a profile $\underline{A} \in D^n$ such that
(i) $\underline{A}(p) = U \in \mathcal{W}_p$
(ii) $\underline{A}(q) = U$
(iii) $\underline{A}(P) = N$.
By the decisiveness of U for p, $p \in f(\underline{A})$ and, by the Pareto property,
$P \subset f(\underline{A})$. Therefore, $q \in f(\underline{A})$ by minimally inconsistency of $P \cup \{p, \neg q\}$. By independence, for any profile $\underline{A}' \in D^n$ such that $\underline{A}'(q) = \underline{A}(q)$, $q \in f(\underline{A}')$ and hence $U \in \mathcal{W}_q$ by strong decisiveness. ■

Iterated application of this lemma establishes the following neutrality property.[4]

Lemma 13 *For a totally blocked agenda, $\mathcal{W}_p = \mathcal{W}_q$ for all propositions $p, q \in X$.*

Hence, one and the same family $\mathcal{W} \subseteq 2^{|N|}$ of decisive sets of individuals determines the social acceptance of every proposition in a totally blocked agenda.

Similar to the ultrafilter proofs of the Arrow theorem, the impossibility result follows immediately from the fact that an ultrafilter on a finite set of individuals is a family $\mathcal{W} = \{U \subseteq N : i \in U\}$ of all supersets of some singleton set $\{i\}$, which is equivalent to the existence of a dictator.

[4] In our framework, this neutrality property is equivalent to the condition of systematicity (or equal treatment of all propositions) with which the first impossibility result in judgment aggregation was derived (List and Pettit 2002).

Theorem 14 *(Nehring and Puppe 2005) If and only if the agenda is totally blocked, an independent and monotonic judgment aggregation rule which satisfies universal domain, collective rationality and sovereignty is dictatorial.*

Our proof exploits the relation between ultrafilters and collections of decisive sets of individuals known as simple games (Neumann and Morgenstern 1944, for a recent reference see Taylor and Zwicker 1999).

Definition 15 *A* ***simple game*** *on the set N of individuals is a collection $\mathcal{W} \subseteq 2^{|N|}$ of subsets of N which satisfies closure under supersets (i.e. $U \in \mathcal{W} \wedge U \subset V \Rightarrow V \in \mathcal{W}$).*
A simple game is ***strong*** *if for any $U \subseteq N$, $U \notin \mathcal{W} \Rightarrow N \backslash U \in \mathcal{W}$.*

Simple games stand in a close relation to ultrafilters which, explicitly following Guilbaud 1952, was established by Monjardet 1978:

Lemma 16 *A strong simple game is an ultrafilter if for any $U, V, W \in \mathcal{W}$, $U \cap V \cap W \neq \emptyset$.*

Proof. (If-part) From Lemma 13 we know that the collection $\mathcal{W} \subseteq 2^{|N|}$ of the decisive sets of individuals is the same for all propositions in a totally blocked agenda. The proof then proceeds by establishing that any such collection $\mathcal{W}$ is
(i) a strong simple game, which is
(ii) an ultrafilter.
(i) To see that any collection $\mathcal{W}$ of decisive sets of individuals is a simple game keep in mind that monotonicity of the judgment aggregation rule implies the closure of $\mathcal{W}$ under supersets. To see that the simple game $\mathcal{W}$ is strong consider that for any profile $\underline{A} \in D^n$, any non-decisive set of individuals $U \notin \mathcal{W}$ and any proposition $p \in X$, $\underline{A}(p) = U$ implies $p \notin f(\underline{A})$ and hence by completeness of the agenda, $\neg p \in f(\underline{A})$, which in turn implies $\underline{A}(\neg p) = N \backslash U \in \mathcal{W}$ (by strong decisiveness).
(ii) To see that the simple game $\mathcal{W}$ is an ultrafilter, verify that for any decisive sets of individuals $U, V, W \in \mathcal{W}$, it must be the case that $U \cap V \cap W \neq \emptyset$. Otherwise, a profile $\underline{A} \in D^n$ can be constructed such that for a pair of propositions $(p, q) \in \vdash^*$ and a minimally inconsistent superset S of $\{p, \neg q\}$ with cardinality strictly larger than two, $\underline{A}(p) = U$, $\underline{A}(S \backslash \{p, \neg q\}) = V$, and $\underline{A}(\neg q) = W$ which contradicts the collective rationality of the aggregation rule.

For the only if-part, consider that an agenda which is not totally blocked can be partitioned into totally blocked subsets such that to different elements of the partition different (local) dictators are assigned. ■

3 Discussion

The ultrafilter proof strategy of impossibility results in judgment aggregation shows that it essentially is the logical structure of the agenda of the collective decision problem which determines the distribution of decisiveness and thus drives the dictatorship result. This is a serious challenge for justifications of democracy which are based on the epistemic properties of the outcome of democratic procedures (like e.g. Estlund, 2008), since the elementary rationality condition of the logical consistency of the collective outcome conflicts with the weakest possible fairness principle for the decision procedure (non-dictatorship). Thus, it might well be that also from an epistemic point of view, democratic legitimacy has to be sought at a purely procedural level (see e.g. Peter 2007).

References

[1] Arrow, K. (1964) *Social Choice and Individual Values,* New York: Wiley

[2] Daniels, T.A. (2006) "*Social Choice and the Logic of Simple Games"* in *Proceedings of the 1st International Workshop on Computational Social Choice (COMSOC-2006)* by U. Endriss and J. Lang, Eds., Universiteit van Amsterdam: Amsterdam, 125-138.

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

[4] Dietrich, F., and Ph. Mongin (2007) "The Premiss-Based Approach to Judgment Aggregation" *mimeo.*

[5] Dokow, E., and R. Holzman (2005) "Aggregation of binary evaluations" *Journal of Economic Theory* forthcoming.

[6] Estlund, D.M. (2008) *Democratic Authority: A Philosophical Framework*, Princeton University Press.

[7] Gärdenfors, P. (2006) "An Arrow-like theorem for voting with logical consequences" *Economics and Philosophy* **22**, 181-190.

[8] Guilbaud, G. Th. (1952) "Les théories de l'intérêt général et le problème logique de l'agrégation" *Economie Appliquée* **15**, 502-584.

[9] Kelly, J. (1988) *Social Choice Theory: An Introduction*, Springer: Berlin.

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

[11] List, C., and C. Puppe (2007) "Judgment aggregation: a survey" *mimeo,* forthcoming in *The Oxford Handbook of Rational and Social Choice* by Anand, P., P.K. Pattanaik, and C Puppe, Oxford University Press

[12] Monjardet, B. (1978) "Une autre preuve du théorème d'Arrow" *R.A.I.R.O. Recherche Opérationelle* **12**, 291-296.

[13] Monjardet, B. (1983) "On the use of ultrafilters in social choice theory" in *Social Choice and Welfare* by Pattanaik, P.K., and M. Salles, Eds., North Holland: Amsterdam, 73-78.

[14] Monjardet, B. (2003) "De Condorcet à Arrow via Guilbaud, Nakamura et les 'jeux simples'" *Mathématiques et Sciences Humaines* **41**,163, 5-32.

[15] Monjardet, B. (2005) "Social choice theory and the "Centre de Mathématique Sociale": some historical notes" *Social Choice and Welfare* **25**, 433-456.

[16] Nehring, K., and C. Puppe (2005) "The Structure of Strategy-Proof Social Choice. Part II: Non-Dictatorship, Anonymity and Neutrality" *mimeo.*

[17] Nehring, K., and C. Puppe (2002) "Strategy-Proof Social Choice on Single-Peaked Domains: Possibility, Impossibility and the Space Between" *mimeo.*

[18] von Neumann, J., and O. Morgenstern (1944) *Theory of Games and Economic Behavior*, Princeton University Press.

[19] Peter, F. (2007) "Democratic legitimacy and proceduralist social epistemology" *Politics, Philosophy & Economics* **6**, 329-353.

[20] Taylor, A.D. and W.S. Zwicker (1999) *Simple Games: Desirability Relations, Trading, Pseudoweightings*, Princeton University Press.