Reason-Based Group Choice under Partial Consensus

Klaus Nehring
Department of Economics, University of California at Davis
Davis, CA 95616, U.S.A.
kdnehring@ucdavis.edu

and

Clemens Puppe
Department of Economics, University of Karlsruhe
D – 76128 Karlsruhe, Germany
puppe@wior.uni-karlsruhe.de

Preliminary version, only for use at the
Cowles Foundation Workshop on
Aggregation of Opinions
September 2006

Please do not quote
Comments welcome

1 Introduction (preliminary sketch)

The famous discursive dilemma (Kornhauser and Sager (1986)) shows that reasonable aggregation rules such as majority voting may lead to inconsistencies in contexts in which a group has to take a decision based on a pre-specified set of criteria, or *reasons*. The first formal impossibility result was proved by List and Pettit (2002)). Several recent papers (see ...) have demonstrated the robustness of the impossibility underlying the discursive dilemma, both with respect to the class of admissible aggregation rules and with respect to the structure of the logical relation between the "decision" and the "reasons." For instance, in Nehring and Puppe (2006b), we have shown that the discursive dilemma extends to all "truth-functional" contexts, i.e. to situations in which the binary decision is completely determined by a judgement on the criteria.

In this paper, we relax the assumption of truth-functionality, allowing for the possibility that an evaluation of all criteria, or reasons, does not necessarily by itself imply a specific decision. As an example, consider the publication process in an academic journal. The binary decision is whether or not to publish a given paper. Suppose that the criteria according to which the referees evaluate the paper are correctness of results, writing style, depth of analysis, and originality of ideas. For simplicity, assume that the criteria are also dichotomous: either the results are correct or not, the writing style is sufficiently clear or not, etc. Suppose that the referees are asked to submit a yes/no-judgement on the four criteria and on the publishability of the paper. Truth-functionality means that, by some commonly agreed upon exogenous rule (a "consensus"), any complete evaluation of the four criteria automatically implies a unique publication recommendation. However, one may well want to expand the range of admissible judgements by allowing for a *partial consensus*, i.e. a partial specification of how the evaluation of the criteria are linked to the publication recommendation.

For instance, it may be commonly agreed upon that both correctness of results and good writing style are necessary for a positive publication decision, and that a positive evaluation on all four criteria is sufficient for publishability. On the other hand, one may wish not to exogenously fix the publication recommendation for papers that are positive on all criteria but on depth of analysis, say. Thus, either a positive and a negative publication recommendation may be considered to be consistent with the affirmation of all criteria except depth of analysis.

We show that relaxing truth-functionality in this manner opens up new possibilities of independent[1] aggregation of individual judgements. In particular, it allows for aggregation without veto power. This is in sharp contrast to the truth-functional case in which all independent aggregation rules entail veto power for at least one individual and are typically even dictatorial, as shown in Nehring and Puppe (2006b). The extent to which one can escape the impossibility results of the truth-functional case depends on the structure of the set of those criteria that are sometimes necessary for the decision in relation to the set of criteria that are sometimes sufficient for the decision.

Specifically, say that a criterion is *relevant for the acceptance* of the decision if there is a combination of all other criteria such that the negation of the criterion in question can turn a necessary acceptance into a rejection. Similarly, say that a criterion is *relevant for the rejection* of the decision if there is a combination of all other criteria such that the affirmation of the criterion in question can turn a necessary rejection into an acceptance. We prove that there exist anonymous and independent aggregation rules without veto power if and only if no criterion is simultaneously relevant for the acceptance and the rejection of the decision. Moreover, in this case the aggregation rule can be chosen in such a way that the decision is determined by majority voting. If exactly one criterion is simultaneously relevant for the acceptance and for the rejection of the decision, then there still exist anonymous rules with majority voting on the decision (but any such rule entails a veto on some of the criteria). If all criteria are simultaneously relevant for the acceptance and for the rejection of the decision, then all aggregation rules are oligarchic, and typically even dictatorial.[2]

2 Notation and Definitions

The set $C = \{c_1, ...c_m\}$ represents a collection of *criteria* ("reasons") for a binary decision d. The elements of $C \cup \{d\}$ are also referred to as *propositions*. An (individual) *judgement* is a subset $J_i \subseteq C \cup \{d\}$ with the interpretation that J_i contains exactly those propositions affirmed by individual i. In particular, $d \in J_i$ (resp. $d \notin J_i$) means that i would like to see d accepted (resp. rejected). Throughout, we assume that any combination of affirmed criteria is logically possible.

There is unanimous agreement among individuals that some combinations of affirmed criteria lead to the acceptance of d; similarly, some other combinations of affirmed criteria lead to the rejection of d. Formally, denote by $\mathcal{A} \subseteq 2^C$ the collection of subsets $A \subseteq C$ for which it is agreed upon that affirmation of exactly the criteria in A leads to the acceptance of the decision d. We call $\mathcal{A}$ the *acceptance region* of d. Similarly, $\mathcal{R} \subseteq 2^C$ is the collection of subsets $R \subseteq C$ for which it is agreed upon that

[1] Here, independence means the requirement that the social evaluation of each proposition (criteria and decision) only depends on the individual evaluations of *that* proposition.

[2] As stated, the results hold if the partial consensus is *monotone* in a sense made precise below.

affirmation of exactly the criteria in R leads to the rejection of the decision d. We call $\mathcal{R}$ the *rejection region* of d. Henceforth, we refer to a pair $(\mathcal{A}, \mathcal{R})$ with $\mathcal{A} \cap \mathcal{R} = \emptyset$ as a *partial consensus*, or a *partial decision rule*. The consensus resp. the decision rule is termed "partial" because there may be a gap between the acceptance and rejection regions, i.e. there may be combinations of affirmed criteria for which it is neither agreed that they lead to acceptance, nor to rejection of the decision. A judgement J_i is *consistent* with the partial consensus if $J_i \cap C \in \mathcal{A}$ implies $d \in J_i$ and $J_i \cap C \in \mathcal{R}$ implies $d \notin J_i$.

Denote by $N = \{1, ..., n\}$ the set of individuals. A *reason-based aggregation rule* is a mapping F that assigns a consistent collective judgement J to each profile $(J_1, ..., J_n)$ of consistent individual judgments. Throughout, we require the following properties. First, F is defined for all logically possible combinations of consistent individual judgments ("unrestricted domain"). Second, any consistent judgement J is the collective result of F for some suitable profile of individual judgements ("voter sovereignty").[3] Finally, we require the following condition of "monotone independence." Consider $(J_1, ..., J_n)$ and $(J'_1, ..., J'_n)$ such that, for some p and all i, $p \in J_i \Rightarrow p \in J'_i$. Then, $p \in F(J_1, ..., J_n) \Rightarrow p \in F(J'_1, ..., J'_n)$. Monotone Independence thus asserts that if the collective judgement entails p, and if the individual support for p increases, then p must remain in the collective judgement.

The partial decision rule is called *monotone* if $\mathcal{A}$ is closed under taking supersets and $\emptyset \notin \mathcal{A}$, and if $\mathcal{R}$ is closed under taking subsets and $C \notin \mathcal{R}$.[4] Monotone decision rules are by far the most relevant ones. Intuitively, they correspond to contexts in which the affirmation of a criterion always has a positive effect on the decision, no matter what other criteria are affirmed. However, although many rules can be assumed to be monotone by choosing the appropriate labeling of the criteria and their respective negations, monotonicity is not without any loss of generality. For instance, a partial consensus in which d is accepted if and only if two criteria are either both affirmed or both negated is not monotone.

The case in which any combination of affirmed criteria either leads to the acceptance or to the rejection of the decision, i.e. the case $\mathcal{A} \cup \mathcal{R} = 2^C$ is referred to as the *truth-functional* case.

By $C_\mathcal{A} \subseteq C$ we denote the set of criteria that are *relevant for the acceptance* of d, i.e. $c \in C_\mathcal{A}$ if and only if there exists a set $A \ni c$ such that $A \in \mathcal{A}$ and $A \setminus \{c\} \notin \mathcal{A}$. Similarly, by $C_\mathcal{R} \subseteq C$ we denote the set of criteria that are *relevant for the rejection* of d, i.e. $c \in C_\mathcal{R}$ if and only if there exists a set $R \not\ni c$ such that $R \cup \{c\} \notin \mathcal{R}$ and $R \in \mathcal{R}$. Without loss of generality, we assume that each criterion is relevant for the acceptance, or relevant for the rejection of d, or both.[5]

Throughout, we assume that d is not logically equivalent to any single criterion. In the monotone case, this is equivalent to the requirement that for no c, $(\{c\} \in \mathcal{A}$ and $C \setminus \{c\} \in \mathcal{R})$.

[3] A slightly stronger condition ("unanimity") would require that J be the collective judgement whenever all individuals agree on J.

[4] Note that if $\mathcal{A}$ is closed under taking supersets and $\emptyset \in \mathcal{A}$, then d is always accepted. Similarly, if $\mathcal{R}$ is closed under taking subsets and $C \in \mathcal{R}$, then d is always rejected.

[5] An independent aggregation rule imposes no restrictions on the aggregation of criteria that are neither relevant for the acceptance nor for the rejection of d.

3 Main Result

The following is the central result. Throughout, a partial consensus is assumed to be monotone, and a reason-based aggregation rule is assumed to satisfy unrestricted domain, voter sovereignty and monotone independence. A voter i is said to have a *veto* on proposition p if i can either force p to be affirmed in the social judgement, or force p to be negated in the social judgement. A voter i is said to be a *local dictator* if, for some proposition p, i can enforce p as well as $\neg p$.

Theorem

1. *A partial consensus admits reason-based aggregation rule without veto if and only if no criterion is simultaneously relevant for acceptance and rejection. In this case, the aggregation rule can be chosen to be anonymous with majority voting over the decision.*

2. *If exactly one criterion is both relevant for acceptance and rejection, then there exist anonymous rules with majority voting on the decision (but any such rule entails a veto on some of the criteria).*

3. *If more than one but not all criteria are both relevant for acceptance and rejection, then there exist non-dictatorial rules (but all of them entail a veto on a criterion and/or on the decision). In this case, any local dictator is a dictator on the decision and on the criteria that are both relevant for acceptance and rejection.*

4. *Suppose finally that all criteria are both relevant for acceptance and rejection. In this case, all aggregation rules are oligarchic (in particular, any aggregation rule entails a veto on the decision). There exist anonymous rules if and only if either the acceptance region contains any non-empty subset of criteria, or the rejection region contains any proper subset of criteria. In all other cases, all aggregation rules are dictatorial.*

4 Proof

The proof of the above result relies on our general characterization of monotonely independent aggregation rules in terms of the Intersection Property (see Nehring and Puppe (2006a)), as follows. A family of *winning coalitions* is a non-empty family $\mathcal{W}$ of subsets of the set N of all individuals satisfying $[W \in \mathcal{W}$ and $W' \supseteq W] \Rightarrow W' \in \mathcal{W}$. Denote by Z the set of all propositions, i.e. $Z = C \cup \{d\}$, and by Z^* the *negation closure* of Z, i.e. $Z^* := Z \cup \{\neg p : p \in Z\}$ where $\neg p$ denotes the negation of p. A *structure of winning coalitions* on Z^* is a mapping $p \mapsto \mathcal{W}_p$ that assigns a family of winning coalitions to each proposition $p \in Z^*$ satisfying the following condition,

$$W \in \mathcal{W}_p \Leftrightarrow (N \setminus W) \notin \mathcal{W}_{\neg p}. \tag{4.1}$$

In words, a coalition is winning for p if and only if its complement is not winning for the negation of p. A reason-based aggregation rule F is called *voting by issues*, or, in our context simply *propositionswise voting*, if for some structure of winning coalitions and all $p \in Z^*$,

$$p \in F(J_1, ..., J_n) \Leftrightarrow \{i : p \in J_i\} \in \mathcal{W}_p.$$

Observe that, so far, nothing guarantees that the outcome judgement $F(J_1, ..., J_n)$ of propositionwise voting is consistent. The necessary and sufficient condition for consistency can be described as follows.

A *critical family* is a minimal subset $Q \subseteq Z^*$ of propositions that is logically inconsistent. The sets $\{p, \neg p\}$ are called *trivial* critical families. A structure of winning coalitions satisfies the *Intersection Property* if for any critical family $\{p_1, ..., p_l\} \subseteq Z^*$, and any selection $W_j \in \mathcal{W}_{p_j}$,

$$\bigcap_{j=1}^{l} W_j \neq \emptyset.$$

In Nehring and Puppe (2006a, Theorem 3), we have shown that a reason-based aggregation rule satisfies unrestricted domain, voter sovereignty and monotone independence if and only if it is propositionwise voting satisfying the Intersection Property.

Using (4.1) and the fact that families of winning coalitions are closed under taking supersets, we obtain

$$\mathcal{W}_{\neg p} = \{W \subseteq N : W \cap W' \neq \emptyset \text{ for all } W' \in \mathcal{W}_p\}. \tag{4.2}$$

The following *conditional entailment relation* plays a central role. For all $p, q \in Z^*$,

$$p \geq^0 q \Leftrightarrow [p \neq \neg q \text{ and there exists a critical family containing } p \text{ and } \neg q]. \tag{4.3}$$

By $\geq$ we denote the transitive closure of $\geq^0$, and by $\equiv$ the symmetric part of $\geq$. Note that $\geq$ is "negation adapted" in the sense that $p \geq q \Leftrightarrow \neg q \geq \neg p$.

The following two lemmas are proved in Nehring and Puppe (2005).

Lemma 1 *Suppose that a structure of winning coalitions satisfies the Intersection Property. Then,* $p \geq q \Rightarrow \mathcal{W}_p \subseteq \mathcal{W}_q$.

Lemma 2 *Suppose that a structure of winning coalitions satisfies the Intersection Property, and assume that* p, q, r *are contained in some critical family. If* $\mathcal{W}_{\neg p} \subseteq \mathcal{W}_q$, *then* $\{i\} \in \mathcal{W}_{\neg r}$, *for some* $i \in N$.

The following lemma characterizes the conditional entailment relation for a monotone partial consensus.

Lemma 3 *For all* $c \in C$, $(c \in C_{\mathcal{A}} \Leftrightarrow c \geq d)$ *and* $(c \in C_{\mathcal{R}} \Leftrightarrow d \geq c)$.

Proof of Lemma 3 Let $c \in C_{\mathcal{A}}$ and consider a minimal subset $A \in \mathcal{A}$ with $c \in A$ such that $A \setminus \{c\} \notin \mathcal{A}$. Note that by monotonicity $A \neq \emptyset$. Since $A \in \mathcal{A}$, $A \cup \{\neg d\}$ forms an inconsistent family of propositions. By consistency of A and minimality of $A \in \mathcal{A}$, $A \cup \{\neg d\}$ is minimally inconsistent, i.e. a critical family; thus, $c \geq d$.

Now suppose conversely that $c \geq d$, i.e. that there exists a critical family $P \subseteq Z^*$ containing c and $\neg d$. Since P is inconsistent, we have $P \cap C \in \mathcal{A}$; thus, by criticality of P, we obtain that c is relevant for the acceptance of d.

The proof of the second part is analogous. If $c \in C_{\mathcal{R}}$, consider a maximal $R \in \mathcal{R}$ with $R \not\ni c$. By monotonicity of the partial consensus, $\{\neg a : a \notin R\} \cup \{d\}$ forms a critical family, thus in particular, $d \geq c$. Conversely, if $d \geq c$ consider a critical family $P \subseteq Z^*$ containing $\neg c$ and d. Then, $(C \setminus \{c : \neg c \in P\}) \in \mathcal{R}$, and thus by criticality, $c \in C_{\mathcal{R}}$.

Lemma 4 *For all* $c \in C$, $c \equiv d \Leftrightarrow c \in C_{\mathcal{A}} \cap C_{\mathcal{R}}$.

Proof of Lemma 4 In view of Lemma 3, it is sufficient to verify that, for no $c \in C$, there is a critical family containing c and d, and that, for no $c \in C$, there is a critical family containing $\neg c$ and $\neg d$. To verify this, suppose, by contradiction, that $P \subseteq Z^*$ is a critical family containing c and d. Then, $(P \cap C) \in \mathcal{R}$ but $[(P \cap C) \setminus \{c\}] \notin \mathcal{R}$, in violation of monotonicity. Similarly, if a critical family P contains $\neg c$ and $\neg d$, then $(P \cap C) \in \mathcal{A}$ and $(P \cap C) \cup \{c\} \notin \mathcal{A}$, again in violation of monotonicity of the partial consensus.

Note that by Lemmas 3 and 4, there are only two types of non-trivial critical families. All critical families that contain d are of the form $\{d, \neg c_1, \neg c_2, ..., \neg c_l\}$ with $c_j \in C_{\mathcal{R}}$ for all $j \in \{1, ..., l\}$, and all critical families that contain $\neg d$ are of the form $\{\neg d, c_1, c_2, ..., c_k\}$ with $c_j \in C_{\mathcal{A}}$ for all $j \in \{1, ..., k\}$.

We are now ready for the proof of the main result.

Proof of Theorem

1. First, suppose that no criterion is simultaneously relevant for acceptance and rejection. Then, $\{C_{\mathcal{A}}, C_{\mathcal{R}}\}$ forms a partition of C. Let $m_{\mathcal{A}}$ be the largest cardinality of a critical family containing some element of $C_{\mathcal{A}}$, and let $m_{\mathcal{R}}$ be the largest cardinality of a critical family containing some element of $C_{\mathcal{R}}$. Using the (anonymous version) of the Intersection Property (see Nehring and Puppe (2006a, Fact 3.4)), it is easily verified that a quota rule that assigns the quota 1/2 to d, the quota $\frac{2m_{\mathcal{A}}-3}{2m_{\mathcal{A}}-2}$ to all $c \in C_{\mathcal{A}}$, and the quota $1 - \frac{2m_{\mathcal{R}}-3}{2m_{\mathcal{R}}-2}$ to all $c \in C_{\mathcal{R}}$ is consistent. If there are sufficiently many individuals, no single voter has a veto.

Now suppose that there is one criterion c that is simultaneously relevant for acceptance and rejection. By Lemma 3, there exists a critical family P containing c and $\neg d$, and another critical family P' containing $\neg c$ and d. By Lemmas 1 and 3, we have $\mathcal{W}_c = \mathcal{W}_d$ and $\mathcal{W}_{\neg c} = \mathcal{W}_{\neg d}$. One of the critical families P or P' must contain at least three elements, since otherwise c and d would be logically equivalent. Without loss of generality, assume $c' \in P$ for some $c' \neq c$. By Lemma 2, for some i, $\{i\} \in \mathcal{W}_{\neg c'}$, i.e. i has a veto.

2. Suppose that $\{c\} = C_{\mathcal{A}} \cap C_{\mathcal{R}}$. We have just shown that necessarily at least one voter has a veto. Using the Intersection Property, the following rule is easily seen to be anonymous and consistent: assign quota 1/2 to d and c, quota 1 to all criteria in $C_{\mathcal{A}} \setminus \{c\}$ and quota 0 to all criteria in $C_{\mathcal{R}} \setminus \{c\}$. In fact, this is the only anonymous rule with majority voting on d. Indeed, by Lemma 1, the quotas for d and c must necessarily be the same; by consequence, majority voting on d forces that all criteria different from c can be affirmed, resp. rejected, only under unanimous consent.

3. Suppose that $C_{\mathcal{A}} \cap C_{\mathcal{R}} = \{c_1, ..., c_k\} \neq C$. Without loss of generality, assume that there exists $c \in C_{\mathcal{A}} \setminus \{c_1, ..., c_k\}$. The aggregation rule according to which $\mathcal{W}_c = \{N\}$ (i.e. affirmation of c requires unanimous consent), and $\{i\} \in \mathcal{W}_p \cap \mathcal{W}_{\neg p}$ for all $p \notin \{c, \neg c\}$ (i.e. voter i is a local dictator on all other propositions) is non-dictatorial and consistent.

If i is a local dictator, then by Lemma 3, i is a dictator over d, and by Lemma 4, also over any criterion that is both relevant for acceptance and rejection of d.

4. By the argument in the first part of this proof, we have $\{i\} \in \mathcal{W}_{\neg c'}$ for some $c' \in C_{\mathcal{A}}$ if there exists a critical family containing $\neg d$ and at least three elements. By assumption, $C_{\mathcal{A}} = C$, hence by Lemma 3, $\{i\} \in \mathcal{W}_{\neg c} = \mathcal{W}_{\neg d}$ for all $c \in C$. Similarly, we obtain $\{i\} \in \mathcal{W}_{c'}$ for some $c' \in C_{\mathcal{R}}$, and thus for all $c' \in C$, if there exists a critical family containing d and at least three elements. Moreover, there is at least one critical family containing at least three elements. Without loss of generality, suppose that $\{i\} \in \mathcal{W}_{\neg c} = \mathcal{W}_{\neg d}$. Define $M := \{h \in N : \{h\} \in \mathcal{W}_{\neg d}\}$. Using the Intersection Property and (4.2), it can be verified that consistency forces the aggregation rule to be oligarchic with oligarchy M (for the precise definition of oligarchic rules, see Nehring and Puppe (2006b)).

Using the Intersection Property once again, it is easily seen from the preceding argument that non-dictatorial rules can only exist if either all critical families containing $\neg d$ have only two elements, or all critical families containing d have only two elements. The first case corresponds to the case $\mathcal{A} = 2^C \setminus \{\emptyset\}$, the second to $\mathcal{R} = 2^C \setminus \{C\}$. In either case, the oligarchic rule can be chosen to be anonymous, i.e. a "unanimity rule."

References

[1] Dietrich, F. (2006), "Judgement Aggregation: (Im)Possibility Theorems," *Journal of Economic Theory* 126, 286-298.

[2] Dietrich, F. and List, C. (2006a), "Arrow's Theorem in Judgement Aggregation," *Social Choice and Welfare*, forthcoming.

[3] Dietrich, F. and List, C. (2006b), "Judgement Aggregation by Quota Rules," *Journal of Theoretical Politics*, forthcoming.

[4] Dokow, E. and Holzman, R. (2005), "Aggregation of Binary Evaluations," *mimeographed.*

[5] Grdenfors, P. (2006), "An Aroow-like Theorem for Voting with Logical Consequences," *Economics and Philosophy* 22, 181-190.

[6] Kornhauser, L. and Sager, L. (1986), "Unpacking the Court," *Yale Law Journal* 96, 82-117.

[7] List, C. (2003), "A Possibility Theorem on Aggregation over Multiple Interconnected Propositions," *Mathematical Social Sciences* 45, 1-13.

[8] List, C. (2004), "A Model of Path Dependence in Decisions over Multiple Propositions," *American Political Science Review*, forthcoming.

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

[10] List, C. and Pettit, P. (2004), "Aggregating Sets of Judgements: Two Impossibility Results Compared," *Synthese*, forthcoming.

[11] Mongin, P. (2005), "Factoring out the Impossibility of Logical Aggregation," *mimeographed.*

[12] Nehring, K. (2003), "Arrow's Theorem as a Corollary," *Economics Letters* 80, 379-382.

[13] Nehring, K. (2006), "The Impossibility of a Paretian Rational ," *mimeographed.*

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

[15] Nehring, K. and Puppe, C. (2005), "Strategy-Proof Social Choice without Dictators," *mimeographed.*

[16] Nehring, K. and Puppe, C. (2006a), "The Structure of Strategy-Proof Social Choice: General Characterization and Possibility Results on Median Spaces," *Journal of Economic Theory*, forthcoming.

[17] Nehring, K. and Puppe, C. (2006b), "Consistent Judgement Aggregation: The Thruth-Functional Case," *mimeographed*

[18] Pauly, M. and van Hees, M. (2006), "Logical Constraints on Judgement Aggregation," *Journal of Philosophical LOgic* 35, 569-585.

[19] Pettit, P. (2001), "Deliberative Democracy and the Discursive Dilemma," *Philosophical Issues* 11, 268-299.