Judgment aggregation: a general theory of collective decisions

Habilitationsschrift

November 2008

Dr. Franz Dietrich

Betreuer: Prof. Dr. Clemens Puppe

Universität Karlsruhe
Fakultät für Wirtschaftswissenschaften

Table of Contents

Chapter 1

An introduction

An introduction

The purpose of this chapter is, first, to give an informal introduction to the judgment aggregation problem; second, to formally define the model of judgment aggregation used in the papers of this thesis; third, to briefly introduce these papers one by one; fourth, to discuss related models and aggregation problems; and fifth, to briefly review the literature. The list of references used in this introductory chapter is included at the chapter end.

1 An informal introduction to the judgment aggregation problem

Collective decisions arise everywhere in modern societies. They arise in the political arena (decisions by parliaments, governments, populations of voters), in the economic arena (decisions by management boards, boards of central banks), in the scientific arena (decision by ethical commissions, expert panels), in the legal arena (courts seeking a verdict), and so on.

In studying collective decision making, one may ask either the empirical question of how groups reach their decisions in reality, or the theoretical (normative) question of how they could or should reach their decisions. We shall take a theoretic focus. The theoretic question of how to reach collective decisions is by no means trivial, let alone because a 'good' collective decision should simultaneously meet two goals: it should suitably reflect the group members' views, and it should be internally coherent. As will be seen, these two goals are often in conflict with each other, and this under several interpretations of the first goal (such as 'one man one vote', or 'listen to the experts') and under several interpretations of the coherence goal (including seemingly undemanding interpretations).

Our theoretical analysis follows the axiomatic approach. This places us in the long tradition of *social choice theory*, going back at least to Arrow's (1951/1963) seminal contribution. In its long history, social choice theory has focussed nearly exclusively on the aggregation of individual preferences over alternatives (candidates, holiday destinations, investment amounts, and so on). A closer look at real-life decisions however reveals that collective decisions do often not take the form of aggregating individual preferences: groups often do not need to rank a set of alternatives but to reach simple 'yes' or 'no' judgments on a set of propositions. For instance, the board of a central bank might need collective positions on the following three propositions:

a :	GDP growth will pick up.
$a \rightarrow b$:	*If* GDP growth will pick up *then* inflation will pick up.
b :	Inflation will pick up.

It is only in recent years that the problem of reaching collective judgments on propositions – *judgment aggregation* – has become studied formally (as discussed later in the literature review). The judgment aggregation problem is genuinely distinct

from (and, as will be seen, more general than) the classic preference aggregation problem. The objects of the decision are arbitrarily interconnected propositions, not mutually exclusive alternatives; and what is being aggregated are sets of judgments, not preference relations.

Just as preference aggregation may lead to acyclic collective preferences if pairwise majority voting is used (*Condorcet's paradox*), so judgment aggregation may lead to logically inconsistent collective judgment sets if propositionwise majority voting is used (the *discursive dilemma*). Various examples of the discursive dilemma could be given; they differ in the type of propositions under consideration. Let me give four such examples now.

Discursive dilemma 1. Assume that in our central bank example the board is split into three camps of equal size, with the following (yes/no) judgments on the propositions.

	a	$a \rightarrow b$	b
1/3 of the board	Yes	Yes	Yes
1/3 of the board	No	Yes	No
1/3 of the board	Yes	No	No
Collective using majority rule	Yes	Yes	No

Note that each camp holds a consistent set of judgments: the first camp believes in growth (a), in growth causing inflation ($a \rightarrow b$), and accordingly in inflation (b); the second camp believes in no growth, in growth causing inflation, and in no inflation; and the third camp believes in growth, in growth not causing inflation, and in no inflation. Yet the propositionwise majority judgments are inconsistent: a majority believes in growth, a majority believes in growth causing inflation, but a majority believes in no inflation.

Of course, this discursive dilemma does not depend on the kind of group and the meaning attached to a and b (and hence to $a \rightarrow b$). Alternatively, an ethical commission might need judgments on a : 'a multi-cultural society is viable', b : 'immigration should continue', and the implication $a \rightarrow b$; or, a government might need judgments on a : 'law x reduces crime', b : 'law x should be introduced', and the implication $a \rightarrow b$; and so on.

Discursive dilemma 2. As an example with a different syntactic structure of propositions, the supervisory board of a loss-making company might debate the following propositions:

a : A factory should be closed down.
b : A new factory should be created.
$a \wedge b$: A factory should be closed down *and* another created.

Suppose the supervisory board consists of three members with the following judgments. (One might imagine that member 1 represents shareholders and wants to restructure, member 2 represents a risk-averse creditor and wants to down-size, and

member 3 represents workers and wants to create jobs.)

	a	b	$a \wedge b$
Board member 1	Yes	Yes	Yes
Board member 2	Yes	No	No
Board member 3	No	Yes	No
Collective using majority rule	Yes	Yes	No

Here again, each board member holds a consistent set of judgments but the set of majority judgments is inconsistent.

Discursive dilemma 3: the doctrinal paradox. The historically first example given of a discursive dilemma is the so-called *doctrinal paradox* in law. According to legal doctrine in case law systems, two conditions are necessary and sufficient for liability of (say) a firm to (say) pay damages for breach of contract. The first condition is that the contract in question was indeed legally valid, and the condition is that the firm has indeed acted against ('broken') the contract. More formally, suppose that the jury in a trial against a firm needs collective judgments on four propositions:

a : The contract was legally valid.
b : The firm has acted against the contract.
$c \leftrightarrow (a \wedge b)$: The firm is liable *if and only if* the contract was legally valid *and* was acted against.
c : The firm is liable.

Imagine that the jury consists of three jurors, who hold the following judgments:

	a	b	$c \leftrightarrow (a \wedge b)$	c
Juror 1	Yes	Yes	Yes	Yes
Juror 2	Yes	No	Yes	No
Juror 3	No	Yes	Yes	No
Collective under majority rule	Yes	Yes	Yes	No

So, all jurors accept the legal doctrine $c \leftrightarrow (a \wedge b)$, whereas they disagree on the three atomic propositions (a, b and c). The majority set of judgments is again logically inconsistent.

Discursive dilemma 4: Condorcet's paradox. Suppose three individuals face three collective alternatives x, y, z (e.g. three holiday destinations, job candidates, and so on). The individuals need to know which alternatives are better than which other alternatives. Suppose that on the three propositions

xPy : x is better than y
yPz : y is better than z
zPx : z is better than x

the individuals hold the following judgments:

	xPy	yPz	zPx
Individual 1 (who prefers x to y to z)	Yes	Yes	No
Individual 2 (who prefers y to z to x)	No	Yes	Yes
Individual 3 (who prefers z to x to y)	Yes	No	Yes
Collective using majority rule	Yes	Yes	Yes

Although each individual holds acyclic betterness judgments, the collective betterness judgments are cyclic. 'Acyclicity' is a form of logical inconsistency (in a suitably defined logic; see Section 2).

Although propositionwise majority rule is prima facie very appealing – it seems to be the most democratic way to form collective judgments – its failure to ensure consistent outcomes is enough of a reason to look for alternatives. In the judgment aggregation literature, four alternatives aggregation rules have received particular attention: the premise-based rule, the conclusion-based rule, quota rules, and distance-based rules. Let me illustrate these aggregation rules using the introductory example with propositions a, $a \rightarrow b$, b. (The rules could also be defined for the other agendas of propositions discussed in Discursive Dilemmas 2-4).

	a	$a \rightarrow b$	b
1/3 of the board	Yes	Yes	Yes
1/3 of the board	No	Yes	No
1/3 of the board	Yes	No	No
Collective using majority rule	Yes	Yes	No
Collective using premise-based rule	Yes	Yes	Yes
Collective using conclusion-based rule			No
Collective using the quota rule with quota $\frac{3}{4}, \frac{3}{4}, \frac{1}{2}$	No	No	No
Collective using the (simplest) distance-based rule	tie	tie	tie

Table 1: A simple judgment aggregation problem and five aggregation rules

The premise-based rule. To define this aggregation rule, one first has to classify the propositions under consideration – in our case, a, $a \rightarrow b$ and b – into *premise* propositions and *non-premise* ('*conclusion*') propositions. What exactly counts as a premise is a matter of interpretation. Typically, the premises are the propositions deemed semantically more fundamental, or the propositions that constitute reasons for (or against) other propositions, or simply the propositions one wishes to prioritise in the aggregation procedure. In our example, it is presumably most natural to count a and $a \rightarrow b$ as the premise propositions and b as the conclusion proposition. The premise-based rule consists in taking a majority vote only on each premise proposition and to decide the conclusion proposition by deductive entailment. More precisely, a premise proposition (a or $a \rightarrow b$) is collectively affirmed if and only if it is majority affirmed, and the conclusion proposition b is collectively affirmed if and only if b follows from the collective judgments on the premise propositions, i.e. if and only if a and $a \rightarrow b$ are each majority affirmed. In Table 1, the premise-based rule leads the collective to affirm a and $a \rightarrow b$, and hence to also affirm b. Some remarks are due:

- The premise-based rule is sensitive to the choice of premises, which opens up possibilities for manipulation.
- The premise-based rule is perfectly 'democratic' (in the majoritarian sense) on each premise, but it may overrule the majority judgment on the conclusion.
- It may happen that the majority judgments on the premises underdetermine the decision on the conclusion; for instance, if a majority negates a and affirms $a \rightarrow b$, then neither b nor its negation follows. There are different variants of how to define the rule's decision on b in such a case: under one variant

the collective makes no judgment on b (a form of incompleteness of collective judgments), under another b is negated (which avoids incompleteness at the cost of building a bias against b into the aggregation rule).

- The premise-based rule is strategically manipulable by voters, assuming that the voters have outcome-oriented preferences, i.e. care only about how the conclusion proposition is judged by the collective (see Chapter 4).

A generalised analysis of premise-based aggregation is given in Chapter 7 and in Dietrich and Mongin (2007).

The conclusion-based rule. Like the premise-based rule, this rule presupposes having classified the propositions into premise propositions and conclusion propositions. Again, I suppose that a and $a \rightarrow b$ count as premises and b as the conclusion. Under the conclusion-based rule, the conclusion b is decided by a majority vote on b (which in Table 1 leads to collective negation of b), and no collective judgment is made on the premises a and $a \rightarrow b$. Again, some remarks are due:

- The conclusion-based rule is, like the premise-based rule, sensitive the choice of what counts as a premise, what as a conclusion.
- The rule is 'democratic' (in a majoritarian sense) on the conclusion.
- The rule does not generate a justification for the collective judgment on the conclusion, in that each premise is neither affirmed nor negated: the central bank board judges that inflation will not pick up while leaving open whether this is because GDP growth will not pick up or because growth does not imply inflation. Whether such 'undertheorised collective judgments' pose a problem depends on the context. If the role of the central bank is seen solely in setting the interest rates, it might be enough to come to a collective judgment on b (i.e. on inflation); but if the central bank must, for instance, give public justifications for its measures, it presumably has to make up its mind on a and on $a \rightarrow b$. The question of whether a democratic society needs collective judgments on premise propositions (and hence whether conclusion-based voting should be used) is ultimately a democracy-theoretic question; it depends on whether a collective as a whole ought to, or ought not to, take positions on fundamental issues that can serve as underlying reasons or justifications for concrete measures without being strictly needed for such collective measures (see List 2006).

Quota rules. Quota rules represent a different deviation from majoritarian democracy than premise- or conclusion-based rules. While premise- or conclusion-based rules (in a sense) retain the idea of majoritarianism but restrict it to particular propositions (the premises a and $a \rightarrow b$ resp. the conclusion b), quota rules retain the idea of deciding each proposition by taking an isolated vote on this proposition but replace the majoritarian quota by some (possibly higher or lower) quota. Formally, let each proposition p be endowed with a quota q_p in $[0, 1]$ and let the collective affirm p if and only if at least a proportion of q_p of the voters affirms p (and negate p otherwise). In the example of Table 1, a has quota $\frac{3}{4}$, $a \rightarrow b$ has quota $\frac{3}{4}$, and b has quota $\frac{1}{2}$, and this leads the collective to negate each of a, $a \rightarrow b$ and b. Again, some comments are due:

- Quota rules may obviously be biased in favour of a proposition (if its acceptance quota is small) or against it (if its acceptance quota is high).

- Quota rules are not strategically manipulable by voters, under suitable assumptions on voters' preferences. See Chapter 4 and Dietrich and List (2007b).
- Whether a quota rule guarantees consistent collective judgments depends on how the quota were specified. How exactly the quota may be set to guarantee consistency depends in a systematic way on how the propositions are logically interconnected, as is shown by Nehring and Puppe (2002/2007b).
- Do the quota specifications in Table 1 (namely $q_a = \frac{3}{4}$, $q_{a \to b} = \frac{3}{4}$, $q_b = \frac{1}{2}$) guarantee consistent collective judgments? As just mentioned, this depends on how a, $a \to b$ and b are logically interconnected. As it turns out, these interconnections are not obvious, as they depend on the interpretation of the conditional '$\to$'. Formal logic has to offer different interpretations of conditionals: material, indicative, subjunctive, strict. The outcome in Table 1 (namely: negating all of a, $a \to b$, b) is inconsistent if '$\to$' denotes a material conditional, but consistent if '$\to$' denotes a subjunctive or strict conditional. Chapter 6 is devoted to consistent judgment aggregation on conditional propositions.

Distance-based rules. Under this entirely different approach, the collective judgments are made in such a way as to minimize the 'distance' to the individual judgments subject to the constraint of the collective judgments being consistent and complete (Konieczny and Pino-Perez 2002 and Pigozzi 2006). More precisely, let $d : \mathbf{A} \times \mathbf{A} \to \mathbf{R}$ be a metric (distance function) on the set $\mathbf{A}$ of sets of judgments. Given any combination of individual sets of judgments (A_i) (with i ranging over the set of individuals), the collective set of judgments A is chosen so as to minimise the sum-total distance $\sum_i d(A, A_i)$ subject to A being consistent and complete. In Table 1, members i of the first, second and third camp have judgment set $A_i = \{a, a \to b, b\}$, $A_i = \{\neg a, a \to b, \neg b\}$ and $A_3 = \{a, \neg(a \to b), \neg b\}$, respectively. Table 1 moreover assumes that d is the *Kemeny* distance, i.e. that the distance between two judgment sets is the number of propositions on which they disagree. The Kemeny distance is of course familiar from standard social choice; e.g. Baigent (1987). In the example of Table 1, there are three judgment sets A that solve the constrained minimisation problem: $A = \{a, a \to b, b\}$, $A = \{\neg a, a \to b, \neg b\}$ and $A = \{a, \neg(a \to b), \neg b\}$. In each case, the resulting sum-total distance is

$$\sum_i d(A, A_i) = \frac{n}{3}0 + \frac{2n}{3}2 = \frac{4}{3}n$$

(where n is the number of individuals), because $n/3$ individuals have a judgment set of Kemeny distance 0 to A (namely A itself) and the remaining $2n/3$ individuals have a judgment set of Kemeny distance 2 to A. Some remarks, again:

- Distance-based rules depend on the choice of the metric d, but otherwise they provide a general solution that does not require deciding what propositions count as premises (as in premise- and conclusion-based rules) or what acceptance thresholds to use for particular propositions (as in quota rules).
- Distance-based rules are not democratic in any local sense: the collective may easily negate a proposition that everyone affirms (or vice versa). But they may be viewed as democratic in a holistic or global sense, as the collective judgment *set* as a whole comes maximally close to the individual judgment *sets*.
- The holistic nature of the rule comes with strategic manipulability by voters. Again, see Chapter 4.

- Without adding a tie-braking method, distance-based rules do not define aggregation rules in the standard (single-valued) sense.

As has become obvious, there exist a number of rival procedures (aggregation rules) to solve the group's decision problem, each of which has some initial plausibility but also some drawbacks. The formal literature on judgment aggregation tackles this dilemma by the axiomatic approach. This approach proposes general conditions (axioms) that a 'good' procedure should satisfy; one might require the procedure to preserve unanimous agreements on propositions, to be not strategically manipulable by voters or agenda setters, to incorporate certain individuals' expert information, to treat all individuals equally, and so on. The axiomatic approach then goes on to investigate different combinations of criteria and to derive whether a combination allows for (i) a single possible procedure (the ideal finding), (ii) a set of more than one possible procedures (so that a subsequent choice from this set must be made), or (iii) no possible procedure (so that some criterion must be given up or weakened). The answer to the questions (i)-(iii) depend not just on the criteria under consideration but also on the kind of logical interconnections between the propositions: the weaker these interconnections, the easier it becomes for procedures to meet the creteria. For very little connected propositions, even propositionwise majority rule (which satisfies most criteria one might think of) is possible. The preference aggregation problem (represented as a judgment aggregation problem) is unlucky enough to have highly interconnected propositions. This is the deeper reason for why the literature on preference aggregation is so much dominated by impossibility results (i.e. results of type (iii)). The picture in judgment aggregation theory is less impossibility-dominated: many realistic judgment aggregation problems (if appropriately modelled using logical connectives that are non-classical rather than truth-functional) lead to possibility results (of type (i) or (ii)).

One might speculate over why social choice theory has focussed for half a century nearly exclusively on choosing alternatives rather than on judging propositions. One reason surely is that collective agency has been construed as being closely tied (and limited) to collective *action*, i.e. to the physical implementation of alternatives. A collective action problem is indeed directly solved by having a collective ordering (which is what preference aggregation generates): it then suffices to implement what is ranked highest among what is feasible. By contrast, the path from collective judgments to collective action can be less direct: although collective judgments often lead to, suggest, or justify, certain collective actions, they often do so not in a unique and deterministic way. This is not to say that judgment aggregation problems never deal with action. Indeed, the propositions on which judgments are formed may talk about action. A government might form a collective 'yes' judgment on the proposition 'the health budget should be increased' and thereafter raise the health budget; and a family might form collective judgments of the form 'holidays in country x is better than holidays in country y' (with x and y ranging over some set of potential holiday destinations) and thereafter travel to a destination it judges better than all other destinations.

The philosophically interested reader might have noticed in these two examples that judgment aggregation handles preferences in a way that is non-standard within economics. Taken literally, the government does not express a preference (desire) for

increasing the health budget but a judgment (belief) that it ought to be increased; and the family does not express preferences for certain holiday destinations over others but beliefs that some destinations are better than others. In short, while traditional economic modelling keeps beliefs and preferences strictly separated, judgment aggregation replaces desires with judgments (beliefs) on normative propositions (that adddress goodness, betterness, desirability, and so on).

One might therefore wonder whether judgment aggregation (as far as applied to normative propositions) is implicitly committed to moral realism.[1] I do not think so: moral anti-realists may re-interpret a judgment of a moral proposition as a way to express a desire, e.g. re-interpret 'war is bad' as 'the speaker does not like war'. Such re-interpretation is in line with the (expressivist) view that someone's moral talk expresses desires, not beliefs. I do not take a position here on the moral realism debate; I merely stress that judgment aggregation, though apparently on the realist side, is in fact not committed to either position.

2 A formal model of judgment aggregation

In this section I define the model of judgment aggregation that is introduced in Dietrich (2007) and is the basis of the papers presented later.[2] The model generalises List and Pettit's (2002) model from classic propositional logic to a general logic. It also relates to other models in the literature, as discussed later.

Individuals. We consider a finite set of individuals $N = \{1, 2, \ldots, n\}$ $(n \geq 2)$.

Logic. A *logic* is given by a *language* and a notion of *consistency*. The *language* is a non-empty set $\mathbf{L}$ of sentences (called *propositions*) closed under negation (i.e., $p \in \mathbf{L}$ implies $\neg p \in \mathbf{L}$, where $\neg$ is the negation symbol). For example, in standard propositional logic, $\mathbf{L}$ contains propositions such as a, b, $a \wedge b$, $a \vee b$, $\neg(a \rightarrow b)$ (where $\wedge$, $\vee$, $\rightarrow$ denote 'and', 'or', 'if-then', respectively). In other logics, the language may involve additional connectives, such as modal operators ('it is necessary/possible that'), deontic operators ('it is obligatory/permissible that'), subjunctive conditionals ('if p were the case, then q would be the case'), or quantifiers ('for all/some'). The notion of *consistency* captures the logical connections between propositions by stipulating that some sets of propositions $S \subseteq \mathbf{L}$ are *consistent* (and the others *inconsistent*). For example, in standard logics, $\{a, a \rightarrow b, b\}$ and $\{a \wedge b\}$ are consistent and $\{a, \neg a\}$ and $\{a, a \rightarrow b, \neg b\}$ inconsistent. Most papers of this thesis require only three regularity conditions on the consistency notion:[3]

- (*self-entailment*) Any pair $\{p, \neg p\} \subseteq \mathbf{L}$ is inconsistent.
- (*monotonicity*) Subsets of consistent sets $S \subseteq \mathbf{L}$ are consistent.
- (*completability*) $\emptyset$ is consistent, and each consistent set $S \subseteq \mathbf{L}$ has a consistent superset $T \subseteq \mathbf{L}$ containing a member of each pair $p, \neg p \in \mathbf{L}$.

[1] That is, to the philosophical position whereby moral facts exist.

[2] Because this section explains the model in some detail, I have not included my paper Dietrich (2007) in this thesis.

[3] These are the conditions I1-I3 in Dietrich (2007) ('I' for 'inconsistent'); they are equivalent to the conditions L1-L3 in Dietrich (2007), which are formulated in terms of the entailment relation rather than the consistency notion.

A set of propositions $A \subseteq \mathbf{L}$ *entails* a proposition $p \in \mathbf{L}$ ($A \vdash p$) if $A \cup \{\neg p\}$ is inconsistent.

Two remarks

- Consistency and entailment can be interpreted either syntactically or semantically. A set A syntactically entails a proposition p if p can be proved (derived) from the propositions in A via the logic's allowed rules of deduction (such as, for instance, *modus ponens*); A semantically entails p if, roughly speaking, p is true whenever all members of A are true (where the meaning of 'whenever' depends on the semantics in question; classic propositional logic, for instance, uses truth-functions). The syntactic and semantic notions are formally equivalent if the logic satisfies 'soundness and completeness'. Interpretationally, the difference is significant. The reader should choose whether he prefers to think of consistency and entailment in syntactic or semantic terms; ultimately, this is a question of one's notion of rationality.
- Our definition of entailment from the consistency notion implicitly assumes that the logic is not paraconsistent, because it implies that an inconsistent set entails *all* propositions. For our (non-paraconsistent) logics, the notions of consistency and entailment are interdefinable, i.e. we could alternatively have taken the entailment relation $\vdash$ to be the primitive notion (as most logicians would have done it, and as done in much of my paper Dietrich 2007). If by contrast we wished to allow for paraconsistent logics, we would have had no choice but to make entailment the primitive notion.

Agenda. The *agenda* is the set of propositions on which judgments are to be made. It is a non-empty set $X \subseteq \mathbf{L}$ expressible as $X = \{p, \neg p : p \in X_+\}$ for some set X_+ of unnegated propositions (this avoids double-negations in X). In our introductory example, the agenda is $X = \{a, \neg a, a \to b, \neg(a \to b), b, \neg b\}$.

Judgment sets. A *judgment set* is a set $A \subseteq X$ of propositions in the agenda, interpreted as the propositions that are accepted (believed) by a given agent (an individual or the group agent). A *profile* is an n-tuple $(A_1, \ldots, A_n)$ of judgment sets across individuals. A judgment set is

- *consistent* if it is consistent as given by the logic**;**
- *complete* if it contains at least one member of each pair $p, \neg p \in X$;
- *opinionated* if it contains precisely one member if each pair $p, \neg p \in X$;
- *deductively closed* if it contains all propositions p in X that it entails;
- *fully rational* if it is complete and consistent (and hence also opinionated and deductively closed, as one can show).

For instance, in the case of our introductory agenda $X = \{a, \neg a, a \to b, \neg(a \to b), b, \neg b\}$, the judgment set

- $\{a\}$ is consistent and deductively closed (but not complete, hence also not opinionated);
- $\{a, a \to b\}$ is still consistent but not anymore deductively closed (b is missing);
- $\{a, a \to b, \neg b\}$ is neither consistent nor deductively closed, but opinionated (hence complete);
- $\{a, a \to b, b\}$ is fully rational.

Many papers included in this thesis require both individuals and the collective to hold fully rational judgment sets, a demanding condition that has been dropped in

some recent work (of other authors or myself).

Aggregation rules. A *domain* is a set D of profiles, interpreted as the admissible inputs to the aggregation. An *aggregation rule* is a function F that maps each profile $(A_1, \ldots, A_n)$ in a given domain D to a collective judgment set $F(A_1, \ldots, A_n) = A \subseteq X$. Simple examples of aggregation rules are:

- *dictatorship* by an individual i, given by $F(A_1, ..., A_n) = A_i$ for all profiles $(A_1, ..., A_n)$ in the domain;
- *majority rule*, given by $F(A_1, ..., A_n) = \{p \in X$: more individuals i have $p \in A_i$ than $p \notin A_i\}$ for all profiles $(A_1, ..., A_n)$ in the domain.

There are several conditions (desiderata) that one might require from the aggregation rule. Following List (2001), let me distinguish three types of conditions:

- *Input conditions*, i.e. conditions on the domain of F. In most impossibility theorems of the literature, the aggregation rule is assumed to have the *universal* domain (which consists of all profiles of consistent and complete judgment sets). On suitably restricted domains, one may achieve possibility results, just as in preference aggregation one achieves possibility results on domains of, for instance, single-peaked preferences.
- *Output conditions*, i.e. conditions on collective judgment sets. In particular, one might require the aggregation rule to always generate *consistent* judgment sets, or *complete* judgment sets, or *deductively closed* judgment sets, or combinations thereof. The consistency requirement should arguably always be retained; but the importance of the other conditions is context-specific.
- *Responsiveness conditions*, i.e. conditions on the relationship between inputs and outputs. A minimal requirement is *unanimity-preservation*: $F(A, ..., A) = A$ for every unanimous profile $(A, ..., A)$ in the domain. A central, and far from 'minimal', condition is that of *propositionwise aggregation* or *independence*: for any proposition $p \in X$ and any two profiles in the domain, $(A_1, ..., A_n)$ and $(A'_1, ..., A'_n)$, if all individuals i have $p \in A_i \Leftrightarrow p \in A'_i$ then also the collective has $p \in F(A_1, ..., A_n) \Leftrightarrow p \in F(A'_1, ..., A'_n)$ (i.e., in short, the collective judgment on any $p \in X$ is a function only of the individual judgments on p). In Chapter 7 I come back to this demanding condition (and replace it by the more flexible condition of *independence of irrelevant information*).

3 The topic of each paper

It is worth giving a brief informal introduction to each of the papers presented in the following chapters.

Chapter 2. Building on a prior result by Nehring and Puppe (2002/2007a), and related to List and Pettit (2004) and also to Nehring (2003), this paper (joint with Christian List) provides an impossibility theorem on judgment aggregation that generalises Arrow's Theorem to the more comprehensive framework of judgment aggregation. More precisely, the theorem states that, under a rather complicated agenda condition, every judgment aggregation rule with four properties – universal domain, collective rationality, independence and unanimity-preservation – is a dictatorship. As applied to the special problem of (strict) preference aggregation (represented as

a judgment aggregation problem), our theorem becomes precisely Arrow's Theorem. That is, our agenda condition reduces to Arrow's condition of having at least three alternatives, and our aggregation conditions reduce to Arrow's aggregation conditions: our independence reduces to Arrow's *independence of irrelevant alternatives*, our unanimity-preservation to the Pareto principle, and so on. It is interesting to see that, while nearly every preference aggregation problem (i.e. each problem with more than two alternatives) is subject to the Arrowian impossibility, many judgment aggregation agendas escape the impossibility. Examples of 'possibility agendas' are studied in Chapter 6.

Essentially the same result is also proven independently by Dokow and Holzman (forthcoming), who moreover show that, importantly, the agenda condition is not only sufficient but also necessary for the impossibility: any judgment aggregation problem that violates the agenda condition – and many do, as just mentioned – leads to possibility.

Chapter 3. Next to Arrow's Theorem, another famous impossibility result of preference aggregation is Sen's *Liberal Paradox* (or *Impossibility of a Paretian Liberal*). Sen's simple but ingenious result highlights a conflict between respecting individual rights and respecting consensus. More precisely, respecting rights is formalised by making each individual decisive on those pairs of alternatives that lie within his 'private sphere' (e.g. that differ only in the clothes this individual wears), and respecting consensus is formalised by the weak Pareto principle (which can of course be applied to pairs of alternatives that do not lie in anyone's private sphere). Sen takes this conflict as a reason for rejecting the Pareto principle, whereas many others prefer to dilute rights.

This chapter's paper (again, joint with Christian List) shows that Sen's Liberal Paradox has, like Arrow's Theorem, an exact analogue within judgment aggregation. We formalise a right of a person as his decisiveness on a proposition, and we show that, under a (necessary and sufficient) agenda condition, a conflict arises between respecting individual rights (on some propositions) and consensus (on other propositions). The agenda condition again requires 'sufficient interconnections' between propsotions, albeit in a different sense than for the Arrow-type impossibility. Our example agenda, $X = \{a, \neg a, a \rightarrow b, \neg(a \rightarrow b), b, \neg b\}$, displays 'sufficient interconnections', hence is vulnerable to a liberal paradox. To illustrate this, suppose some person is given the right to decide a, and another to decide $a \rightarrow b$. In respecting these rights, the collective may be led to accept both a and $a \rightarrow b$, which may clash with a unanimous 'no' judgment on b.

If our Sen-type result is applied to preference aggregation (as a special judgment aggregation problem), we obtain precisely Sen's result. The Sen-type result has, like the Arrow-type result, less of an impossibility flavour in the general judgment aggregation context than in the special preference aggregation context: indeed, many judgment aggregation problems are not vulnerable to, i.e. do not satisfy the agenda condition of, a liberal paradox. Incidentally, the agenda condition for a liberal paradox is, in general, neither weaker nor stronger than that for the Arrow-type impossibility. But it is weaker in the special case of preference aggregation: here it holds as soon as there are more than *one* (not more than *two*) alternatives.

We interpret our result not only in terms of 'liberal' rights (as for Sen's result), but

also in terms of expert rights: an individual may be made decisive on a proposition on the grounds of his special expertise on that proposition, for instance if the individual is a physicist and the proposition is a physical hypothesis.

Chapter 4. In this paper (once more, joint with Christian List), we investigate the question of whether and when voters can manipulate the outcome of aggregation by misrepresenting their judgments. This question is of obvious importance in view of implementing a judgment aggregation rule. But, unlike in the preference aggregation model, it is not obvious how to even define when an aggregation rule is immune (or not immune) to voter manipulation. The difficulty lies in that an 'incentive' is a preference-theoretic notion, but preferences are not part of the judgment aggregation model. We have a two-fold response to this problem, namely by distinguishing between opportunities for manipulation and incentives for manipulation. The former is a preference-free notion and can be formalised within the judgment aggregation model as it stands. The latter leads us to enrich the model by introducing preferences of individuals (over collective judgment sets). It is not obvious what assumptions to make about an individual's preferences; in fact, it is not even clear whether he would like the collective to take over his own judgment set: just imagine that the propositions speak about the effect of pollution and that an individual, who happens to own a polluting fabric, believes that pollution is harmful but, worried about his private business, does not like the collective to believe this.

We derive formal results on the (im)possibility of aggregation without opportunities resp. without incentives to manipulate. These results (especially the incentive-based ones) are closely related to results by Nehring and Puppe (2002) in a different framework. We also prove that under certain assumptions on voters' motivations/preferences – assumptions that notably exclude the motivation of the above-mentioned factory owner – the opportunity-based and the incentive-based approach to manipulation are equivalent.

We further compare the premise- and the conclusion-based rule from the perspective of voter manipulation. Under the motivational assumption that the individuals are *outcome-oriented* – i.e. do not care about the collective's judgment on premises but would like the collective to follow their own judgment on the conclusion – the premise-based rule is strategically manipulable, whereas the conclusion-based rule is not.

Incidentally, while Arrow's and Sen's impossibility theorems have their counterparts in judgment aggregation, the third celebrated impossibility theorem of preference aggregation, the Gibbard-Satterthwaite Theorem on strategy-proofness, cannot be expected to have an exact analogue within judgment aggregation theory. The simple reason is that this theorem is, strictly speaking, not about aggregating individual preferences into a social preference but about aggregating individual preferences into single social choices, and this input-output asymmetry makes the latter aggregation problem be not a special case of judgment aggregation.

Chapter 5. This paper analyses agenda manipulation, which next to voter manipulation constitutes a second threat to reaching desirable collective judgments. For example, consider our introductory example, whose agenda contains the propositions a, $a \rightarrow b$ and b (and their negations), and assume the premise-based rule is used, leading to collective acceptance of b. Then an agenda setter who would prefer the

collective to reject b might replace the premise propositions a and $a \to b$ (and their negations) by new premise propositions a' and $a' \to b$ (and their negations), where a' might be the proposition 'The US Dollar will appreciate against the local currency', in the hope that the new premises a' and $a' \to b$ do not both receive majorities, so that the premise-based rule now leads to rejection of b. More generally, an agenda setter might add or remove propositions from the agenda X in the hope of either (i) achieving certain collective judgments on the added propositions, or (ii) preventing certain collective judgments on the removed propositions, or (iii) changing the collective judgments on the propositions that were in the old and in the new agenda (like the proposition b in the last example), or (iv) changing judgments in the *deductive closure* of the collective judgment set across all propositions of the language (including those outside the old and new agenda). There exist even other forms of agenda manipulation. For each form of agenda manipulation, the question arises as to what aggregation rules are vulnerable to such manipulation. Some forms of manipulation – such as (i) and (ii) – can never be prevented. Other forms can be prevented, as I show in the paper. A property that is shown to play a crucial role in preventing agenda manipulation is the property of independence (i.e. of propositionwise aggregation). I also use different variants of this condition, obtained by aggregating propositionwise only on certain propositions. Part of the paper is also devoted to generalising the premise-based rule.

Chapter 6. Ever since the early (informal) beginnings of the judgment aggregation literature, propositions of the conditional form 'if p then q' or of the biconditional form 'p if, and only if q' have been used as typical examples of propositions on which real-life groups can disagree. A natural class of judgment aggregation problems is indeed given by the agendas that contain two types of propositions (and their negations): *atomic* propositions (typically stating simple facts or norms) and (bi)conditionals $p \to q$ or $p \leftrightarrow q$ between atomic propositions in the agenda or between conjunctions thereof. In this paper, I analyse such agendas, which a call *implication agendas*. (More generally, one might also allow (bi)conditionals between any Boolean combinations of atomic propositions in the agenda, not necessarily conjunctions.)

In formal logic, the conditional connective '$\to$' (and its bidirectional variant '$\leftrightarrow$') can be given different interpretations. According to the material interpretation, $p \to q$ is logically equivalent to the disjunction $\neg p \vee q$ (either p is false or q is true). In particular, the truth value of the material implication $p \to q$ is uniquely determined by the *actual* truth values of p and q. According to the subjunctive interpretation, $p \to q$ means 'if p *were* true, q *would* be true', which is a statement not about p's and q's actual truth values but about b's true value in a *hypothetical* (perhaps counterfactual) world in which p is true. To illustrate the difference, '*if* Karlsruhe becomes the capital of Europe *then* it becomes insignificant' is true as a material implication (because Karlsruhe does not become the capital of Europe, sadly enough), but false as a subjunctive implication because in the world (case) in which Karlsruhe becomes Europe's capital Karlsruhe will become known in all corners of the planet. Clearly, the subjunctive interpretation comes closer to our intution here. As argued in the paper, the (bi)implications that occur in real-life judgment aggregation problems are typically intended as subjunctive implications (or as other non-truthfunctional implications). The paper goes on to study the possibility or impossibility of judgment

aggregation by quota rules for the just-defined implication agendas; it turns out that much depends on whether we interpret the (bi)implications in these agendas materially or subjunctively. While the material interpretation leads to the impossibility of aggregating by quota rules (unless the agenda is 'very small'), the subjunctive interpretation always leads to possibility. I also show exactly which combinations of quota (acceptance thresholds) across propositions are allowed, i.e. guarantee consistent collective judgment sets.

Informally, the paper's main message is that it matters to adequately represent the propositions in an agenda: the logical connectives we use should be faithful to the intended meaning in real life. This typically requires the use of non-classical (i.e. non-truth-functional) logical connectives. Misrepresentations create unnatural and often too strong logical interconnections between propositions, which may lead to artificial impossibilities of aggregation.

Chapter 7. Most authors (including myself in many of my contributions) impose in their theorems a strong condition on the aggregation rule, the already-mentioned independence condition, whereby aggregation is performed on a propositionwise basis: the collective judgment on any proposition in the agenda is a function of the individual judgments on that proposition alone. On the other hand, the perhaps most popular response to the discursive dilemma, the premise-based rule, is not a propositionwise rule (as is seen from the way the conclusion proposition is decided); nor are distance-based rules. There has indeed been a divergence between the formal and the informal developments in the field, i.e. between the direction that the 'industry of theorems' has taken (by focussing on propositionwise aggregation) and the ongoing informal discussion on how a 'good' or 'democratic' procedure should look like.

The independence condition is not without arguments in its favour. In addition to the role it plays in preventing manipulation by voters (see Chapter 4) or agenda setters (see Chapter 5), a normative defence could be based on a *local* notion of democratic aggregation, which underlies for example most real-world systems of direct democracy. Under a local understanding of democracy, a collective decision on a given issue counts as 'democratic' if it reflects people's positions – that is: positions on that issue – while the presence of other determinants (such as a coin toss or indeed people's positions on other issues) undermine democratic legitimacy. But if by contrast one adopts a more *holistic* notion, e.g. a premise-based approach, independence loses its appeal and becomes objectionable. Further, decision problems involving non-binary issues (such as: estimating temperature on a real scale) can be reasonably represented in the judgment aggregation model only by relaxing the independence condition, as explained in the paper.

Motivated by a more holistic approach, by non-binary issues, and by the 'impossibility message' of several existing theorems on propositionwise aggregation, the present paper gives up the independence condition. In the absence of independence, one is at first faced with a large – far too large! – set of possible aggregation rules. This calls for a new condition that 'disciplines' aggregation and narrows down the class of possibilities. To this end, I introduce a general informational restriction on aggregation: *independence of irrelevant information* (III), whereby the collective judgment on any proposition p depends only on the individuals' judgments on those propositions that are relevant to p. This condition has a parameter, the notion of

relevance employed. Formally, a notion of relevance is captured by a binary relation $\mathcal{R}$ on the the agenda, where $q\mathcal{R}p$ is read 'q is relevant to (the decision on) p'. Nearly every plausible informational constraint on aggregation is a special case of our III condition with some suitable relevance relation $\mathcal{R}$. For instance, the classic independence relation is equivalent to III if we allow only self-relevance (i.e. $q\mathcal{R}p \Leftrightarrow p = q$ for all $p, q \in X$); whereas III permits the premise-based rule if premise propositions are relevant to conclusion propositions. In fact, if the relevance relation is interpreted as a relation of premisehood, the III condition becomes precisely the condition of premise-based aggregation in a generalised sense. The paper explores different variants of how to define relevance. It also shows that impossibilities of aggregation re-emerge under certain conditions on the interplay between logical interconnections and relevance connections.

4 Related models and aggregation problems

For the interested reader, I now briefly discuss related models and aggregation problems.

Semantic representation of propositions. The judgment aggregation problem and work on it could be formulated semantically rather than syntactically. That is, instead of defining $\mathbf{L}$ as a set of sentences (sequences of symbols), let $\mathbf{L}$ be a Boolean algebra of subsets of some underlying set $\Omega \neq \emptyset$ of 'possible worlds'. The agenda is still a non-empty negation-closed set $X \subseteq \mathbf{L}$, where 'negation-closed' now means 'complement-closed'. As for the consistency notion, a set of propositions $A \subseteq \mathbf{L}$ is inconsistent if and only if its intersection $\cap_{p \in A} p$ is empty; and it entails a proposition $q \in \mathbf{L}$ if and only if $\cap_{p \in A} p \subseteq q$. Although my work on judgment aggregation follows the syntactic representation of propositions, I do not take a position in the epistemological debate over whether the objects of beliefs are sentences or semantic propositions. What makes it perhaps more natural to represent propositions syntactically *in our aggregative context* is that, after all, the propositions in the agenda must be communicated by the agenda setter to the individuals, and communication works syntactically, namely though verbal or written speach. From a technical perspective, the two approaches are essentially, but not totally, isomorphic.[4] Non-truthfunctional operators are represented by suitable functions in $\mathbf{L}$; for instance, the knowledge operator 'it is known that' is represented by a function $K : \mathbf{L} \to \mathbf{L}$ satisfying suitable 'knowledge axioms'.

Finally, the semantic approach could alternatively be formulated in a purely abstract way, without invoking (sets of) worlds, namely by letting $\mathbf{L}$ be an *abstract* Boolean algebra (i.e. a lattice containing a bottom, a top, and the negation of any element).

[4] A syntactic model can be converted into a semantic one by defining the worlds as the complete and consistent sets of sentences and identifying each sentence p with the set of worlds containing p. This maps the logic $\mathbf{L}$ to an algebra over the set of worlds (provided the logic contains conjuntction). This mapping preserves the consistency notion. Technically, it defines a homororphism, but not an isomorphism because distinct but logically equivalent sentences are mapped to the same set of worlds. In short, the move to a semantic model preserves logical relations but looses the syntax, hence involves a limited loss of information.

Abstract aggregation theory. Most theorems of the literature on judgment aggregation can be stated in an abstract and logic-free model, which is indeed done by different authors. There are many variants of the abstract model; they involve no underlying logic $\mathbf{L}$ but in one form or another an agenda. Let me here present one variant, Dokow and Holzman's *binary evaluations* model (other variants being Wilson's 1976 model and Nehring and Puppe's 2002 *property space* model). Let K be a non-empty set of *issues*, let an *evaluation* be a function $v : K \to \{0, 1\}$ (assigning a position, 0 or 1, to each issue), and let $\mathbf{E}$ be a non-empty set of '*admissible*' evaluations. One may then look for aggregation functions $f : \mathbf{E}^n \to \mathbf{E}$, mapping profiles $(v_1, ..., v_n)$ of admissible individual evaluations to admissible collective evaluations (where n is the number of individuals). To see the connection to judgment aggregation, let K be the set of non-negated propositions in the agenda X (so that K consists of exactly one member of each proposition-negation pair $p, \neg p$ in X), identify every evaluation v with the opinionated judgment set $A \subseteq X$ that contains each p with $v(p) = 1$ and each $\neg p$ with $v(p) = 0$, and call an evaluation v admissible if the corresponding opinionated judgment set is consistent (hence fully rational). Under these identifications, abstract aggregation functions $f : \mathbf{E}^n \to \mathbf{E}$ correspond uniquely to judgment aggregation functions that have universal domain and generate fully rational collective judgment sets. Adaptations of this abstract model can also cope with rationality violations (on the individual or collective level), in particular with 'abstentions' (corresponding to incompleteness of judgment sets).

It is worth noting that a loss of information is involved in moving from the judgment aggregation model to an abstract model. It becomes impossible to (i) refer to propositions outside the agenda, and (ii) refer to a proposition's syntactic form, e.g. distinguish atomic from compound propositions. (The loss of the syntax actually also happens when moving to a semantic model.) This informational loss detaches the model to some extent from philosophical questions that one may raise about judgments and their aggregation; and one cannot anymore state conditions or theorems that draw on additional information (such as Mongin's (forthcoming) independence condition restricted to atomic propositions). However, most conditions and theorems proved up to now do not draw on additional information, and from this perspective one might regard the informational slimness of abstract models as an appealing feature.

General attitude aggregation. We are back now to the logic-based framework, with a language $\mathbf{L}$ and an agenda $X \subseteq \mathbf{L}$ of propositions under consideration. The standard model allows exactly two attitudes on each proposition in X: acceptance or rejection. But belief need not be a binary on-off affair. This calls for a generalisation. Following Dietrich and List (forthcoming), consider an arbitrary non-empty set $\mathbf{V}$ of possible attitudes: in the binary case $\mathbf{V} = \{0, 1\}$, in the case of probabilistic attitudes $\mathbf{V} = [0, 1]$, in the case of Spohnian ranks $\mathbf{V} = \{0, 1, ...\} \cup \{\infty\}$, and so on. Let $\mathbf{F}$ be a set of functions $f : \mathbf{L} \to \mathbf{V}$, the *valuation* functions; in the binary case these are the truth functions, in the probabilistic case the probability functions (or perhaps the Dempster-Schaefer belief functions or the capacities),[5] in the case of Spohnian

[5] Probability functions (that is, fininitely addive ones) require a Bollean algebra as their domain. To ensure $\mathbf{L}$ forms an (abstract) Boolean algebra (modulo logical equivalence), I here make a small additional richness assumption on the language: conjunctions can be formed (i.e. $p, q \in \mathbf{L}$ implies

ranks the Spohnian ranking functions, and so on. An *attitude function* is a function $A : X \to \mathbf{V}$; it is *rational* if it is extendable to a valuation function, i.e. to a function $\mathbf{L} \to \mathbf{V}$ in $\mathbf{F}$. So, in the binary case rationality means extendability to a truth function, in the probabilistic case it means extendability to a probability function, and so on. The problem then is to aggregate a profile of individual attitude functions $(A_1, ..., A_n)$ into a collective attitude function; again, the output of aggregation should ideally be rational, democratically responsive to the individual inputs, and defined on a large domain of possible input profiles. Dietrich and List (forthcoming) propose to study this general aggregation problem and offer some first results. This problem unifies different concrete aggregation problems in the literature, notably judgment, preference, and probability aggregation.

Of course, this non-binary aggregation problem again has a semantic variant (in which $\mathbf{L}$ is an algebra over some set of worlds Ω), and an abstract variant (which involves no logic $\mathbf{L}$; see Dokow and Holzman 2007).

5 The related literature

In the course of this introduction I have repeatedly referred to the literature, and the papers in later chapters all contain their own literature reviews. I therefore only briefly sketch here the developments in the field. The judgment aggregation theory has diverse origins. List and Pettit's (2002) seminal contribution marks the beginning of the formal social-choice-theoretic literature. The model I presented in Dietrich (2007) was intended as a generalisation of List and Pettit's original classical-logical model, which in turn was an abstraction of the informal literature on judgment aggregation and the discursive dilemma in law and political philosophy (e.g. Kornhauser and Sager 1985 and Pettit 2001). Computer scientists have already for a long time worked on the *belief merging* problem, which is closely related in that also here sets of logical propositions are being aggregated (e.g. Konieczny and Pino-Perez 2002); the two perhaps most notable differences are that belief merging often has different applications in mind (e.g. merging data bases) and does not focus on propositionwise aggregation (and on resulting impossibilities). Guilbaud's (1966) 'logical problem of aggregation' can be viewed as an early precursor to the judgment aggregation theory. Abstract aggregation theory goes back at least to Wilson (1975), whose impossibility result already generalises Arrow's Theorem; I have already discussed the close link between abstract and judgment aggregation.

A series of theorems have by now established that, for sufficiently interconnected agendas, every propositionwise aggregation rule violates certain desiderata (such as to generate consistent outputs, to be non-dictatorial, anonymous, unanimity-preserving, and so on), where the theorems differ in the chosen desiderata and in the agenda conditions. While the early theorems (List and Pettit 2002, Pauly and van Hees 2006, Dietrich 2006a, Gärdenfors 2006) did not provide minimal agenda conditions, Nehring and Puppe (2002/2007a, 2007b, 2008) were the first to give necessary and sufficient agenda conditions for impossibilities of propositionwise aggregation; in fact, their results close to exhaustively treat the case in which the propositionwise aggregation rule

$p \wedge q \in \mathbf{L}$, where as usual $p \wedge q$ is logically equivalent to the pair $\{p, q\}$ in the sense of mutual entailment).

is required to generate fully rational outputs and to be monotonic and unanimity-preserving. I have already discussed Dietrich and List's (2007c) and Dokow and Holzman's (forthcoming) Arrow-type impossibility results. Allowing incompleteness in collective judgment sets does not open up genuine possibilities of propositionwise aggregation (e.g. Dietrich and List 2008 and Dokow and Holzman 2006). Applying propositionwise aggregation only to certain propositions, interpretable as the premise propositions, may or may not lead to possibility results, depending on the choice of premise propositions (Dietrich 2006a, Mongin forthcoming, Nehring 2006, Dietrich and Mongin 2007). Restricting the domain of the aggregation rule to profiles in which disagreements between people take a suitably systematic form opens up possibilities of propositionwise, even majoritarian, aggregation (List 2003, Dietrich and List 2007d). Other contributions give up propositionwise aggregation altogether, for example in favour of sequential rules (e.g. List 2004), fusion operators (Konieczny and Pino-Perez 2002 and Pigozzi 2006), or aggregation rules based on relevant information (see Chapter 7). For treatments of individual or subgroup rights, and of voter manipulation, see the papers of Chapters 3 and 4.

6 References

Arrow, K. (1951/1963) *Social Choice and Individual Values*, New York (Wiley)

Baigent, N. (1987) Preference proximity and anonymous social choice, *Quarterly Journal of Economics* 102(1): 161-170

Dietrich, F. (2006a) Judgment aggregation: (im)possibility theorems, *Journal of Economic Theory* 126(1): 286-298

Dietrich, F. (2006b) Aggregation theory and the relevance of some issues to others, *working paper*, London School of Economics

Dietrich, F. (2007) A generalised model of judgment aggregation, *Social Choice and Welfare* 28(4): 529-565

Dietrich, F. (forthcoming) The possibility of judgment aggregation on agendas with subjunctive implications, *Journal of Economic Theory*

Dietrich, F., List, C. (2007a) Strategy-proof judgment aggregation, *Economics and Philosophy* 23: 269-300

Dietrich, F., List, C. (2007b) Judgment aggregation by quota rules: majority voting generalized, *Journal of Theoretical Politics* 19(4): 391-424

Dietrich, F., List, C. (2007c) Arrow's theorem in judgment aggregation, *Social Choice and Welfare* 29(1): 19-33

Dietrich, F., List, C. (2007d) Majority voting on restricted domains, *working paper*, London School of Economics

Dietrich, F., List, C. (2008) Judgment aggregation without full rationality, *Social Choice and Welfare* 31(1): 15-39

Dietrich, F., List, C. (forthcoming) The aggregation of propositional attitudes: towards a general theory, *Oxford Studies in Epistemology*

Dietrich, F., Mongin, P. (2007) The premise-based approach to judgment aggregation, *working paper*, HEC, Paris

Dokow, E., Holzman, R. (forthcoming) Aggregation of binary evaluations, *Journal of Economic Theory*

Dokow, E., Holzman, R. (2006) Aggregation of binary evaluations with abstentions,

working paper, Technion Israel Institute of Technology

Dokow, E., Holzman, R. (2007) Aggregation of non-binary evaluations: towards a general result, *working paper*, Technion Israel Institute of Technology

Gärdenfors, P. (2006) An Arrow-like theorem for voting with logical consequences, *Economics and Philosophy* 22(2): 181-190

Guilbaud, G. Th. (1966) Theories of the general interest, and the logical problem of aggregation, in P. F. Lazarsfeld and N. W. Henry (eds.), *Readings in Mathematical Social Science*, Cambridge/MA (MIT Press): 262-307

Konieczny, S., Pino-Perez, R. (2002) Merging information under constraints: a logical framework, *Journal of Logic and Computation* 12: 773-808

Kornhauser, L. A., Sager, L. G. (1986) Unpacking the court, *Yale Law Journal* 96(1): 82-117

List, C. (2001) *Mission Impossible? The Problem of Democratic Aggregation in the Face of Arrow's Theorem*, DPhil-thesis in Politics, University of Oxford

List, C. (2003) A possibility theorem on aggregation over multiple interconnected propositions, *Mathematical Social Sciences* 45(1): 1-13 (Corrigendum in *Mathematical Social Sciences* 52:109-110)

List, C. (2004) A model of path dependence in decisions over multiple propositions, *American Political Science Review* 98(3): 495-513

List, C. (2006) The discursive dilemma and public reason, *Ethics* 116(2): 362-402

List, C., Pettit, P. (2002) Aggregating sets of judgments: an impossibility result, *Economics and Philosophy* 18: 89-110

List, C., Pettit, P. (2004) Aggregating sets of judgments: two impossibility results compared, *Synthese* 140(1-2): 207-235

Mongin, P. (forthcoming) Factoring out the impossibility of logical aggregation, *Journal of Economic Theory*

Nehring, K. (2003) Arrow's theorem as a corollary, *Economics Letters* 80(3): 379-382

Nehring, K. (2006) The impossibility of a Paretian rational, *working paper*, University of California at Davies

Nehring, K., Puppe, C. (2002) Strategy-proof social choice on single-peaked domains: possibility, impossibility and the space between, *working paper*, University of California at Davies

Nehring, K., Puppe, C. (2007a) The structure of strategy-proof social choice, part I: general characterization and possibility results on median spaces, *Journal of Economic Theory* 135: 269-305

Nehring, K., Puppe, C. (2007b) Abstract Arrowian aggregation, *working paper*, University of California at Davies

Nehring, K., Puppe, C. (2008) Consistent judgement aggregation: the truth-functional case, *Social Choice and Welfare* 31: 41-57

Pigozzi, G. (2006) Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation, *Synthese* 152(2): 285-298

Pettit, P. (2001) Deliberative democracy and the discursive dilemma, *Philosophical Issues* (supplement 1 of *Nous*) 11: 268-95

Wilson, R. (1975) On the theory of aggregation, *Journal of Economic Theory* 10: 89-99

Chapter 2

A generalisation of Arrow's Theorem

Paper: Arrow's theorem in judgment aggregation (with C. List), *Social Choice and Welfare* 29: 19-33, 2007

Arrow's theorem in judgment aggregation

Franz Dietrich and Christian List[1]

In response to recent work on the aggregation of individual judgments on logically connected propositions into collective judgments, it is often asked whether judgment aggregation is a special case of Arrowian preference aggregation. We argue for the converse claim. After proving two impossibility theorems on judgment aggregation (using "systematicity" and "independence" conditions, respectively), we construct an embedding of preference aggregation into judgment aggregation and prove Arrow's theorem (stated for strict preferences) as a corollary of our second result. Although we thereby provide a new proof of Arrow's theorem, our main aim is to identify the analogue of Arrow's theorem in judgment aggregation, to clarify the relation between judgment and preference aggregation, and to illustrate the generality of the judgment aggregation model.
JEL Classification: D70, D71

1 Introduction

The problem of "judgment aggregation" has recently received much attention: How can the judgments of several individuals on logically connected propositions be aggregated into corresponding collective judgments? To illustrate, suppose a three-member committee has to make collective judgments (acceptance/rejection) on three connected propositions:

a: "Carbon dioxide emissions are above the threshold x."

$a \rightarrow b$: "If carbon dioxide emissions are above the threshold x, then there will be global warming."

b: "There will be global warming."

	a	$a \rightarrow b$	b
Individual 1	True	True	True
Individual 2	True	False	False
Individual 3	False	True	False
Majority	True	True	False

Table 1: The discursive paradox

[1]We thank Richard Bradley, Ruvin Gekker, Ron Holzman, Philippe Mongin, Klaus Nehring, Clemens Puppe and the referees for comments and suggestions. Addresses for correspondence: F. Dietrich, Department of Quantitative Economics, University of Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands; C. List, Department of Government, London School of Economics, London WC2A 2AE, U.K.

As shown in Table 1, the first committee member accepts all three propositions; the second accepts a but rejects $a \rightarrow b$ and b; the third accepts $a \rightarrow b$ but rejects a and b. Then the judgments of each committee member are individually consistent, and yet the majority judgments on the propositions are inconsistent: a majority accepts a, a majority accepts $a \rightarrow b$, but a majority rejects b.

This so-called *discursive paradox* (Pettit 2001) has led to a growing literature on the possibility of consistent judgment aggregation under various conditions. List and Pettit (2002) have provided a first model of judgment aggregation based on propositional logic and proved that no aggregation rule generating consistent collective judgments can satisfy some conditions inspired by (but not equivalent to) Arrow's conditions on preference aggregation. This impossibility result has been extended and strengthened by Pauly and van Hees (forthcoming; see also van Hees forthcoming), Dietrich (2006), and Gärdenfors (forthcoming).[2] Abstracting from propositional logic, Dietrich (forthcoming) has provided a model of judgment aggregation in general logics, which we use in the present paper, that can represent aggregation problems involving propositions formulated in richer logical languages. Drawing on the related model of "property spaces", Nehring and Puppe (2002, 2005) have proved the first agenda characterization results, identifying necessary and sufficient conditions under which an agenda of propositions gives rise to an impossibility result under certain conditions.

Although judgment aggregation is different from the more familiar problem of preference aggregation, the recent results resemble earlier results in social choice theory. The discursive paradox resembles Condorcet's paradox of cyclical majority preferences, and the various recent impossibility theorems resemble Arrow's and other theorems on preference aggregation. This raises the question of how the work on judgment aggregation is related to earlier work in social choice theory. Provocatively expressed, is it just a reinvention of the wheel?

It can be replied that the logic-based model of judgment aggregation generalizes Arrow's classical model of preference aggregation. Specifically, preference aggregation problems can be modelled as special cases of judgment aggregation problems by representing preference orderings as sets of binary ranking judgments in predicate logic (List and Pettit 2001/2004; List 2003).[3] Less formally, this way of thinking about preferences goes back to Condorcet himself (see also Guilbaud 1966).

In this paper, we reinforce this argument. After introducing the judgment aggregation model in general logics and proving two impossibility results (using "systematicity" and "independence" conditions, respectively), we construct

[2] Possibility results, obtained by relaxing some of the conditions of these impossibility results, have been proved by List (2003, 2004); Dietrich (2006), Pigozzi (forthcoming), and Dietrich and List (2005). The relationship with the Condorcet jury theorem has been investigated by Bovens and Rabinowicz (2006) and List (2005).

[3] This embedding works only for the ordinal preference-relation-based part of Arrowian social choice theory, not for the cardinal welfare-function-based part. Wilson's (1975) aggregation model, as discussed in our concluding remarks, is another generalization of ordinal preference aggregation.

an explicit embedding of preference aggregation into judgment aggregation and prove Arrow's theorem (for strict preferences) as a corollary of our second impossibility result. We also point out that our first impossibility result has corollaries for the aggregation of other binary relations (such as partial orderings or equivalence relations).

Although we thereby provide a new proof of Arrow's theorem, our primary aim is to identify the analogue of Arrow's theorem in judgment aggregation, to clarify the logical relation between judgment and preference aggregation, and to illustrate the generality of the judgment aggregation model.

Related results were given by List and Pettit (2001/2004), who derived a simple impossibility theorem on preference aggregation from their (2002) impossibility result on judgment aggregation, and Nehring (2003), who derived an Arrow-like impossibility theorem from Nehring and Puppe's (2002) characterization result in the related model of "property spaces". But neither result exactly matches Arrow's theorem. Compared to Arrow's original theorem, List and Pettit's result requires additional neutrality and anonymity conditions, but no Pareto principle; Nehring's result requires an additional monotonicity condition. We highlight the connections of our present results with these and other results (including recent results by Dokow and Holzman 2005) throughout the paper.

2 The judgment aggregation model

We consider a group of individuals $1, 2, \ldots, n$ ($n \geq 2$). The group has to make collective judgments on logically connected propositions.

Formal logic. Propositions are represented in an appropriate logic. A *logic (with negation symbol $\neg$)* is an ordered pair $(\mathbf{L}, \vDash)$, where (i) $\mathbf{L}$ is a non-empty set of formal expressions (*propositions*) closed under negation (i.e., if $p \in \mathbf{L}$ then $\neg p \in \mathbf{L}$), and (ii) $\vDash$ is an *entailment relation*, where, for each set $A \subseteq \mathbf{L}$ and each proposition $p \in \mathbf{L}$, $A \vDash p$ is read as "A entails p" (we write $p \vDash q$ to abbreviate $\{p\} \vDash q$).[4]

A set $A \subseteq \mathbf{L}$ is *inconsistent* if $A \vDash p$ and $A \vDash \neg p$ for some $p \in \mathbf{L}$, and *consistent* otherwise; $A \subseteq \mathbf{L}$ is *minimal inconsistent* if it is inconsistent and every proper subset $B \subsetneq A$ is consistent. A proposition $p \in \mathbf{L}$ is *contingent* if $\{p\}$ and $\{\neg p\}$ are consistent.

We require the logic to satisfy the following minimal conditions:

(L1) For all $p \in \mathbf{L}$, $p \vDash p$ (self-entailment).

(L2) For all $p \in \mathbf{L}$ and $A \subseteq B \subseteq \mathbf{L}$, if $A \vDash p$ then $B \vDash p$ (monotonicity).

(L3) $\emptyset$ is consistent, and each consistent set $A \subseteq \mathbf{L}$ has a consistent superset $B \subseteq \mathbf{L}$ containing a member of each pair $p, \neg p \in \mathbf{L}$ (completability).

[4] Formally, $\vDash \subseteq \mathcal{P}(\mathbf{L}) \times \mathbf{L}$, where $\mathcal{P}(\mathbf{L})$ is the power set of $\mathbf{L}$.

Many different logics satisfy conditions L1 to L3, including standard propositional logic, standard modal and conditional logics and, for the purpose of representing preferences, predicate logic, as defined below. For example, in standard propositional logic, $\mathbf{L}$ contains propositions such as a, b, $a \wedge b$, $a \vee b$, $a \rightarrow b$, $\neg(a \wedge b)$, and $\vDash$ satisfies $\{a, a \rightarrow b\} \vDash b$, $b \vDash a \vee b$, but not $b \vDash a \wedge b$.

The agenda. The *agenda* is a non-empty subset $X \subseteq \mathbf{L}$, interpreted as the set of propositions on which judgments are to be made, where X is a union of proposition-negation pairs $\{p, \neg p\}$ (with p not itself a negated proposition). For simplicity, we assume that double negations cancel each other out, i.e., $\neg\neg p$ stands for p.[5] In the discursive paradox example above, the agenda is $X = \{a, \neg a, b, \neg b, a \rightarrow b, \neg(a \rightarrow b)\}$, with $\rightarrow$ interpreted either as the material conditional in standard propositional logic or as a subjunctive conditional in a suitable conditional logic.

Agenda richness. Whether or not judgment aggregation gives rise to serious impossibility results depends on how the propositions in the agenda are interconnected. We consider agendas X with different types of interconnections. Our basic agenda assumption, which significantly generalizes the one in List and Pettit (2002), is *minimal connectedness*. An agenda X is *minimally connected* if (i) it has a minimal inconsistent subset $Y \subseteq X$ with $|Y| \geq 3$, and (ii) it has a minimal inconsistent subset $Y \subseteq X$ such that $(Y \backslash Z) \cup \{\neg z : z \in Z\}$ is consistent for some subset $Z \subseteq Y$ of even size.[6]

As Ron Holzman has indicated to us, part (ii) of minimal connectness is equivalent to Dokow and Holzman's (2005) assumption that the set of admissible yes/no views on the propositions in X is a non-affine subset of $\{0,1\}^X$.[7]

To obtain a more demanding agenda assumption, we define *path-connectedness*, a variant of Nehring and Puppe's (2002) assumption of *total blockedness*.[8] For any $p, q \in X$, we write $p \vDash^* q$ if $\{p, \neg q\} \cup Y$ is inconsistent for some $Y \subseteq X$ consistent with p and with $\neg q$.[9] Now an agenda X is *path-connected* if, for every contingent $p, q \in X$, there exist $p_1, p_2, ..., p_k \in X$ (with $p = p_1$ and $q = p_k$) such that $p_1 \vDash^* p_2$, $p_2 \vDash^* p_3$, ..., $p_{k-1} \vDash^* p_k$.

The agenda of our example above is minimally connected, but not path-connected. As detailed below, preference aggregation problems can be represented by agendas that are both minimally connected and path-connected. The

[5] When we use the negation symbol $\neg$ hereafter, we mean a modified negation symbol $\sim$, where $\sim p := \neg p$ if p is unnegated and $\sim p := q$ if $p = \neg q$ for some q.

[6] Note that the set Y can be different in parts (i) and (ii).

[7] In the first version of this paper, we had used a more restrictive version of part (ii), requiring Z to be of size two rather than even size. The present version of part (ii) was introduced in Dietrich (forthcoming).

[8] For a compact logic, path-connectedness is equivalent to total blockedness; in the general case, path-connectedness is weaker.

[9] For non-paraconsistent logics (in the sense of L4 in Dietrich forthcoming), $\{p, \neg q\} \cup Y$ is inconsistent if and only if $\{p\} \cup Y \vDash q$.

aggregation of many other binary relations can be represented by minimally connected agendas.

Individual judgment sets. Each individual i's *judgment set* is a subset $A_i \subseteq X$, where $p \in A_i$ means that individual i accepts proposition p. A judgment set A_i is *consistent* if it is a consistent set as defined above; A_i is *complete* if, for every proposition $p \in X$, $p \in A_i$ or $\neg p \in A_i$. A *profile (of individual judgment sets)* is an n-tuple $(A_1, \ldots, A_n)$.

Aggregation rules. A *(judgment) aggregation rule* is a function F that assigns to each admissible profile $(A_1, \ldots, A_n)$ a single collective judgment set $F(A_1, \ldots, A_n) = A \subseteq X$, where $p \in A$ means that the group accepts proposition p. The set of admissible profiles is called the *domain* of F, denoted $Domain(F)$. Examples of aggregation rules are the following.

- *Propositionwise majority voting.* For each $(A_1, ..., A_n)$, $F(A_1, ..., A_n) = \{p \in X : \text{more individuals } i \text{ have } p \in A_i \text{ than } p \notin A_i\}$.
- *Dictatorship of individual i.* For each $(A_1, ..., A_n)$, $F(A_1, ..., A_n) = A_i$.
- *Inverse dictatorship of individual i.* For each $(A_1, ..., A_n)$, $F(A_1, ..., A_n) = \{\neg p : p \in A_i\}$.

Regularity conditions on aggregation rules. We impose the following conditions on the inputs and outputs of aggregation rules.

Universal domain. The domain of F is the set of all possible profiles of consistent and complete individual judgment sets.

Collective rationality. F generates consistent and complete collective judgment sets.

Propositionwise majority voting, dictatorships and inverse dictatorships satisfy universal domain, but only dictatorships generally satisfy collective rationality. As the discursive paradox example of Table 1 shows, propositionwise majority voting sometimes generates inconsistent collective judgment sets. Inverse dictatorships satisfy collective rationality only in special cases (i.e., when the agenda is *symmetrical*: for every consistent $Z \subseteq X$, $\{\neg p : p \in Z\}$ is also consistent).

3 Two impossibility theorems on judgment aggregation

Are there any non-dictatorial judgment aggregation rules satisfying universal domain and collective rationality? The following conditions are frequently used in the literature.

Independence. For any proposition $p \in X$ and profiles $(A_1, \ldots, A_n)$, $(A_1^*, \ldots, A_n^*) \in \mathit{Domain}(F)$, if [for all individuals i, $p \in A_i$ if and only if $p \in A_i^*$] then [$p \in F(A_1, \ldots, A_n)$ if and only if $p \in F(A_1^*, \ldots, A_n^*)$].

Systematicity. For any propositions $p, q \in X$ and profiles $(A_1, \ldots, A_n)$, $(A_1^*, \ldots, A_n^*) \in \mathit{Domain}(F)$, if [for all individuals i, $p \in A_i$ if and only if $q \in A_i^*$] then [$p \in F(A_1, \ldots, A_n)$ if and only if $q \in F(A_1^*, \ldots, A_n^*)$].

Unanimity principle. For any profile $(A_1, \ldots, A_n) \in \mathit{Domain}(F)$ and any proposition $p \in X$, if $p \in A_i$ for all individuals i, then $p \in F(A_1, \ldots, A_n)$.

Independence requires that the collective judgment on each proposition should depend only on individual judgments on that proposition. Systematicity strengthens independence by requiring in addition that the same pattern of dependence should hold for all propositions (a neutrality condition). The unanimity principle requires that if all individuals accept a proposition then this proposition should also be collectively accepted. The following result holds.

Proposition 1. For a minimally connected agenda X, an aggregation rule F satisfies universal domain, collective rationality, systematicity and the unanimity principle if and only if it is a dictatorship of some individual.

Proof. All proofs are given in the appendix. ■

Proposition 1 is related to an earlier result by Dietrich (forthcoming), which requires an additional assumption on the agenda X but no unanimity principle (the additional assumption is that X is also *asymmetrical*: for some inconsistent $Z \subseteq X$, $\{\neg p : p \in Z\}$ is consistent). This result, in turn, generalizes an earlier result on systematicity by Pauly and van Hees (forthcoming).

From Proposition 1, we can derive two new results of interest. The first is a generalization of List and Pettit's (2002) theorem on the non-existence of an aggregation rule satisfying universal domain, collective rationality, systematicity and anonymity (i.e., invariance of the collective judgment set under permutations of the given profile of individual judgment sets). Our result extends the earlier impossibility result to any minimally connected agenda and weakens anonymity to the requirement that there is no dictator or inverse dictator.

Theorem 1. For a minimally connected agenda X, every aggregation rule F satisfying universal domain, collective rationality and systematicity is a (possibly inverse) dictatorship of some individual.

The agenda assumption of Theorem 1 cannot be weakened further if the agenda is finite or the logic is compact (and $n \geq 3$ and X contains at least one contingent proposition), i.e., minimal connectedness is also necessary (and not

just sufficient) for giving rise to (possibly inverse) dictatorships by the conditions of Theorem 1.[10]

The second result we can derive from Proposition 1 is the analogue of Arrow's theorem in judgment aggregation, from which we subsequently derive Arrow's theorem on (strict) preference aggregation as a corollary. We use the following lemma, which strengthens an earlier lemma by Nehring and Puppe (2002) by not requiring monotonicity.

Lemma 1. For a path-connected agenda X, an aggregation rule F satisfying universal domain, collective rationality, independence and the unanimity principle also satisfies systematicity.

Let us call an agenda *strongly connected* if it is both minimally connected and path-connected. Using Lemma 1, Proposition 1 now implies the following impossibility result.

Theorem 2. For a strongly connected agenda X, an aggregation rule F satisfies universal domain, collective rationality, independence and the unanimity principle if and only if it is a dictatorship of some individual.

Dokow and Holzman (2005) have independently shown that (for a finite agenda containing only contingent propositions) strong connectedness (in the form of the conjunction of non-affineness and total blockedness) is both necessary and sufficient for characterizing dictatorships by the conditions of Theorem 2 (assuming $n \geq 3$). A prior closely related result is Nehring and Puppe's (2002) characterization result, using total blockeness alone but imposing an additional monotonicity condition. In fact, within the general logics framework, the necessity holds if the agenda is finite or the logic is compact (and X contains at least one contingent proposition; again assuming $n \geq 3$).

Proposition 1 and Theorems 1 and 2 continue to hold under generalized definitions of minimally connected and strongly connected agendas.[11]

Of course, it is debatable whether and when independence or systematicity are plausible requirements on judgment aggregation. The literature contains

[10] It can then be shown that, if X is not minimally connected, there exists an aggregation rule that satisfies universal domain, collective rationality and systematicity and is not a (possibly inverse) dictatorship. Let M be a subset of $\{1, ..., n\}$ of odd size at least 3. If part (i) of minimal connectedness is violated, then majority voting among the individuals in M satisfies all requirements. If part (ii) is violated, the aggregation rule F with universal domain defined by $F(A_1, ..., A_n) := \{p \in X : \text{the number of individuals } i \in M \text{ with } p \in A_i \text{ is odd}\}$ satisfies all requirements. The second example is inspired by Dokow and Holzman (2005).

[11] In the definition of minimal connectedness, (i) can be weakened to the following: (i*) there is an inconsistent set $Y \subseteq X$ with pairwise disjoint subsets Z_1, Z_2, Z_3 such that $(Y \backslash Z_j) \cup \{\neg p : p \in Z_j\}$ is consistent for any $j \in \{1, 2, 3\}$ (Dietrich forthcoming). In the definition of strong connectedness (by (i), (ii) and path-connectedness), (i) can be dropped altogether, as path-connectedness implies (i*). In the definitions of minimal connectedness and strong connectedness, (ii) can be weakened to (ii*) in Dietrich (forthcoming).

extensive discussions of these conditions and their possible relaxations. In our view, the importance of Theorems 1 and 2 lies not so much in establishing the impossibility of consistent judgment aggregation, but rather in indicating what conditions must be relaxed in order to make consistent judgment aggregation possible. The theorems describe boundaries of the logical space of possibilities.

4 Arrow's theorem

We now show that Arrow's theorem (stated here for strict preferences) can be restated in the judgment aggregation model, where it is a direct corollary of Theorem 2. We consider a standard Arrowian preference aggregation model, where each individual has a strict preference ordering (asymmtrical, transitive and connected, as defined below) over a set of options $K = \{x, y, z, ...\}$ with $|K| \geq 3$. We embed this model into our judgment aggregation model by representing preference orderings as sets of binary ranking judgments in a simple predicate logic, following List and Pettit (2001/2004). Although we consider strict preferences for simplicity, we note that a similar embedding is possible for weak preferences.[12]

A simple predicate logic for representing preferences. We consider a predicate logic with constants $x, y, z, ... \in K$ (representing the options), variables v, w, v_1, v_2 , ..., identity symbol $=$, a two-place predicate P (representing strict preference), logical connectives $\neg$ (not), $\wedge$ (and), $\vee$ (or), $\rightarrow$ (if-then), and universal quantifier $\forall$. Formally, $\mathbf{L}$ is the smallest set such that

- $\mathbf{L}$ contains all propositions of the forms $\alpha P \beta$ and $\alpha = \beta$, where α and β are constants or variables, and
- whenever $\mathbf{L}$ contains two propositions p and q, then $\mathbf{L}$ also contains $\neg p$, $(p \wedge q)$, $(p \vee q)$, $(p \rightarrow q)$ and $(\forall v)p$, where v is any variable.

Notationally, we drop brackets when there is no ambiguity. The entailment relation $\vDash$ is defined as follows. For any set $A \subseteq \mathbf{L}$ and any proposition $p \in \mathbf{L}$,

$$A \vDash p \text{ if and only if } \begin{array}{l} A \cup Z \text{ entails } p \text{ in the standard} \\ \text{sense of predicate logic,} \end{array}$$

where Z is the set of rationality conditions on strict preferences:

$(\forall v_1)(\forall v_2)(v_1 P v_2 \rightarrow \neg v_2 P v_1)$ (asymmetry);
$(\forall v_1)(\forall v_2)(\forall v_3)((v_1 P v_2 \wedge v_2 P v_3) \rightarrow v_1 P v_3)$ (transitivity);
$(\forall v_1)(\forall v_2)(\neg\, v_1 = v_2 \rightarrow (v_1 P v_2 \vee v_2 P v_1))$ (connectedness).[13]

[12] If we represent *weak* preference aggregation in the judgment aggregation model using the embedding indicated below, the independence condition and the unanimity principle become stronger than Arrow's independence of irrelevant alternatives and the weak Pareto principle. So, in the case of weak preferences unlike that of strict ones, Theorem 2 only implies a slightly weaker form of Arrow's theorem.

[13] For technical reasons, Z also contains, for each pair of distinct constants x, y, the condition $\neg\, x = y$, reflecting the mutual exclusiveness of the options.

To represent weak preferences rather than strict ones, Z simply needs to be redefined as the set of rationality conditions on weak preferences (i.e., reflexivity, transitivity, and connectedness); see also Dietrich (forthcoming).[14] Binary relations with other properties can be represented analogously, by defining Z as the set of appropriate rationality conditions, e.g., the set containing reflexivity (respectively, asymmetry) and transitivity for weak (respectively, strict) partial orderings, and the set containing reflexivity, transitivity and symmetry for equivalence relations.

The agenda. The *preference agenda* is the set X of all propositions of the forms $xPy, \neg xPy \in \mathbf{L}$, where x and y are distinct constants.[15] Note the following lemma (which holds for strict as well as weak preferences). The path-connectedness part of the result is equivalent to a lemma by Nehring (2003).

Lemma 2. The preference agenda X is strongly connected.

The correspondence between preference orderings and judgment sets. It is easy to see that each (asymmetrical, transitive and connected) preference ordering over K can be represented by a unique consistent and complete judgment set in X and vice-versa, where individual i strictly prefers x to y if and only if $xPy \in A_i$. For example, if individual i strictly prefers x to y to z, this is uniquely represented by the judgment set $A_i = \{xPy, yPz, xPz, \neg yPx, \neg zPy, \neg zPx\}$.

The correspondence between Arrow's conditions and conditions on judgment aggregation. For the preference agenda, the conditions of universal domain, collective rationality, independence ("independence of irrelevant alternatives") and the unanimity principle ("the weak Pareto principle"), as stated above, exactly match the standard conditions of Arrow's theorem, where an Arrowian preference aggregation rule is represented by a judgment aggregation rule.

As the preference agenda is strongly connected, Arrow's theorem now follows from Theorem 2.

Corollary 1. (Arrow's theorem) For the preference agenda X, an aggregation rule F satisfies universal domain, collective rationality, independence and the unanimity principle if and only if it is a dictatorship of some individual.

[14] Transitivity and connectedness are as defined above. Reflexivity can be stated by the proposition $(\forall v)(vPv)$. For aesthetic reasons, one might also replace the predicate symbol P by R in the logic.

[15] xPy is interpreted as "x is better than/preferable to y". Note that this represents preference aggregation as the aggregation of *beliefs of betterness/preferability*. One might argue that preferences are desire-like rather than belief-like and thus object to re-interpreting them as beliefs of preferability. To respond to this objection, we might, for example, interpret xPy as "x is socially preferred to y", and interpret an individual judgment set $A_i \subseteq X$ as the set of propositions that individual i *desires* (rather than *believes*), while interpreting a collective judgment set $A \subseteq X$ as a set of propositions about social preference.

Corollary 1 strengthens Nehring's (2003) corollary by dropping monotonicity; it also strengthens List and Pettit's (2001/2004) corollary by weakening systematicity to independence and (effectively) anonymity to non-dictatorship, at the expense of imposing, in addition, the unanimity principle.

The correspondence between preference and judgment aggregation concepts under the constructed embedding is summarized in Table 2.

Preference aggregation	Judgment aggregation
Preference ordering over a set of options	Judgment set in the preference agenda
Three or more options	Strong connectedness of the preference agenda
Asymmetry, transitivity and connectedness of the preference ordering	Consistency and completeness of the judgment set
Preference aggregation rule	Judgment aggregation rule
Universal domain	Universal domain
Collective rationality	Collective rationality
Independence of irrelevant alternatives	Independence
Weak Pareto principle	Unanimity principle
Arrowian dictator	(Judgment) dictator
Arrow's theorem	Corollary of Theorem 2

Table 2: The embedding of concepts

5 Concluding remarks

After proving two impossibility theorems on judgment aggregation – Theorem 1 with systematicity and a weak agenda assumption, Theorem 2 with independence and a stronger agenda assumption – we have shown that Arrow's theorem (for strict preferences) is a corollary of Theorem 2, applied to the aggregation of binary ranking judgments in a simple predicate logic. In the case of binary relations other than preference orderings, Theorem 2 does not necessarily apply, as the resulting agenda is not necessarily path-connected. For example, if the binary relations in question are partial orderings or equivalence relations (as briefly mentioned above), the agenda is merely minimally connected; but Theorem 2 still yields an immediate corollary for the aggregation of profiles of such binary relations into corresponding collective binary relations: here every aggregation rule satisfying universal domain, collective rationality and systematicity is a (possibly inverse) dictatorship of some individual.

These findings illustrate the generality of judgment aggregation. Impossibility and possibility results such as Theorems 1 and 2 can apply to a large class

of aggregation problems formulated in a suitable logic – any logic satisfying conditions L1 to L3 – of which a predicate logic for representing preferences is a special case. Other logics to which the results apply are propositional, modal or conditional logics, some fuzzy logics as well as different predicate logics.

An alternative, very general model of aggregation is the one introduced by Wilson (1975) and used by Dokow and Holzman (2005), where a group has to determine its yes/no views on several issues based on the group members' views on these issues (subject to feasibility constraints). Wilson's model can also be represented in our model; Dokow and Holzman's results for Wilson's model apply to a logic satisfying L1 to L3 and a finite agenda.[16]

Although we have constructed an explicit embedding of preference aggregation into judgment aggregation, we have not proved the impossibility of a converse embedding. We suspect that such an embedding is hard to achieve, as Arrow's standard model cannot easily capture the different informational basis of judgment aggregation. It is unclear what an embedding of judgment aggregation into preference aggregation would look like. In particular, it is unclear how to specify the *options* over which individuals have preferences. The *propositions* in an agenda are not candidates for options, as propositions are usually not mutually exclusive. Natural candidates for options are perhaps entire *judgment sets* (consistent and complete), as these are mutually exclusive and exhaustive. But in a preference aggregation model with options thus defined, individuals would feed into the aggregation rule not a single judgment set (option), but an entire preference ordering over all possible judgment sets (options). This would be a different informational basis from the one in judgment aggregation. In addition, the explicit logical structure within each judgment set would be lost under this approach, as judgment sets in their entirety, not propositions, would be taken as primitives. However, the construction of a useful converse embedding or the proof of its non-existence remains a challenge.

6 References

Bovens L, Rabinowicz W (2006) Democratic Answers to Complex Questions: An Epistemic Perspective. Synthese 150(1): 131-153

Dietrich F (2006) Judgment Aggregation: (Im)Possibility Theorems. Journal of Economic Theory 126(1): 286-298

Dietrich F (forthcoming) A generalized model of judgment aggregation. Social Choice and Welfare

Dietrich F, List C (2005) Judgment aggregation by quota rules. Working paper, LSE

[16] In Wilson's model, the notion of consistency (feasibility) rather than that of entailment is a primitive. While the notion of entailment in our model fully specifies a notion of consistency, the converse does not hold for all logics satisfying L1 to L3.

Dokow E, Holzman R (2005) Aggregation of binary evaluations, Working paper, Technion Israel Institute of Technology

Gärdenfors P (forthcoming) An Arrow-like theorem for voting with logical consequences. Economics and Philosophy

Guilbaud GT (1966) Theories of the General Interest, and the Logical Problem of Aggregation. In Lazarsfeld PF, Henry NW (eds.) Readings in Mathematical Social Science, Cambridge/MA (MIT Press): 262-307

List C (2003) A Possibility Theorem on Aggregation over Multiple Interconnected Propositions. Mathematical Social Sciences 45(1): 1-13

List C (2004) A Model of Path Dependence in Decisions over Multiple Propositions. American Political Science Review 98(3): 495-513

List C (2005) The Probability of Inconsistencies in Complex Collective Decisions. Social Choice and Welfare 24: 3-32

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

List C, Pettit P (2001/2004) Aggregating Sets of Judgments: Two Impossibility Results Compared. Social and Political Theory Paper W20 (technical report ID 931), Australian National University, published in Synthese 140(1-2): 207-235

Nehring K (2003) Arrow's theorem as a corollary. Economics Letters 80: 379-382

Nehring K, Puppe C (2002) Strategyproof Social Choice on Single-Peaked Domains: Possibility, Impossibility and the Space Between. Working paper, University of California at Davis

Nehring K, Puppe C (2005) Consistent Judgment Aggregation: A Characterization. Working paper, University of Karlsruhe

Pauly M, van Hees M (forthcoming) Logical Constraints on Judgment Aggregation. Journal of Philosophical Logic

Pettit P (2001) Deliberative Democracy and the Discursive Dilemma. Philosophical Issues 11: 268-299

Pigozzi G (forthcoming) Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. Synthese

van Hees M (forthcoming) The limits of epistemic democracy. Social Choice and Welfare

Wilson R (1975) On the Theory of Aggregation. Journal of Economic Theory 10: 89-99

A Appendix

Proof of Proposition 1. Let X be minimally connected and let F be any aggregation rule. Put $N := \{1, ..., n\}$. If F is dictatorial, F obviously satisfies universal domain, collective rationality, systematicity and the unanimity principle. Now assume F satisfies the latter conditions. Then there is a set

$\mathcal{C}$ of ("winning") coalitions $C \subseteq N$ such that, for every $p \in X$ and every $(A_1, ..., A_n) \in Domain(F)$, $F(A_1, ..., A_n) = \{p \in X : \{i : p \in A_i\} \in \mathcal{C}\}$. For every consistent set $Z \subseteq X$, let A_Z be some consistent and complete judgment set such that $Z \subseteq A_Z$.

Claim 1. $N \in \mathcal{C}$, and, for every coalition $C \subseteq N$, $C \in \mathcal{C}$ if and only if $N \backslash C \notin \mathcal{C}$.

The first part of the claim follows from the unanimity principle, and the second part follows from collective rationality together with universal domain.

Claim 2. For any coalitions $C, C^* \subseteq N$, if $C \in \mathcal{C}$ and $C \subseteq C^*$ then $C^* \in \mathcal{C}$.

Let $C, C^* \subseteq N$ with $C \in \mathcal{C}$ and $C \subseteq C^*$. Assume for contradiction that $C^* \notin \mathcal{C}$. Then $N \backslash C^* \in \mathcal{C}$. Let Y be as in part (ii) of the definition of minimally connected agendas, and let Z be a *smallest* subset of Y such that $(Y \backslash Z) \cup \{\neg z : z \in Z\}$ is consistent and Z has even size. We have $Z \neq \emptyset$, since otherwise the (inconsistent) set Y would equal the (consistent) set $(Y \backslash Z) \cup \{\neg z : z \in Z\}$. So, as Z has even size, there are two distinct propositions $p, q \in Z$. Since Y is minimal inconsistent, $(Y \backslash \{p\}) \cup \{\neg p\}$ and $(Y \backslash \{q\}) \cup \{\neg q\}$ are each consistent. This and the consistency of $(Y \backslash Z) \cup \{\neg z : z \in Z\}$ allow us to define a profile $(A_1, ..., A_n) \in Domain(F)$ as follows. Putting $C_1 := C^* \backslash C$ and $C_2 := N \backslash C^*$ (note that $\{C, C_1, C_2\}$ is a partition of N), let

$$A_i := \begin{cases} A_{(Y \backslash \{p\}) \cup \{\neg p\}} & \text{if } i \in C \\ A_{(Y \backslash Z) \cup \{\neg z : z \in Z\}} & \text{if } i \in C_1 \\ A_{(Y \backslash \{q\}) \cup \{\neg q\}} & \text{if } i \in C_2. \end{cases} \quad (1)$$

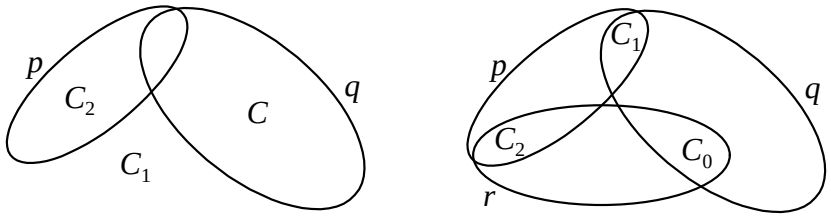

Figure 1: The profiles constructed in the proofs of claims 2 (left) and 3 (right).

By (1), we have $Y \backslash Z \subseteq F(A_1, ..., A_n)$ as $N \in \mathcal{C}$. Also by (1), we have $q \in F(A_1, ..., A_n)$ as $C \in \mathcal{C}$, and $p \in F(A_1, ..., A_n)$ as $C_2 = N \backslash C^* \in \mathcal{C}$. In summary, writing $Z^* := Z \backslash \{p, q\}$, we have (*) $Y \backslash Z^* \subseteq F(A, ..., A_n)$. We distinguish two cases.

Case $C_1 \notin \mathcal{C}$. Then $C \cup C_2 = N \backslash C_1 \in \mathcal{C}$. So $Z^* \subseteq F(A_1, ..., A_n)$ by (1), which together with (*) implies $Y \subseteq F(A_1, ..., A_n)$. But then $F(A_1, ..., A_n)$ is inconsistent, a contradiction.

Case $C_1 \in \mathcal{C}$. So $\{\neg z : z \in Z^*\} \subseteq F(A_1, ..., A_n)$ by (1). This together with (*) implies that $(Y \backslash Z^*) \cup \{\neg z : z \in Z^*\} \subseteq F(A_1, ..., A_n)$. So $(Y \backslash Z^*) \cup \{\neg z :$

$z \in Z^*\}$ is consistent. As Z^* also has even size, the minimality condition in the definition of Z is violated.

Claim 3. For any coalitions $C, C^* \subseteq N$, if $C, C^* \in \mathcal{C}$ then $C \cap C^* \in \mathcal{C}$.

Consider any $C, C^* \in \mathcal{C}$. Let $Y \subseteq X$ be as in part (i) of the definition of minimally connected agendas. As $|Y| \geq 3$, there are pairwise distinct propositions $p, q, r \in Y$. As Y is minimally inconsistent, each of the sets $(Y \backslash \{p\}) \cup \{\neg p\}$, $(Y \backslash \{q\}) \cup \{\neg q\}$ and $(Y \backslash \{r\}) \cup \{\neg r\}$ is consistent. This allows us to defined a profile $(A_1, ..., A_n) \in Domain(F)$ as follows. Putting $C_0 := C \cap C^*$, $C_1 := C^* \backslash C$ and $C_2 := N \backslash C^*$ (note that $\{C_0, C_1, C_2\}$ is a partition of N), let

$$A_i := \begin{cases} A_{(Y \backslash \{p\}) \cup \{\neg p\}} & \text{if } i \in C_0 \\ A_{(Y \backslash \{r\}) \cup \{\neg r\}} & \text{if } i \in C_1 \\ A_{(Y \backslash \{q\}) \cup \{\neg q\}} & \text{if } i \in C_2. \end{cases} \tag{2}$$

By (2), $Y \backslash \{p, q, r\} \subseteq F(A_1, ..., A_n)$ as $N \in \mathcal{C}$. Again by (2), we have $q \in F(A_1, ..., A_n)$ as $C_0 \cup C_1 = C^* \in \mathcal{C}$. As $C \in \mathcal{C}$ and $C \subseteq C_0 \cup C_2$, we have $C_0 \cup C_2 \in \mathcal{C}$ by claim 2. So, by (2), $r \in F(A_1, ..., A_n)$. In summary, $Y \backslash \{p\} \subseteq F(A_1, ..., A_n)$. As Y is inconsistent, $p \notin F(A_1, ..., A_n)$, and hence $\neg p \in F(A_1, ..., A_n)$. So, by (2), $C_0 \in \mathcal{C}$.

Claim 4. There is a dictator.

Consider the intersection of all winning coalitions, $\widetilde{C} := \cap_{C \in \mathcal{C}} C$. By claim 3, $\widetilde{C} \in \mathcal{C}$. So $\widetilde{C} \neq \emptyset$, as by claim 1 $\emptyset \notin \mathcal{C}$. Hence there is a $j \in \widetilde{C}$. As j belongs to every winning coalition $C \in \mathcal{C}$, j is a dictator: indeed, for each profile $(A_1, ..., A_n) \in Domain(F)$ and each $p \in X$, if $p \in A_j$ then $\{i : p \in A_i\} \in \mathcal{C}$, so that $p \in F(A_1, ..., A_n)$; and if $p \notin A_i$ then $\neg p \in A_i$, so that $\{i : \neg p \in A_i\} \in \mathcal{C}$, implying $\neg p \in F(A_1, ..., A_n)$, and hence $p \notin F(A_1, ..., A_n)$. ■

Proof of Theorem 1. Let X be minimally connected, and let F satisfy universal domain, collective rationality and systematicity. If F satisfies the unanimity principle, then, by Proposition 1, F is dictatorial. Now suppose F violates the unanimity principle.

Claim 1. X is symmetrical, i.e., if $A \subseteq X$ is consistent, so is $\{\neg p : p \in A\}$.

Let $A \subseteq X$ be consistent. Then there exists a consistent and complete judgment set B such that $A \subseteq B$. As F violates the unanimity principle (but satisfies systematicity), the set $F(B, ..., B)$ contains no element of B, hence contains no element of A, hence contains all elements of $\{\neg p : p \in A\}$ by collective rationality. So, again by collective rationality, $\{\neg p : p \in A\}$ is consistent.

Claim 2. The aggregation rule $\widehat{F}$ with universal domain defined by $\widehat{F}(A_1, ..., A_n) := \{\neg p : p \in F(A_1, ..., A_n)\}$ is dictatorial.

As F satisfies collective rationality and systematicity, so does $\widehat{F}$, where the consistency of collective judgment sets follows from claim 1. $\widehat{F}$ also satisfies the unanimity principle: for any $p \in X$ and any $(A_1, ..., A_n)$ in the universal domain, where $p \in A_i$ for all i, $p \notin F(A_1, ..., A_n)$, hence $\neg p \in F(A_1, ..., A_n)$, and so $p = \neg\neg p \in \widehat{F}(A_1, ..., A_n)$. Now Proposition 1 applies to $\widehat{F}$, and hence $\widehat{F}$ is dictatorial.

Claim 3. F is inverse dictatorial.
The dictator for $\widehat{F}$ is an inverse dictator for F. ■

Proof of Lemma 1. Let X and F be as specified. To show that F is systematic, consider any $p, q \in X$ and any $(A_1, ..., A_n), (A_1^*, ..., A_n^*) \in Domain(F)$ such that $C := \{i : p \in A_i\} = \{i : q \in A_i^*\}$, and let us prove that $p \in F(A_1, ..., A_n)$ if and only if $q \in F(A_1^*, ..., A_n^*)$. If p and q are both tautologies ($\{\neg p\}$ and $\{\neg q\}$ are inconsistent), the latter holds since (by collective rationality) $p \in F(A_1, ..., A_n)$ and $q \in F(A_1^*, ..., A_n^*)$. If p and q are both contradictions ($\{p\}$ and $\{q\}$ are inconsistent), it holds since (by collective rationality) $p \notin F(A_1, ..., A_n)$ and $q \notin F(A_1^*, ..., A_n^*)$. It is impossible that one of p and q is a tautology and the other a contradiction, because then one of $\{i : p \in A_i\}$ and $\{i : q \in A_i^*\}$ would be N and the other $\emptyset$.

Now consider the remaining case where both p and q are contingent. We say that C is *winning for* r $(\in X)$ if $r \in F(B_1, ..., B_n)$ for some (hence by independence any) profile $(B_1, .., B_n) \in Domain(F)$. with $\{i : r \in B_i\} = C$. We have to show that C is winning for p if and only if C is winning for q. Suppose C is winning for p, and let us show that C is winning for q (the converse implication can be shown analogously). As X is path-connected and p and q are contingent, there are $p = p_1, p_2, ..., p_k = q \in X$ such that $p_1 \vDash^* p_2$, $p_2 \vDash^* p_3$, ..., $p_{k-1} \vDash^* p_k$. We show by induction that C is winning for each of $p_1, p_2, ..., p_k$. If $j = 1$ then C is winning for p_1 by $p_1 = p$. Now let $1 \leq j < k$ and assume C is winning for p_j. We show that C is winning for p_{j+1}. By $p_j \vDash^* p_{j+1}$, there is a set $Y \subseteq X$ such that (i) $\{p_j\} \cup Y$ and $\{\neg p_{j+1}\} \cup Y$ are each consistent, and (ii) $\{p_j, \neg p_{j+1}\} \cup Y$ is inconsistent. Using (i) and (ii), the sets $\{p_j, p_{j+1}\} \cup Y$ and $\{\neg p_j, \neg p_{j+1}\} \cup Y$ are each consistent. So there exists a profile $(B_1, ..., B_n) \in Domain(F)$ such that $\{p_j, p_{j+1}\} \cup Y \subseteq B_i$ for all $i \in C$ and $\{\neg p_j, \neg p_{j+1}\} \cup Y \subseteq B_i$ for all $i \notin C$. Since $Y \subseteq A_i$ for all i, $Y \subseteq F(A_1, ..., A_n)$ by the unanimity principle. Since $\{i : p_j \in A_i\} = C$ is winning for p_j, we have $p_j \in F(A_1, ..., A_n)$. So $\{p_j\} \cup Y \subseteq F(A_1, ..., A_n)$. Hence, using collective rationality and (ii), we have $\neg p_{j+1} \notin F(A_1, ..., A_n)$, and so $p_{j+1} \in F(A_1, ..., A_n)$. Hence, as $\{i : p_{j+1} \in A_i\} = C$, C is winning for p_{j+1}. ■

Proof of Lemma 2. Let X be the preference agenda. X is minimally connected, as, for any pairwise distinct constants x, y, z, the set $Y = \{xPy, yPz, zPx\} \subseteq X$ is minimal inconsistent, where $\{\neg xPy, \neg yPz, zPx\}$ is consistent.

To prove path-connectedness, note that, by the axioms of our predicate logic for representing preferences, (*) $\neg xPy$ and yPx are equivalent (i.e., entail each other) for any distinct $x, y \in K$. Now consider any (contingent) $p, q \in X$, and let us construct a sequence $p = p_1, p_2, ..., p_k = q \in X$ with $p \models^* p_2, ..., p_{k-1} \models^* q$. By (*), if p is a negated proposition $\neg xPy$, then p is equivalent to the non-negated proposition yPx ; and similarly for q. So we may assume without loss of generality that p and q are non-negated propositions, say p is xPy and q is $x'Py'$. We distinguish three cases, each with subcases.

Case $x = x'$. If $y = y'$, then $xPy \vDash^* xPy = x'Py'$ (take $Y = \emptyset$). If $y \neq y'$, then $xPy \vDash^* xPy' = x'Py'$ (take $Y = \{yPy'\}$).

Case $x = y'$. If $y = x'$, then, taking any $z \in K \backslash \{x, y\}$, we have $xPy \vDash^* xPz$ (take $Y = \{yPz\}$), $xPz \vDash^* yPz$ (take yPx), and $yPz \vDash^* yPx = x'Py'$ (take $Y = \{zPx\}$). If $y \neq x'$, then $xPy \vDash^* x'Py$ (take $Y = \{x'Px\}$) and $x'Py \vDash^* x'Py'$ (take $Y = \{yPy'\}$).

Case $x \neq x', y'$. If $y = x'$, then $xPy \vDash^* xPy'$ (take $Y = \{yPy'\}$) and $xPy' \vDash^* x'Py'$ (take $Y = \{x'Px\}$). If $y = y'$, then $xPy \vDash^* x'Py = x'Py'$ (take $Y = \{x'Px\}$). If $y \neq x', y'$, then $xPy \vDash^* x'Py'$ (take $Y = \{x'Px, yPy'\}$). ■

Chapter 3

A generalisation of Sen's Impossibility of a Paretian Liberal

Paper: A liberal paradox for judgment aggregation (with C. List), *Social Choice and Welfare* 31(1): 59-78, 2008

A Liberal Paradox for Judgment Aggregation

Franz Dietrich and Christian List[1]

In the emerging literature on judgment aggregation over logically connected propositions, expert rights or liberal rights have not been investigated yet. A group making collective judgments may assign individual members or subgroups with expert knowledge on, or particularly affected by, certain propositions the right to determine the collective judgment on those propositions. We identify a problem that generalizes Sen's 'liberal paradox'. Under plausible conditions, the assignment of rights to two or more individuals or subgroups is inconsistent with the unanimity principle, whereby unanimously accepted propositions are collectively accepted. The inconsistency can be avoided if individual judgments or rights satisfy special conditions.

1 Introduction

Groups frequently make collective judgments on certain propositions. Examples are legislatures, committees, courts, juries, expert panels and entire populations deciding what propositions to accept as true (thus forming *collective beliefs*) and what propositions to make true through their actions (thus forming *collective desires*). When a group forms collective beliefs, some group members or subgroups may have expert knowledge on certain propositions and may therefore be granted the right to be decisive on those propositions (an *expert right*). Legislatures or expert panels, for example, may grant such rights to specialist members or subcommittees so as to rely on their expertise or to achieve a division of labour. When a group forms collective desires, some group members or subgroups may be particularly affected by certain propositions, for example when those propositions concern their private sphere(s), and may also be granted the right to be decisive on those propositions (a *liberal right*).

How does the assignment of rights constrain a group's collective judgments? In this paper, we identify a problem that generalizes Sen's 'liberal paradox' (1970), the result that individual rights may conflict with the Pareto principle (for recent contributions, see Deb, Pattanaik and Razzolini 1997; van Hees 1999, 2004; Dowding and van Hees 2003). Consider the following two examples.[2]

[1]F. Dietrich, Dept. of Quant. Economics, University of Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands; C. List, Dept. of Government, London School of Economics, London WC2A 2AE, U.K. Earlier versions of this paper were presented at the LGS-4 Conference in Caen, June 2005, and at the World Congress of the Econometric Society in London, August 2005. We are grateful for the comments we received at both occasions as well as from anonymous referees. Franz Dietrich acknowledges financial support from the European Commission-DG Research Sixth Framework Programme (CIT-2-CT-2004-506084 / Polarization and Conflict Project). Christian List acknowledges the hospitality of the Social and Political Theory Program, RSSS, Australian National University.

[2]In the expert rights example, accepted propositions are interpreted as propositions *believed* to be true; in the liberal rights example, as propositions *desired* to be true.

Example 1: expert rights.[3] An expert committee has to make judgments on the following propositions:

a: Carbon dioxide emissions are above some critical threshold.

b: There will be global warming.

$a \rightarrow b$: If carbon dioxide emissions are above the threshold, then there will be global warming.

Half of the committee members are experts on a, the other half experts on $a \rightarrow b$. So the committee assigns to the first half the right to determine the collective judgment on a and to the second a similar right on $a \rightarrow b$. The committee's constitution further stipulates that unanimous individual judgments must be respected. Now suppose that all the experts on a judge a to be true, and all the experts on $a \rightarrow b$ judge $a \rightarrow b$ to be true. In accordance with the expert rights, the committee accepts both a and $a \rightarrow b$. We may therefore expect it to accept b as well. But when a vote is taken on b, *all* committee members reject b. How can this happen? Table 1 shows the committee members' judgments on all propositions.

	a	$a \rightarrow b$	b
Experts on a	True	False	False
Experts on $a \rightarrow b$	False	True	False

Table 1: A paradox of expert rights

The experts on a accept a, but reject $a \rightarrow b$ and b. The experts on $a \rightarrow b$ accept $a \rightarrow b$, but reject a and b. So all committee members are individually consistent. Nonetheless, respecting the rights of the experts on a and $a \rightarrow b$ is inconsistent with respecting the committee's unanimous judgment on b. To achieve consistency, the committee must either restrict the expert rights or overrule its unanimous judgment on b.

Example 2: liberal rights.[4] The two members of a small society, Lewd and Prude, each have a personal copy of the book *Lady Chatterley's Lover*. Consider three propositions:

l: Lewd reads the book.

p: Prude reads the book.

$l \rightarrow p$: If Lewd reads the book, then so does Prude.

Lewd desires to read the book himself, and that, if he reads it, then Prude read it too, as he anticipates that his own pleasure of reading the book will be enhanced by the thought of Prude finding the book offensive. Prude, by contrast, desires not to read the book, and that Lewd not read it either, as he fears that the book would corrupt Lewd's moral outlook. But he also desires

[3] A structurally similar example was given by Pauly and van Hees (2006).

[4] This example is inspired by Sen's example. While in Sen's example there is only one copy of the book – to be borrowed and read by at most one individual – in ours there are two copies; so the book may be read by both individuals, by one, or by neither.

that, if Lewd reads the book, then he read it too, so as to be informed about the dangerous material Lewd is exposed to. Table 2 shows Lewd's and Prude's desires on the propositions.[5]

	l	p	$l \rightarrow p$
Lewd	True	True	True
Prude	False	False	True

Table 2: A paradox of liberal rights

Society assigns to each individual the liberal right to determine the collective desire on those propositions that concern only the individual's private sphere. Since l and p are such propositions for Lewd and Prude, respectively, society assigns to Lewd the right to determine the collective desire on l, and to Prude a similar right on p. Further, according to society's constitution, unanimous desires of all individuals must be collectively respected. But because of Lewd's liberal right on l, l is collectively accepted; because of Prude's liberal right on p, p is collectively rejected; and yet, by unanimity, $l \rightarrow p$ is collectively accepted, an inconsistent collective set of desires. To achieve consistency, society must either restrict the liberal rights of the individuals or relax its constitutional principle of respecting unanimous desires.

In both examples, there is a conflict between some individuals' rights on some propositions and all individuals' unanimous judgments on others. This conflict is not accidental. We show that, as soon as the relevant propositions exhibit mild interconnections, no consistent mapping from individual to collective judgments can generally respect the rights of two or more individuals or subgroups and preserve unanimous judgments. Except in special cases, which we discuss later, respecting such rights may require overruling unanimity. We also derive Sen's original result as a corollary of our new result.

We present our result within the model of judgment aggregation on logically connected propositions, initially proposed by List and Pettit (2002), which combines axiomatic social choice theory and formal logic. Much of this literature has focused on generalizations of, and solutions to, another paradox, the 'doctrinal' or 'discursive' paradox (Kornhauser and Sager 1986, Pettit 2001), which is similar in spirit to Condorcet's famous paradox of cyclical majority preferences and consists in the fact that majority voting on logically connected propositions may lead to inconsistent majority judgments (for generalizations, see, e.g., List and Pettit 2002, 2004; Pauly and van Hees 2006; van Hees 2007;

[5] Conditional desires, like Lewd's and Prude's desire of p given l, can be represented in various ways, which are controversially discussed in deontic logic. Our example represents a conditional desire of p given l as a desire of the implication $l \rightarrow p$, as distinct from a desire of p on the supposition/condition that l. A further question is whether '$\rightarrow$' should be a material or subjunctive conditional (our example works either way). See, e.g., Hintikka (1971), Wagner Decew (1981), Bradley (1999).

Dietrich 2006, 2007a; Nehring and Puppe 2006; Dietrich and List 2007; Dokow and Holzman 2005; for proposed solutions, see, e.g., List 2003, 2004a; Pigozzi 2006; Dietrich forthcoming, 2007b).[6] This paper, however, presents the first extension of Sen's liberal paradox to judgment aggregation. The use of formal logic illuminates the logical structure of the paradox and highlights its robustness. All proofs are given in the appendix.

2 The model

We consider a group of individuals $N = \{1, 2, ..., n\}$ ($n \geq 2$). The propositions on which judgments are made are represented in logic (following List and Pettit 2002, 2004; we use Dietrich's 2007a generalization).

Logic. Let $\mathbf{L}$ be a set of sentences, called *propositions*, closed under negation (i.e., if $p \in \mathbf{L}$ then $\neg p \in \mathbf{L}$, where $\neg$ denotes 'not'), and stipulate that each subset $S \subseteq \mathbf{L}$ is either *consistent* or *inconsistent*, subject to standard axioms.[7] In standard propositional logic, $\mathbf{L}$ contains propositions such as a, b, $a \wedge b$, $a \vee b$, $\neg(a \rightarrow b)$ (where $\wedge$, $\vee$, $\rightarrow$ denote 'and', 'or', 'if-then', respectively). Examples of consistent sets are $\{a, a \rightarrow b, b\}$ and $\{a \wedge b\}$, examples of inconsistent ones $\{a, \neg a\}$ and $\{a, a \rightarrow b, \neg b\}$. A proposition $p \in \mathbf{L}$ is a *tautology* if $\{\neg p\}$ is inconsistent and a *contradiction* if $\{p\}$ is inconsistent.

Agenda. The *agenda* is the set of propositions on which judgments are made, defined as a non-empty subset $X \subseteq \mathbf{L}$ expressible as $X = \{p, \neg p : p \in X_+\}$ for a set $X_+ \subseteq \mathbf{L}$ of unnegated propositions. We assume that X contains no tautologies or contradictions[8] and that double negations cancel each other out (i.e., $\neg\neg p$ stands for p).[9] In our examples, $X = \{a, \neg a, a \rightarrow b, \neg(a \rightarrow b), b, \neg b\}$ and $X = \{l, \neg l, l \rightarrow p, \neg(l \rightarrow p), p, \neg p\}$ (in standard propositional or conditional logic).

Individual judgment sets. Each individual i's judgment set is the set $A_i \subseteq X$ of propositions that he or she accepts. On a belief interpretation, A_i is the set of propositions believed by individual i to be true; on a desire interpretation, the set of propositions desired by individual i to be true. A judgment set is

[6] Related contributions are those on abstract aggregation theory (Wilson 1975, Rubinstein and Fishburn 1986, Nehring and Puppe 2002) and belief merging in computer science (Konieczny and Pino-Perez 2002).

[7] C1: For any $p \in \mathbf{L}$, $\{p, \neg p\}$ is inconsistent. C2: If $S \subseteq \mathbf{L}$ is inconsistent, then so is any superset $T \supseteq S$ (in $\mathbf{L}$). C3: $\emptyset$ is consistent, and each consistent $S \subseteq \mathbf{L}$ has a consistent superset $T \supseteq S$ (in $\mathbf{L}$) containing a member of each pair $p, \neg p \in \mathbf{L}$. See Dietrich (2007a).

[8] This assumption is only needed in theorem 4 (where it could be avoided, for instance, by supposing that different individuals have disjoint rights sets).

[9] Hereafter, when we write $\neg p$ and p is already of the form $\neg q$, we mean q (rather than $\neg\neg q$).

consistent if it is a consistent set in **L** and *complete* if it contains a member of each proposition-negation pair $p, \neg p \in X$. A *profile* is an n-tuple $(A_1, ..., A_n)$ of individual judgment sets.

Aggregation functions. An *aggregation function* is a function F that maps each profile $(A_1, ..., A_n)$ from some domain of admissible ones to a collective judgment set $F(A_1, ..., A_n) = A \subseteq X$, the set of propositions that the group as a whole accepts. The collective judgment set A can be interpreted as the set of propositions collectively believed to be true or as the set collectively desired to be true. Below we impose minimal conditions on aggregation functions (including on the domain of admissible profiles). Standard examples of aggregation functions are *majority voting* (where $F(A_1, ..., A_n)$ is the set of propositions $p \in X$ for which the number of individuals with $p \in A_i$ exceeds that with $p \notin A_i$) and *dictatorships* (where $F(A_1, ..., A_n) = A_i$ for some antecedently fixed individual $i \in N$).

3 Impossibility results

We first state an impossibility result on the assignment of (expert or liberal) rights to individuals; we then state a similar result on the assignment of rights to subgroups. Following Sen's (1970) account of rights, we formalize rights in terms of a suitable notion of decisiveness. In the next section, we show that Sen's result is a corollary of ours.

Our impossibility results hold for all agendas exhibiting 'mild' interconnections in the following sense. Call propositions $p, q \in X$ *conditionally dependent* if there exist $p^* \in \{p, \neg p\}$ and $q^* \in \{q, \neg q\}$ such that $\{p^*, q^*\} \cup Y$ is inconsistent for some $Y \subseteq X$ consistent with each of p^* and q^*. The agenda X is *connected* if any two propositions $p, q \in X$ are conditionally dependent. Notice that the agendas in the two examples above are connected in this sense.

3.1 Individual rights

Call individual i *decisive* on a set of propositions $Y \subseteq X$ (under the aggregation function F) if any proposition in Y is collectively accepted if and only if it is accepted by i, formally

$$F(A_1, ..., A_n) \cap Y = A_i \cap Y.$$

Suppose we want to find an aggregation function with the following properties:

Universal Domain. The domain of F is the set of all possible profiles of consistent and complete individual judgment sets.

Minimal Rights. There exist (at least) two individuals who are each decisive on (at least) one proposition-negation pair $\{p, \neg p\} \subseteq X$.

Unanimity Principle. For any profile $(A_1, ..., A_n)$ in the domain of F and any proposition $p \in X$, if $p \in A_i$ for all individuals i, then $p \in F(A_1, ..., A_n)$.

Like Sen's (1970) condition of *minimal liberalism*, minimal rights is a weak requirement that leaves open which individuals have rights and to which propositions these rights apply. By using an undemanding rights requirement, our impossibility result becomes stronger. In a later section, we introduce explicit rights systems and state a stronger rights requirement.

Theorem 1 *If (and only if) the agenda is connected, there exists no aggregation function (generating consistent collective judgment sets) that satisfies universal domain, minimal rights and the unanimity principle.*[10]

So a group whose aggregation function has universal domain cannot *both* assign (liberal or expert) rights to more than one individual *and* respect unanimous judgments.

The result does not require complete collective judgment sets, only consistent ones. But, like all later results except theorem 4, it continues to hold if we add the completeness requirement on collective judgment sets. Further, theorem 1 continues to hold if decisiveness in minimal rights is weakened to *positive* decisiveness, where individual i is *positively decisive* on a set of propositions $Y \subseteq X$ (under the aggregation function F) if $F(A_1, ..., A_n) \cap Y \supseteq A_i \cap Y$. It also continues to hold if F is required to generate consistent *and* complete judgment sets and decisiveness in minimal rights is weakened to *negative* decisiveness (the presence of veto power), where individual i is *negatively decisive* on a set of propositions $Y \subseteq X$ (under the aggregation function F) if $F(A_1, ..., A_n) \cap Y \subseteq A_i \cap Y$. (Decisiveness *simpliciter* is the conjunction of positive and negative decisiveness.) Without a connected agenda, a modified impossibility holds in which minimal rights is strengthened to the requirement that there exist (at least) two individuals who are each decisive on (at least) one proposition-negation pair in X such that these two pairs are conditionally dependent.

3.2 Subgroup rights

A *subgroup* is a non-empty subset $M \subseteq N$. Call M *decisive* on a set of propositions $Y \subseteq X$ (under the aggregation function F) if any proposition in Y accepted by all members of M is also collectively accepted and any proposition in Y rejected by all members of M is also collectively rejected, formally

$$\bigcap_{i \in M} (A_i \cap Y) \subseteq F(A_1, ..., A_n) \cap Y \text{ and } \bigcap_{i \in M} (Y \backslash A_i) \subseteq Y \backslash F(A_1, ..., A_n).$$

[10] In this and later results, some parts are put in brackets in order to focus the attention on the other parts. The requirement of consistent collective judgment sets is left implicit in some of the informal discussion that follows.

If M is singleton, this definition reduces to the one in the individual case. In the interest of strength of the next theorem, we have deliberately given an undemanding definition of subgroup decisiveness. For a subgroup to be decisive on a set of propositions, it suffices that the subgroup can determine the collective judgments on them when its members unanimously agree on them; without unanimity, there are no constraints. Stronger forms of subgroup decisiveness are imaginable. One may require, for example, that the subgroup can determine the collective judgment on the relevant propositions by taking majority votes on them. However, are there any aggregation functions that satisfy the following rights condition with decisiveness defined in the present weak sense?

Minimal Subgroup Rights. There exist (at least) two disjoint subgroups that are each decisive on (at least) one proposition-negation pair $\{p, \neg p\} \subseteq X$.

Theorem 2 *If (and only if) the agenda is connected, there exists no aggregation function (generating consistent collective judgment sets) that satisfies universal domain, minimal subgroup rights and the unanimity principle.*

So a group whose aggregation function has universal domain cannot *both* assign (liberal or expert) rights to more than one subgroup *and* respect unanimous judgments among its members. Theorem 2 strengthens theorem 1, because minimal subgroup rights is less demanding than minimal rights (the latter implies the former – take singleton subgroups – but not vice-versa).[11] As in the case of theorem 1, theorem 2 continues to hold if the notion of decisiveness in minimal subgroup rights is weakened to *positive* decisiveness (the first conjunct in the definition above) or (when collective judgment sets are also required to be complete) to *negative* decisiveness (the second conjunct in the definition).

4 Sen's liberal paradox

To show that our main result generalizes Sen's 'liberal paradox' (1970), we apply theorem 1 to the aggregation of (strict) preference relations (using a construction in Dietrich and List 2007; see also List and Pettit 2004). For this purpose, we define a simple predicate logic $\mathbf{L}$, with

- a two-place predicate P (representing strict preference), and
- a set of (two or more) constants $K = \{x, y, z, ...\}$ (representing alternatives),

where any set $S \subseteq \mathbf{L}$ is *inconsistent* if and only if $S \cup Z$ is inconsistent in the standard sense of predicate logic, with Z defined as the set of rationality axioms on strict preferences:

[11] Except in the special case $n = 2$, where the two conditions are equivalent.

$$Z = \left\{ \begin{array}{c} (\forall v_1)(\forall v_2)(v_1 P v_2 \rightarrow \neg v_2 P v_1) \text{ (asymmetry)}, \\ (\forall v_1)(\forall v_2)(\forall v_3)((v_1 P v_2 \wedge v_2 P v_3) \rightarrow v_1 P v_3) \text{ (transitivity)}, \\ (\forall v_1)(\forall v_2)(\neg\, v_1 = v_2 \rightarrow (v_1 P v_2 \vee v_2 P v_1)) \text{ (connectedness)} \end{array} \right\}.^{12}$$

Thus the atomic propositions in **L** are binary ranking propositions of the form xPy, yPz etc.; examples of compound propositions are the axioms in Z. We discuss the interpretation in terms of preferences below. Sets such as $\{xPy, yPz\}$ are consistent, while sets such as $\{xPy, \neg xPy\}$, $\{xPy, yPx\}$, $\{xPy, yPz, zPx\}$, $\{\neg xPy, \neg yPx\}$ are inconsistent (the first set contains a proposition-negation pair; the second, third and fourth conflict with the first, second and third rationality axioms in Z, respectively).

The *preference agenda* is the set $X = \{xPy, \neg xPy \in \mathbf{L} : x, y \in K \text{ with } x \neq y\}$. The mapping that assigns to each fully rational (i.e., asymmetric, transitive and connected) preference relation $\succ$ on K the judgment set $A = \{xPy, \neg yPx \in X : x \succ y\}$ establishes a bijection between the set of all fully rational preference relations and the set of all consistent and complete judgment sets. More generally, any consistent judgment set $A \subseteq X$ represents an acyclic preference relation $\succ$ on K given by $x \succ y$ if and only if $xPy \in A$ or $\neg yPx \in A$ (for any $x, y \in K$).

What does accepting some binary ranking proposition xPy mean? On a belief interpretation, it means to believe that x is preferable to y; thus judgments on the preference agenda are beliefs on propositions of the form 'x is preferable to y'. On a desire interpretation, to accept xPy means to desire that, given a choice between x and y, x be chosen over y; here judgments on the preference agenda are desires on propositions of the form 'given a choice between x and y, x is chosen over y'.[13]

To represent Sen's original example in this way, let $N = \{1, 2\}$ be a two-member society consisting of Lewd and Prude, and let the set of alternatives be $K = \{l, p, n\}$, with the interpretation:

l: Lewd reads the book.
p: Prude reads the book.
n: No-one reads the book.[14]

Table 3 shows the two individuals' judgments on the ranking propositions lPn, nPp and pPl; the preference relations represented by these judgments are shown in brackets.

[12] For technical reasons, Z additionally contains, for each pair of distinct contants $x, y \in K$, $\neg\, x{=}y$ (exclusiveness).

[13] The two proposed interpretations – which correspond to *cognitivist* and *emotivist* interpretations of preferences – thus differ both in the meaning of the predicate P and in the meaning of 'accepting' a proposition. On a cognitivist interpretation, xPy means that x is preferable to/better than y, and the question is whether or not to *believe* such a proposition. On an emotivist interpretation, xPy means that x is chosen over y in a binary choice, and the question is whether or not to *desire* such a proposition. The two interpretations illustrate our broader point that judgment aggregation can be viewed either as the aggregation of belief sets or as that of desire sets.

[14] For convenience, we use the symbol n here, which elsewhere in the paper denotes the group size.

	lPn	nPp	pPl
Lewd ($p \succ l \succ n$)	True	False	True
Prude ($n \succ p \succ l$)	False	True	True

Table 3: Sen's example

Society assigns to Lewd the right to determine the collective judgment on lPn. On a belief interpretation, this means that Lewd is given an expert right on whether or not Lewd-reading-the-book is preferable to no-one-reading-the-book; on a desire interpretation, that he is given a liberal right on whether or not, in a choice between these two alternatives, Lewd-reading-the-book is chosen over no-one reading the book. Similarly, society assigns to Prude the right to determine the collective judgment on nPp, interpretable analogously. Given the individual judgments in table 3, respecting these rights means that society must accept both lPn and nPp; and since both individuals accept pPl, the Pareto principle requires the collective acceptance of pPl. But the resulting judgment set $\{lPn, nPp, pPl\}$ is inconsistent: it represents a cyclical preference relation. More generally, we can apply theorem 1 to the preference agenda.

Lemma 1 *The preference agenda is connected.*

This lemma has a straightforward proof (given in the appendix); for instance, propositions xPy and $x'Py'$ for pairwise distinct alternatives $x, y, x', y' \in K$ are conditionally dependent, as is seen by conditionalizing on $Y = \{yPx', y'Px\}$.

Corollary 1 *(Sen 1970) For the preference agenda, there exists no aggregation function (generating consistent collective judgment sets) that satisfies universal domain, minimal rights and the unanimity principle.*

Note that an aggregation function for the preference agenda with universal domain and generating consistent collective judgment sets represents a preference aggregation function that maps any possible profile of fully rational preference relations to an acyclic one, and the conditions of minimal rights and the unanimity principle correspond to Sen's conditions of minimal liberalism and the Pareto principle.

5 Possibility results

We now consider conditions under which the conflict between (expert or liberal) rights and the unanimity principle does not arise. For simplicity, we focus on individual rights, but our results can be generalized to subgroup rights too. To state our possibility results, we first refine our account of rights. The condition of minimal rights above does not specify which individuals have rights on which propositions. We now make the assignment of rights more 'targeted' by introducing explicit rights systems.

A *rights system* is an n-tuple $(R_1, ..., R_n)$, where each R_i is a (possibly empty) subset of X containing pairs $p, \neg p$. For each i, we call R_i individual i's *rights set*. On a belief interpretation, the elements of R_i are the propositions on which individual i is the expert; on a desire interpretation, the propositions that belong to i's private sphere. An aggregation function respects a rights system if it satisfies the following condition.

Rights. Every individual i is decisive on the rights set R_i.

It is easy to see that this condition can be met by a well-behaved aggregation function only if the rights system is consistent in a minimal sense. Call a rights system $(R_1, ..., R_n)$ *consistent* if $B_1 \cup ... \cup B_n$ is consistent for any consistent subsets $B_1, ..., B_n$ of $R_1, ..., R_n$, respectively.

Proposition 1 *If and only if the rights system is consistent, there exists an aggregation function F (generating consistent collective judgment sets) that satisfies universal domain and rights.*

But even for a consistent rights system, theorem 1 immediately implies that, if the agenda is connected and two or more distinct R_i's each contain at least one proposition-negation pair, respecting rights is inconsistent with universal domain and the unanimity principle in an aggregation function generating consistent collective judgment sets. We now show that the inconsistency can be avoided if individual judgments fall into a suitably restricted domain or the rights system (together with the agenda) has a particular property.[15]

5.1 Special domains: deferring/empathetic judgments

Let a rights system be given. When one individual adopts the judgments of another whenever those judgments concern propositions in the other's rights set, we say that the first individual *defers* to the judgments of the second (if the rights in question are expert rights) or is *empathetic* towards them (if the rights are liberal rights). Formally, individual i is *deferring/empathetic* in profile $(A_1, ..., A_n)$ if $A_i \cap R_j = A_j \cap R_j$ for all $j \neq i$, and a profile $(A_1, ..., A_n)$ is *deferring/empathetic* if every individual is deferring/empathetic in it. Deferring/empathetic profiles exhibit unanimous agreement on every proposition in some individual's rights set, a strong restriction. Our possibility theorem, however, is based on a less demanding restriction. A profile $(A_1, ..., A_n)$ is *minimally deferring/empathetic* if some individual is deferring/empathetic in it.

Minimally Deferring/Empathetic Domain. The domain of F is the set of all minimally deferring/empathetic profiles of consistent and complete individual judgment sets.

[15] For an overview of domain restrictions in response to the original liberal paradox in preference aggregation, including preference-based definitions of 'empathy' and 'tolerance', see Sen (1983); see also Craven (1982), Gigliotti (1986).

If more than one individual i has a non-empty rights set R_i, the minimally deferring/empathetic domain is a proper subset of the universal domain.[16]

Theorem 3 *For any agenda and any rights system, there exists an aggregation function (generating consistent collective judgment sets) that satisfies minimally deferring/empathetic domain, rights and the unanimity principle.*

Surprisingly, the result does not require a consistent rights system $(R_1, ..., R_n)$. But if $(R_1, ..., R_n)$ is inconsistent, how could a single deferring/empathetic individual prevent the other individuals from exercising their rights in an inconsistent way, leading to an inconsistent collective judgment set by respecting rights? The answer is that individual i's deferral/empathy *does* prevent such inconsistencies, albeit in a technical sense. Inconsistencies in the exercise of the others' rights would (by the definition of deferral/empathy) lead individual i to have an inconsistent judgment set A_i, something excluded by the minimally deferring/empathetic domain. Our definition of this domain thus restricts individuals $j \neq i$ in their exercise of rights so as to allow individual i to be both deferring/empathetic and consistent. To avoid this feature of the definition, one could redefine a deferring/empathetic individual as one who adopts the others' judgments (where they have rights) unless these judgments are mutually inconsistent; formally, one may define individual i to be *deferring/empathetic* in profile $(A_1, ..., A_n)$ if $[A_i \cap R_j = A_j \cap R_j$ for all $j \neq i]$ whenever $\bigcup_{j \neq i} [A_j \cap R_j]$ is consistent. Under this modified definition, theorem 3 continues to hold provided the rights system $(R_1, ..., R_n)$ is consistent.

5.2 Special domains: agnostic/tolerant judgments

When one individual makes no judgment on propositions in another's rights set, we say that the first individual is *agnostic* about the judgments of the second (if the rights in question are expert rights) or *tolerant* towards them (if the rights are liberal rights). We define agnosticism/tolerance as the requirement that an individual's judgment set be consistent with any possible consistent exercise of rights by others. Formally, individual i with judgment set A_i is *agnostic/tolerant* if A_i is consistent with every consistent set of the form $B_1 \cup ... \cup B_{i-1} \cup B_{i+1} \cup ... \cup B_n$, where, for each individual $j \neq i$, $B_j \subseteq R_j$. A profile $(A_1, ..., A_n)$ is *agnostic/tolerant* if every individual is agnostic/tolerant in it. A profile $(A_1, ..., A_n)$ is *minimally agnostic/tolerant* if some individual is agnostic/tolerant in it. Our possibility theorem requires only minimally agnostic/tolerant profiles.

[16] If there exists only one individual i with $R_i \neq \emptyset$, then i is trivially deferring/empathetic in every profile. If there exists no individual i with $R_i \neq \emptyset$, then every individual is trivially deferring/empathetic in every profile. So, if $R_i \neq \emptyset$ for at most one individual i, then the minimally deferring/empathetic domain coincides with the universal domain.

Minimally Agnostic/Tolerant Domain. The domain of F is the set of all minimally agnostic/tolerant profiles of consistent individual judgment sets.

The minimally agnostic/tolerant domain does not require complete judgment sets, and hence is not a subset of the universal domain. In fact, an agnostic/tolerant individual cannot have a complete judgment set (unless all other individuals have an empty rights set), since agnosticism/tolerance forces an individual to make no judgments on propositions in other individuals' rights sets. If *at least two* individuals have a non-empty rights set, then the universal domain neither contains, nor is contained by, the minimally agnostic/tolerant domain.[17]

Theorem 4 *For any agenda and any consistent rights system, there exists an aggregation function (generating consistent collective judgment sets) that satisfies minimally agnostic/tolerant domain, rights and the unanimity principle.*

Unlike our result on the minimally deferring/empathetic domain, the present result explicitly requires a consistent rights system. Also, in this theorem (unlike in all others) it is essential that we allow incomplete collective judgment sets: respecting rights forces the collective to take over any incompleteness of any individual's judgments within his or her rights set. If we wish to ensure complete collective judgment sets in theorem 4 we may *either* weaken people's rights by making each individual i merely *positively* decisive on R_i *or* restrict the domain by allowing only those minimally agnostic/tolerant profiles $(A_1, ..., A_n)$ in which each A_i is complete within R_i (i.e., each A_i contains a member of every proposition-negation pair in R_i). In such a restricted domain, each individual may refrain from making judgments only outside his or her rights set.

5.3 Special agendas and rights systems

Instead of restricting the domain, we now consider special rights systems, namely ones we call *disconnected*. We have seen in proposition 1 that consistency of a rights system is sufficient for the existence of aggregation functions satisfying universal domain and rights, yet the unanimity principle may be violated. We now strengthen the consistency requirement on the rights system so as to make it sufficient for the existence of aggregation functions satisfying universal domain, rights and the unanimity principle.

For a finite agenda or compact logic,[18] our definition of a disconnected rights system can be stated as follows (in the appendix we give a more general statement). The rights system $(R_1, ..., R_n)$ is *disconnected* (in X) if no proposition

[17] Again, if $R_i \neq \emptyset$ for only one individual i, then i is trivially agnostic/tolerant in every profile; and if $R_i \neq \emptyset$ for no individual i, then every individual is trivially agnostic/tolerant in every profile. So, if $R_i \neq \emptyset$ for at most one individual i, then the minimally agnostic/tolerant domain contains the universal domain.

[18] A logic is *compact* if every inconsistent set of propositions has a finite inconsistent subset.

in any R_i is conditionally dependent of any proposition in any R_j ($j \neq i$). Informally, a disconnected rights system is one in which the rights of different individuals are not 'entangled' with each other conditional on other propositions in the agenda. Note that a disconnected rights system in which more than one individual has a non-empty rights set can exist only if the agenda is *not* connected. The following theorem holds.

Theorem 5 *If (and only if) the rights system is disconnected, there exists an aggregation function (generating consistent collective judgment sets) that satisfies universal domain, rights and the unanimity principle.*

However, while the domain is not restricted – there need not be any deferring/empathetic or agnostic/tolerant individuals – disconnectedness is a severe constraint on a rights system and satisfiable (if more than one individual is to have a non-empty rights set) only for special agendas.

6 Discussion

We have identified a liberal paradox for judgment aggregation. If the agenda of propositions under consideration is connected, then, under universal domain, the assignment of (expert or liberal) rights to two or more individuals or subgroups is inconsistent with the unanimity principle. The inconsistency arises because propositions on which unanimous judgments are reached are sometimes logically constrained by other propositions that lie in some individual's or subgroup's sphere of rights. The inconsistency does not arise for the restricted domains of deferring/empathetic judgments or agnostic/tolerant judgments or for a disconnected rights system – which requires an agenda that is not connected, if more than one individual or subgroup is to have rights. For example, if different individuals (or subgroups) each live on their own Robinson Crusoe island, where the propositions relevant to different islands are not conditionally dependent on each other, then rights can be assigned to them without violating the unanimity principle. But such scenarios are rare; almost all realistic collective decision problems presuppose some interaction between different agents, which makes it plausible to expect connections between different individuals' rights sets.

Our results have implications for the design of mechanisms that groups (societies, legislatures, committees, expert panels, management boards, organizations) can use for making decisions on multiple interconnected propositions. For some groups or decision problems, the existence of agnostic/tolerant or deferring/empathetic group members may avoid the paradox. But there is no guarantee that such attitudes will exist, and constitutional provisions may be needed to deal with the possible occurrence of the paradox. Ultimately, the group faces the constitutional choice between either relaxing the (democratic) unanimity principle or relaxing (expert or liberal) rights of individuals or subgroups. Let us briefly discuss each option.

If it is deemed unacceptable to weaken any rights, violations of the unanimity principle will have to be allowed in collective decision making – an option advocated, among others, by Sen (1976) in the context of preference aggregation. The overruling of unanimous judgments may be defended on the grounds of unacceptable individual *motivations* behind such judgments, which disregard the rights of other individuals. Individual judgments driven by such unacceptable motivations may be seen as the counterpart in judgment aggregation of the so-called *meddlesome* preferences in preference aggregation (Blau 1975).

On the other hand, if the unanimity principle is deemed indispensible, then some weakening of rights is necessary. One possibility is to assign such rights in a suitably disconnected way, so that different rights never conflict with each other or with unanimous judgments on other propositions. Alternatively, rights can be made *alienable*, i.e., conditional on not conflicting with other rights or unanimous judgments. Dowding and van Hees (2003) have suggested that rights may sometimes be overruled by other considerations; in particular, different rights may carry a different threshold of being respected, which may vary from right to right and from context to context.

The choice of whether or not to give rights priority over the unanimity principle also depends on whether these rights are expert rights or liberal rights. In the case of liberal rights, the choice is ultimately a normative one, which depends on how much weight we give to individual liberty as a value relative to other values such as certain democratic decision principles. In the case of expert rights, by contrast, the choice is not just normative. If the propositions are factually either true or false, then it becomes an epistemological question which aggregation function is better at tracking their truth-values: one that respects expert rights or one that satisfies the unanimity principle. The answer to this question – which we cannot provide here – depends on several factors, such as how competent the experts and non-experts are on the various propositions and whether different individuals' judgments are mutually dependent or independent. The literature on the Condorcet jury theorem can be modified to address this question (Bovens and Rabinowicz 2006, List 2004b).

As the liberal paradox continues to be discussed in social choice theory and game theory, we hope that our findings will help to extend this discussion to the emerging theory of judgment aggregation and inspire further work.

7 References

Blau JH (1975) Liberal values and independence. *Review of Economic Studies* 42: 395-402

Bovens L, Rabinowicz W (2006) Democratic Answers to Complex Questions – An Epistemic Perspective. *Synthese* 150: 131-153

Bradley R (1999) Conditional Desirability. *Theory and Decision* 47: 23-55

Craven J (1982) Liberalism and individual preferences. *Theory and Decision* 14: 351-360

Deb R, Pattanaik PK, Razzolini L (1997) Game forms, rights, and the efficiency of social outcomes. *Journal of Economic Theory* 72: 74-95

Dietrich F (2006) Judgment Aggregation: (Im)Possibility Theorems. *Journal of Economic Theory* 126(1): 286-298

Dietrich F (2007a) A generalised model of judgment aggregation. *Social Choice and Welfare* 28(4): 529-565

Dietrich F (2007b) Aggregation theory and the relevance of some issues to others. Working paper, London School of Economics

Dietrich F (forthcoming) The possibility of judgment aggregation on agendas with subjunctive implications. *Journal of Economic Theory*

Dietrich F, List C (2007) Arrow's theorem in judgment aggregation. *Social Choice and Welfare* 29(1): 19-33

Dokow E, Holzman R (2005) Aggregation of binary evaluations. Working paper, Technion Israel Institute of Technology

Dowding K, van Hees M (2003) The construction of rights. *American Political Science Review* 97: 281-293

Gigliotti GA (1986) Comment on Craven. *Theory and Decision* 21: 89-95

van Hees M (1999) Liberalism, efficiency, and stability: some possibility results. *Journal of Economic Theory* 88: 294-309

van Hees M (2004) Freedom of choice and diversity of options: some difficulties. *Social Choice and Welfare* 22: 253-266

van Hees, M. (2007) The limits of epistemic democracy. *Social Choice and Welfare* 28(4): 649-666

Hintikka J (1971) Some Main Problems of Deontic Logic. In Hilpinen R (1971) *Deontic Logic: Introductory and Systematic Readings*, Dordrecht (D. Reidel): 59-104

Konieczny S, Pino-Perez R (2002) Merging information under constraints: a logical framework. *Journal of Logic and Computation* 12: 773-808

Kornhauser LA, Sager LG (1986) Unpacking the Court. *Yale Law Journal* 96(1): 82-117

List C (2003) A Possibility Theorem on Aggregation over Multiple Interconnected Propositions. *Mathematical Social Sciences* 45(1): 1-13 (Corrigendum in *Mathematical Social Sciences* 52:109-110)

List C (2004a) A Model of Path Dependence in Decisions over Multiple Propositions. *American Political Science Review* 98(3): 495-513

List C (2004b) The Probability of Inconsistencies in Complex Collective Decisions. *Social Choice and Welfare* 24(1):3-32

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

List C, Pettit P (2004) Aggregating Sets of Judgments: Two Impossibility Results Compared. *Synthese* 140(1-2): 207-235

Nehring K, Puppe C (2002) Strategy-Proof Social Choice on Single-Peaked Domains: Possibility, Impossibility and the Space Between. Working paper, University of California at Davies

Nehring K, Puppe C (2006) Consistent Judgement Aggregation: The Truth-Functional Case. Working paper, University of Karlsruhe

Pauly M, van Hees M (2006) Logical Constraints on Judgment Aggregation. *Journal of Philosophical Logic* 35: 569-585

Pettit P (2001) Deliberative Democracy and the Discursive Dilemma. *Philosophical Issues* 11: 268-299

Pigozzi G (2006) Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese* 152(2): 285-298

Rubinstein A, Fishburn P (1986) Algebraic Aggregation Theory. *Journal of Economic Theory* 38: 63-77

Sen AK (1970) The Impossibility of a Paretian Liberal. *Journal of Political Economy* 78: 152-157

Sen AK (1976) Liberty, Unanimity and Rights. *Economica* 43: 217-245

Sen AK (1983) Liberty and Social Choice. *Journal of Philosophy* 80: 5-28

Wagner Decew J (1981) Conditional Obligation and Counterfactuals. *Journal of Philosophical Logic* 10(1): 55-72

Wilson R (1975) On the Theory of Aggregation. *Journal of Economic Theory* 10: 89-99

A Appendix: proofs

We write $Domain(F)$ for the domain of F. As mentioned earlier, theorems 1, 2, 3 and 5 and proposition 1 continue to hold if completeness of collective judgment sets is also required. To turn our proofs of these results into proofs of the results with the added completeness condition, one must modify the constructed aggregation function F in each proof (specifically, in one direction of the implication) by replacing every consistent output $F(A_1, ..., A_n)$ by a consistent *and complete* superset of it.

Proof of theorem 1. 1. First assume the agenda X is connected. Suppose the aggregation function F satisfies minimal rights, the unanimity principle and universal domain. We show that F generates an inconsistent collective judgment set on some profile. By minimal rights, some individual i is decisive on some $\{p, \neg p\} \subseteq X$, and some other individual j is decisive on some $\{q, \neg q\} \subseteq X$. As X is connected, there exist propositions $p^* \in \{p, \neg p\}$ and $q^* \in \{q, \neg q\}$ and a set $Y \subseteq X$ inconsistent with the pair p^*, q^* but consistent with p^* and with q^*. As the sets $\{p^*\} \cup Y$ and $\{q^*\} \cup Y$ are each consistent, they can each be extended to a consistent and complete judgment set. Consider a profile $(A_1, ..., A_n)$ of complete and consistent judgment sets such that A_i extends $\{p^*\} \cup Y$, A_j extends $\{q^*\} \cup Y$, and each A_k, $k \neq i, j$, extends either $\{p^*\} \cup Y$ or $\{q^*\} \cup Y$. By universal domain, $(A_1, ..., A_n) \in Domain(F)$. $F(A_1, ..., A_n)$ contains p^* by i's decisiveness on $\{p, \neg p\}$, contains q^* by j's decisiveness on $\{q, \neg q\}$, and contains all $y \in Y$ by the unanimity principle. So $\{p^*, q^*\} \cup Y \subseteq F(A_1, ..., A_n)$. Hence $F(A_1, ..., A_n)$ is inconsistent.

2. Now assume X is not connected. Then there are propositions $p, q \in X$ that are not conditionally dependent. Let F be the aggregation function with universal domain given by

$$F(A_1, ..., A_n) := (A_1 \cap \{p, \neg p\}) \cup (A_2 \cap \{q, \neg q\}) \cup (A_1 \cap ... \cap A_n)$$

for all $(A_1, ..., A_n) \in Domain(F)$. We show that F satisfies all requirements.

First, F satisfies the unanimity principle because, for all $(A_1, ..., A_n) \in Domain(F)$, $A_1 \cap ... \cap A_n \subseteq F(A_1, ..., A_n)$.

To show that F satisfies minimal rights, we show that individuals 1 and 2 are decisive, respectively, on $\{p, \neg p\}$ and $\{q, \neg q\}$. For all $(A_1, ..., A_n) \in Domain(F)$, we have

$$F(A_1, ..., A_n) \cap \{p, \neg p\} = A_1 \cap \{p, \neg p\}$$

because $\{p, \neg p\} \cap \{q, \neg q\} = \emptyset$ (otherwise p and q would be conditionally dependent, in fact dependent conditionally on $\emptyset$). So individual 1 is decisive on $\{p, \neg p\}$. For analogous reasons, individual 2 is decisive on $\{q, \neg q\}$.

Finally, we consider any profile $(A_1, ..., A_n) \in Domain(F)$ and show that $F(A_1, ..., A_n)$ is consistent. Note that $F(A_1, ..., A_n) = \{p^*, q^*\} \cup Y$, where p^* is the member of $A_1 \cap \{p, \neg p\}$, q^* the member of $A_2 \cap \{q, \neg q\}$, and Y the set $A_1 \cap ... \cap A_n$. By $\{p^*\} \cup Y \subseteq A_1$, $\{p^*\} \cup Y$ is consistent. By $\{q^*\} \cup Y \subseteq A_2$, $\{q^*\} \cup Y$ is consistent. So, as p and q are not conditionally dependent, $\{p^*, q^*\} \cup Y$ is consistent, i.e. $F(A_1, ..., A_n)$ is consistent. ■

Proof of theorem 2. If the agenda X is not connected then there exists an aggregation function with the relevant properties, by Theorem 1 and since minimal rights implies minimal subgroup rights (take singleton subgroups). The converse implication follows by straightforwardly adapting part 1 of the proof of Theorem 1. ■

Proof of lemma 1. Consider any two proposition p and q in the preference agenda $X = \{xPy, \neg xPy : x, y \in K, x \neq y\}$. Without loss of generality, we may assume that p and q are of the non-negated form xPy, because any negated proposition $\neg xPy \in X$ is logically equivalent to the non-negated proposition yPx. So let p be xPy, and y be $x'Py'$. To show that xPy and $x'Py'$ are conditionally dependent, we have to choose propositions $p^* \in \{xPy, \neg xPy\}$ and $q^* \in \{x'Py', \neg x'Py'\}$ and a set $Y \subseteq X$ such that $\{p^*\} \cup Y$ and $\{q^*\} \cup Y$ are consistent, and $\{p^*, q^*\} \cup Y$ is inconsistent (in fact, represents a cycle). The choices of p^*, q^*, Y depend on whether $x \in \{x', y'\}$ and whether $y \in \{x', y'\}$.

Case $x \neq x', y' \& y \neq x', y'$: $p^* = xPy$, $q^* = x'Py'$, $Y = \{yPx', y'Px\}$.
Case $y = y' \& x \neq x', y'$: $p^* = xPy$, $q^* = \neg x'Py$ $(\equiv yPx')$, $Y = \{x'Px\}$.
Case $y = x' \& x \neq x', y'$: $p^* = xPy$, $q^* = yPy'$, $Y = \{y'Px\}$.
Case $x = x' \& y \neq y', x'$: $p^* = \neg xPy$ $(\equiv yPx)$, $q^* = xPy'$ $Y = \{y'Py\}$.
Case $x = y' \& y \neq x', y'$: $p^* = xPy$, $q^* = x'Px$, $Y = \{yPx'\}$.
Case $x = x' \& y = y'$: $p^* = xPy$, $q^* = \neg xPy$ $(\equiv yPx)$, $Y = \emptyset$.
Case $x = y' \& y = x'$: $p^* = xPy$, $q^* = yPx$, $Y = \emptyset$. ■

Proof of proposition 1. (i) First, assume the rights system $(R_1, ..., R_n)$ is consistent. Let F be the aggregation function with universal domain defined by

$$F(A_1, ..., A_n) = (A_1 \cap R_1) \cup ... \cup (A_n \cap R_n)$$

for any profile $(A_1, ..., A_n) \in Domain(F)$. Obviously, F satisfies rights. To show collective consistency, note that, for any consistent sets $A_1, ..., A_n \subseteq X$, also $A_1 \cap R_1, ..., A_n \cap R_n$ are consistent, hence have a consistent union as the rights system is consistent.

(ii) Now assume the aggregation function F has all properties. To show that the rights system $(R_1, ..., R_n)$ is consistent, let $B_1, ..., B_n$ be consistent subsets of, respectively, $R_1, ..., R_n$. As each B_i is consistent, it may be extended to a consistent and complete judgment set A_i. The so-defined profile $(A_1, ..., A_n)$ belongs to the (universal) domain of F. By rights, $B_i \cap F(A_1, ..., A_n) = B_i$ for all individuals i, and so

$$\begin{aligned} B_1 \cup ... \cup B_n &= [B_1 \cap F(A_1, ..., A_n)] \cup ... \cup [B_n \cap F(A_1, ..., A_n)] \\ &= [B_1 \cup ... \cup B_n] \cap F(A_1, ..., A_n). \end{aligned}$$

So $B_1 \cup ... \cup B_n$ is a subset of the consistent set $F(A_1, ..., A_n)$, hence is itself consistent. ■

Proof of theorem 3. For each minimally deferring/empathetic profile $(A_1, ..., A_n)$, define $F(A_1, ..., A_n)$ as the judgment set A_i of some deferring/empathetic individual i (if there are several such individuals, choose any one of them). The so-defined aggregation function satisfies all conditions, because the collective judgment set, by being the judgment set of a deferring/empathetic individual, is consistent, matches the judgments of any individual within this individual's rights set (so that F satisfies rights), and contains each proposition that every individual accepts (so that F satisfies the unanimity principle). ■

Proof of theorem 4. Suppose the rights system $(R_1, ..., R_n)$ is consistent. For every minimally agnostic/tolerant profile $(A_1, ..., A_n)$, since each A_i is consistent, so is each $A_i \cap R_i$. Hence, by the consistency of the rights system, the union $\cup_i (A_i \cap R_i)$ is consistent. So, as $(A_1, ..., A_n)$ is minimally agnostic/tolerant, there exists an (agnostic/tolerant) individual j such that A_j is consistent with $\cup_{i \neq j} (A_i \cap R_i)$, i.e. such that the set

$$A_j \cup [\cup_{i \neq j} (A_i \cap R_i)]$$

is consistent. Let $F(A_1, ..., A_n)$ be this set. To show that the so-definied aggregation function F satisfies all properties, note first that F by construction satisfies minimally agnostic/tolerant domain, and consistent collective judgment sets. Also the unanimity principle holds: for all minimally agnostic/tolerant profiles $(A_1, ..., A_n)$, $F(A_1, ..., A_n)$ is by definition a superset of $A_1 \cap ... \cap A_n$.

To show rights, consider a minimally agnostic/tolerant profile $(A_1, ..., A_n)$. Then there is an agnostic/tolerant individual j such that

$$F(A_1, ..., A_n) = A_j \cup [\cup_{i \neq j}(A_i \cap R_i)].$$

Individual j's rights are respected since

$$F(A_1, ..., A_n) \cap R_j = A_j \cap R_j,$$

where we use the fact that the sets $R_1, ..., R_n$ are pairwise disjoint by the consistency of the rights system (and since we have excluded tautologies and contradictions). To see that the rights of any individual $k \neq j$ are also respected, note first that

$$F(A_1, ..., A_n) \cap R_k = (A_j \cap R_k) \cup (A_k \cap R_k),$$

again using that $R_1, ..., R_n$ are pairwise disjoint. But $A_j \cap R_k$ is empty: otherwise A_j would not be consistent with all consistent subsets of R_k, hence j would not be agnostic/tolerant. Hence

$$F(A_1, ..., A_n) \cap R_k = A_k \cap R_k,$$

as desired. ■

In the main text, we have stated the definition of a disconnected rights system in the case that X is finite or the logic is compact. The general definition is as follows. The rights system $(R_1, ..., R_n)$ is *disconnected* (in X) if there are no sets $B \subseteq R_i$ and $C \subseteq R_j$ with $i \neq j$ such that $B \cup C$ is inconsistent with some set $Y \subseteq X$ that is consistent with B and with C. This definition is closely related to the previous one: if we restrict the sets B and C to be singletons, we obtain the previous definition. We now prove the equivalence of the two definitions.

Lemma 2 *For a rights system $(R_1, ..., R_n)$,*

(a) *if X is finite or belongs to a compact logic, the two disconnectedness definitions are equivalent;*

(b) *in general, disconnectedness in the new sense implies disconnectedness in the old sense, and is equivalent to the following condition:*

- *the sets $R_1, ..., R_n$ are logically independent conditional on any set $B \subseteq X \backslash (R_1 \cup ... \cup R_n)$, i.e., for every set $B \subseteq X \backslash (R_1 \cup ... \cup R_n)$, $B_1 \cup ... \cup B_n$ is consistent with B whenever each $B_i \subseteq R_i$ is.*

Proof of lemma 2. We denote by D1 the condition defining disconnectedness in the main text, by D2 the condition defining disconnectedness in the appendix, and by D3 the condition stated in lemma 2.

We first prove part (b).

'D2⇒D1'. Assume D1 does not hold. We show that D2 does not hold. As D1 is violated, there are $p \in R_i$ and $q \in R_j$ $(i \neq j)$ that are conditionally dependent, that is: for some $p^* \in \{p, \neg p\}$, $q^* \in \{q, \neg q\}$ and $Y \subseteq X$, $\{p^*, q^*\} \cup Y$ is inconsistent but each of $\{p^*\} \cup Y$ and $\{q^*\} \cup Y$ is consistent. So D2 is violated: take $B := \{p^*\}$ and $C := \{q^*\}$.

'D2⇒D3'. Suppose D3 does not hold. We show that D2 does not hold. As D3 does not hold, there are sets $B_1 \subseteq R_1, ..., B_n \subseteq R_n, B \subseteq X \backslash (R_1 \cup ... \cup R_n)$ such that each $B_i \cup B$ is consistent but $(\cup_{i=1,...,n} B_i) \cup B$ is inconsistent. Among all sets of individuals $K \subseteq \{1, ..., n\}$ such that $(\cup_{k \in K} B_k) \cup B$ is inconsistent (there is at least one), let K be one of smallest size. We have $|K| \geq 2$, since otherwise some $B_k \cup B$ would be inconsistent. So there are distinct individuals $i, j \in K$. To find a counterexample to D2, let $C := B_i$, $D := B_j$ and $Y := (\cup_{k \in K \backslash \{i,j\}} B_k) \cup B$. The sets $Y \cup C = (\cup_{k \in K \backslash \{j\}} B_k) \cup B$ and $Y \cup D = (\cup_{k \in K \backslash \{i\}} B_k) \cup B$ are each consistent (by the minimality of K), but the set $Y \cup C \cup D = (\cup_{k \in K} B_k) \cup B$ is inconsistent, as desired.

'D3⇒D2'. Assume D3. Suppose for a contradiction that $B \subseteq R_i$, $C \subseteq R_j$ $(i \neq j)$, and $Y \subseteq X$, and that $B \cup C \cup Y$ is inconsistent but $B \cup Y$ and $C \cup Y$ are consistent. Put $Z := B \cup C \cup Y$. Then (*) Z is inconsistent, and (**) $Z \backslash B$ and $Z \backslash C$ are each consistent. By D3, the sets $R_1, ..., R_n$ are pairwise disjoint: otherwise they would be logically dependent conditional on $B = \emptyset$ (since some pair $p, \neg p$ would belong to two of the sets $R_1, ..., R_n$, so that we could choose consistent subsets of $R_1, ..., R_n$, respectively, whose union contains the pair $p, \neg p$, hence is inconsistent). So, among the sets $B_1 := Z \cap R_1, ..., B_n := Z \cap R_n$, all except B_i are disjoint with B, and all except B_j are disjoint with C. Hence each of $B_1, ..., B_n$ is a subset of $Z \backslash B$ or of $Z \backslash C$. So, as $D := Z \backslash (R_1 \cup ... \cup R_n)$ is a subset of $Z \backslash B$ *and* of $Z \backslash C$, each of the sets $B_1 \cup D, ..., B_n \cup D$ is a subset of $Z \backslash B$ or of $Z \backslash C$, hence is consistent by (**). But the union

$$\begin{aligned} B_1 \cup ... B_n \cup D &= [(Z \cap R_1) \cup ... \cup (Z \cap R_n)] \cup [Z \backslash (R_1 \cup ... \cup R_n)] \\ &= [Z \cap (R_1 \cup ... \cup R_n)] \cup [Z \backslash (R_1 \cup ... \cup R_n)] = Z \end{aligned}$$

is inconsistent by (*). This contradicts D3.

To prove part (a), it remains to show the following implication, assuming that X is finite or the logic compact.

'D1⇒D2'. Suppose for a contradiction that D1 holds but D2 does not. As D2 is violated, there are sets $B \subseteq R_i$ and $C \subseteq R_j$ with $i \neq j$ and $Y \subseteq X$ such that $B \cup C \cup Y$ is inconsistent but $B \cup Y$ and $C \cup Y$ are each consistent. As X is finite or the logic compact, $B \cup C \cup Y$ has a minimal inconsistent subset Z. By Z's inconsistency, Z is neither a subset of $C \cup Y$ nor of $B \cup Y$. So there is a $p \in B \cap Z$ and a $q \in C \cap Z$. Let $Z' := Z \backslash \{p, q\}$. By D1, p and q are not conditionally dependent, hence are distinct. So $\{p\} \cup Z'$ and $\{q\} \cup Z'$ are each proper subsets of Z, so are consistent; but $\{p, q\} \cup Z' = Z$ is inconsistent. Hence p and q are conditionally dependent, violating D1. ■

Proof of theorem 5. 1. First let the rights system $(R_1, ..., R_n)$ be disconnec-

ted. Define F as as the aggregation function with universal domain given, for all $(A_1, ..., A_n) \in Domain(F)$, by

$$F(A_1, ..., A_n) := B_1 \cup ... \cup B_n \cup B,$$

where

$$B_i := A_i \cap R_i,\ i = 1, ..., n,$$

and

$$B := (A_1 \cap ... \cap A_n) \backslash (R_1 \cup ... \cup R_n).$$

We now show that F satisfies all relevant properties.

First, each outcome $F(A_1, ..., A_n)$ is consistent: defining $B_1, ..., B_n, B$ as before, each $B_i \cup B$ is consistent (by being a subset of the consistent set A_i), whence the union $B_1 \cup ... \cup B_n \cup B$ $(= F(A_1, ..., A_n))$ is consistent by part (b) of lemma 2.

Second, F satisfies rights since $F(A_1, ..., A_n) \cap R_i = A_i \cap R_i$ for all individuals i and profiles $(A_1, ..., A_n) \subseteq Domain(F)$.

Finally, F satisfies the unanimity principle since, for all profiles $(A_1, ..., A_n) \in Domain(F)$, $F(A_1, ..., A_n)$ contains each member of $A_1 \cap ... \cap A_n$, whether it belongs to some R_i (hence to $A_i \cap R_i$) or to no R_i (hence to $(A_1 \cap ... \cap A_n) \backslash (R_1 \cup ... \cup R_n)$).

2. Conversely, assume that F is an aggregation function with all the required properties. To prove that the rights system is disconnected, it suffices by part (b) of lemma 2 to consider sets $B_1 \subseteq R_1, ..., B_n \subseteq R_n$ consistent with a set $B \subseteq X \backslash (R_1 \cup ... \cup R_n)$, and to show that $B_1 \cup ... \cup B_n \cup B$ is consistent. As each $B_i \cup B$ is consistent, it can be extended to a complete and consistent judgment set $A_i \subseteq X$. The collective judgment set $F(A_1, ..., A_n)$ contains all $p \in B_1 \cup ... \cup B_n$ (by rights) and all $p \in B$ (by the unanimity principle). So $B_1 \cup ... \cup B_n \cup B \subseteq F(A_1, ..., A_n)$. Hence, as $F(A_1, ..., A_n)$ is consistent, so is $B_1 \cup ... \cup B_n \cup B$, as desired. ■

Chapter 4

Voter manipulation

Paper: Strategy-proof judgment aggregation (with C. List), *Economics and Philosophy* 23: 269-300, 2007

Strategy-proof judgment aggregation

Franz Dietrich and Christian List[1]

Which rules for aggregating judgments on logically connected propositions are manipulable and which not? In this paper, we introduce a preference-free concept of non-manipulability and contrast it with a preference-theoretic concept of strategy-proofness. We characterize all non-manipulable and all strategy-proof judgment aggregation rules and prove an impossibility theorem similar to the Gibbard-Satterthwaite theorem. We also discuss weaker forms of non-manipulability and strategy-proofness. Comparing two frequently discussed aggregation rules, we show that "conclusion-based voting" is less vulnerable to manipulation than "premise-based voting", which is strategy-proof only for "reason-oriented" individuals. Surprisingly, for "outcome-oriented" individuals, the two rules are strategically equivalent, generating identical judgments in equilibrium. Our results introduce game-theoretic considerations into judgment aggregation and have implications for debates on deliberative democracy.

1 Introduction

How can a group of individuals aggregate their individual judgments (beliefs, opinions) on some logically connected propositions into collective judgments on these propositions? In particular, how can a group do this under conditions of pluralism, i.e., when individuals disagree on the propositions in question? This problem – *judgment aggregation* – is discussed in a growing literature in philosophy, economics and political science and generalizes earlier problems of social choice, notably preference aggregation in the Condorcet-Arrow tradition.[2] The problem arises in many different decision making bodies, ranging from legislative committees and multi-member courts to expert advisory panels and monetary policy committees of a central bank.

Judgment aggregation is often illustrated by a paradox: the *discursive* (or *doctrinal*) *paradox* (Kornhauser and Sager 1986; Pettit 2001; Brennan 2001). To illustrate, suppose a university committee responsible for a tenure decision has to make collective judgments on three propositions:[3]

a: The candidate is good at teaching.

b: The candidate is good at research.

c: The candidate deserves tenure.

According to the university's rules, c (the "conclusion") is true if and only if a and b (the "premises") are both true, formally $c \leftrightarrow (a \wedge b)$ (the "connection rule"). Suppose the committee has three members with judgments as shown in Table 1.

If the committee takes a majority vote on each proposition, then a and b are each accepted and yet c is rejected (each by two thirds), despite the (unanimous)

[1] F. Dietrich, Dept. of Quant. Econ., Univ. of Maastricht, P.O. Box 616, 6200 MD Maastricht, NL. C. List, Dept. of Govt., LSE, London WC2A 2AE, UK. This paper was presented at the University of Konstanz (6/2004), the Social Choice and Welfare Conference in Osaka (7/2004), the London School of Economics (10/2004), Université de Caen (11/2004), the University of East Anglia (1/2005), Northwestern University (5/2005), the 2005 SAET Conference in Vigo (6/2005), the University of Hamburg (10/2005), IHPST, Paris (1/2006). We thank the participants at these occasions, the anonymous referees of this paper and the editor, Bertil Tungodden, for comments.

[2] Preference aggregation becomes a case of judgment aggregation by expressing preference relations as sets of binary ranking propositions in predicate logic (List and Pettit 2004; Dietrich and List 2007a).

[3] This example is due to Bovens and Rabinowicz (2006).

	a	b	$c \leftrightarrow (a \wedge b)$	c
Individual 1	Yes	Yes	Yes	Yes
Individual 2	Yes	No	Yes	No
Individual 3	No	Yes	Yes	No
Majority	Yes	Yes	Yes	No

Table 1: The discursive paradox

acceptance of $c \leftrightarrow (a \wedge b)$. The discursive paradox shows that judgment aggregation by propositionwise majority voting may lead to inconsistent collective judgments, just as Condorcet's paradox shows that preference aggregation by pairwise majority voting may lead to intransitive collective preferences.

In response to the discursive paradox, two aggregation rules have been proposed to avoid such inconsistencies (e.g., Pettit 2001; Chapman 1998, 2002; Bovens and Rabinowicz 2006). Under *premise-based voting*, majority votes are taken on a and b (the premises), but not on c (the conclusion), and the collective judgment on c is derived using the connection rule $c \leftrightarrow (a \wedge b)$: in Table 1, a, b and c are all accepted. Premise-based voting captures the deliberative democratic idea that collective decisions on outcomes should be made on the basis of collectively decided reasons. Here reasoning is "collectivized", as Pettit (2001) describes it. Under *conclusion-based voting*, a majority vote is taken only on c, and no collective judgments are made on a or b: in Table 1, c is rejected and other propositions are left undecided. Conclusion-based voting captures the minimal liberal idea that collective decisions should be made only on (practical) outcomes and that the reasons behind such decisions should remain private. Here collective decisions are "incompletely theorized" in Sunstein's (1994) terms. (For a comparison between minimal liberal and comprehensive deliberative approaches to decision making, see List 2006.)

Abstracting from the discursive dilemma, List and Pettit (2002, 2004) have formalized judgment aggregation and proved that no judgment aggregation rule ensuring consistency can satisfy some conditions inspired by Arrow's conditions on preference aggregation. This impossibility result has been strengthened and extended by Pauly and van Hees (2006; see also van Hees 2007), Dietrich (2006), Gärdenfors (2006) and Dietrich and List (2007). Drawing on the model of "property spaces", Nehring and Puppe (2002, 2005) have offered the first characterizations of agendas of propositions for which impossibility results hold (for a subsequent contribution, see Dokow and Holzman 2005). Possibility results have been obtained by List (2003, 2004), Pigozzi (2006) and Osherson and Vardi (forthcoming). Dietrich (2007) has developed an extension of the judgment aggregation model to richer logical languages for expressing propositions, which we use in this paper. Related bodies of literature include those on abstract aggregation theory (Wilson 1975)[4] and on belief merging in computer science (Konieczny and Pino-Perez 2002).

But one important question has received little attention in the literature on judg-

[4]Wilson's (1975) aggregation problem, where a group has to form yes/no views on several issues based on individual views on them (subject to feasibility constraints), can be represented in judgment aggregation. Unlike judgment aggregation, Wilson's model cannot fully generally represent logical entailment: its primitive is a consistency (feasibility) notion, from which an entailment relation can be retrieved only for certain logical languages (Dietrich 2007).

ment aggregation: Which aggregation rules are manipulable by strategic voting and which are strategy-proof? The answer is not obvious, as strategy-proofness in the familiar sense in economics is a preference-theoretic concept and preferences are not primitives of judgment aggregation models. Yet the question matters for the design and implementation of an aggregation rule in a collective decision making body such as in the examples above. Ideally, we would like to find aggregation rules that lead individuals to reveal their judgments truthfully. Indeed, if an aggregation rule captures the normatively desirable functional relation between individual and collective judgments, then truthful revelation of these individual judgments (which are typically private information) is crucial for the (direct) implementation of that functional relation.[5]

In this paper, we address this question. We first introduce a simple condition of non-manipulability and characterize the class of non-manipulable judgment aggregation rules. We then show that, under certain motivational assumptions about individuals, our condition is equivalent to a game-theoretic strategy-proofness condition similar to the one introduced by Gibbard (1973) and Satterthwaite (1975) for preference aggregation.[6] Our characterization of non-manipulable aggregation rules then yields a characterization of strategy-proof aggregation rules. The relevant motivational assumptions hold if agents want the group to make collective judgments that match their own individual judgments (e.g., want the group to make judgments that match what they consider the truth). In many other cases, such as that of "reason-oriented" individuals (as defined in Section 5), non-manipulability and strategy-proofness may come significantly apart.

By introducing both a non-game-theoretic condition of non-manipulability and a game-theoretic condition of strategy-proofness, we are able to distinguish between *opportunities* for manipulation (which depend only on the aggregation rule in question) and *incentives* for manipulation (which depend also on the motivations of the decision-makers).

We prove that, for a general class of aggregation problems including the tenure example above, there exists no non-manipulable judgment aggregation rule satisfying universal domain and some other mild conditions, an impossibility result similar to the Gibbard-Satterthwaite theorem on preference aggregation. Subsequently, we identify various ways to avoid the impossibility result. We also show that our default conditions of non-manipulability and strategy-proofness fall into general families of conditions and discuss other conditions in these families. In the case of strategy-proofness, these conditions correspond to different motivational assumptions about the decision makers. In the tenure example, conclusion-based voting is strategy-proof in a strong sense, but produces no collective judgments on the premises. Premise-based voting satisfies only the weaker condition of strategy-proofness for "reason-oriented" individuals. Surprisingly, although premise- and conclusion-based voting are regarded in the literature as two diametrically opposed aggregation rules, they are strategically equivalent if individuals are "outcome-oriented", generating identical

[5] A functional relation between individual and collective judgments could be deemed normatively desirable for a variety of reasons, such as epistemic or democratic legitimacy goals. The axiomatic approach to social choice theory translates these goals into formal requirements on aggregation.

[6] Our definition of strategy-proofness in judgment aggregation draws on List (2002b, 2004), where sufficient conditions for strategy-proofness in (sequential) judgment aggregation are given.

judgments in equilibrium. Our results not only introduce game-theoretic considerations into the theory of judgment aggregation, but they are also relevant to debates on democratic theory as premise-based voting has been advocated, and conclusion-based voting rejected, by proponents of deliberative democracy (Pettit 2001).

There is, of course, a related literature on manipulability and strategy-proofness in preference aggregation, following Gibbard's and Satterthwaite's classic contributions (e.g., Taylor 2002, 2005; Saporiti and Thomé 2005). An important branch of this literature, from which several corollaries for judgment aggregation can be derived, has considered preference aggregation over options that are vectors of binary properties (Barberà et al. 1993, 1997; Nehring and Puppe 2002). A parallel to judgment aggregation can be drawn by identifying propositions with properties; a disanalogy lies in the structure of the informational input to the aggregation rule. While judgment aggregation rules collect a single judgment set from each individual (expressed in a possibly rich logical language), preference aggregation rules collect an entire preference ordering over vectors of properties. Whether or not an individual's most-preferred vector of properties (in preference aggregation) can be identified with her judgment set (in judgment aggregation) depends precisely on the motivational assumptions we make about this individual.

Another important related literature is that on the paradox of multiple elections (Brams et al. 1997, 1998; Kelly 1989). Here a group also aggregates individual votes on multiple propositions, and the winning combination can be one that no voter individually endorses. However, given the different focus of that work, the propositions in question are not explicitly modelled as logically interconnected as in our present model of judgment aggregation. The formal proofs of all the results reported in the main text are given in the Appendix.

2 The basic model

We consider a group of individuals $N = \{1, 2, \ldots, n\}$, where $n \geq 2$.[7] The group has to make collective judgments on logically connected propositions.

2.1 Representing propositions in formal logic

Propositions are represented in a *logical language*, defined by two components:

- a non-empty set $\mathbf{L}$ of formal expressions representing *propositions*; the language has a negation symbol $\neg$ ("not"), where for each proposition p in $\mathbf{L}$, its negation $\neg p$ is also contained in $\mathbf{L}$.
- an *entailment relation* $\vDash$, where, for each set of propositions $A \subseteq \mathbf{L}$ and each proposition $p \in \mathbf{L}$, $A \vDash p$ is read as "A logically entails p".[8]

We call a set of propositions $A \subseteq \mathbf{L}$ *inconsistent* if $A \vDash p$ and $A \vDash \neg p$ for some $p \in \mathbf{L}$, and *consistent* otherwise. We require the logical language to have certain

[7] Although no discursive paradox arises for $n = 2$, our results below still hold: Under Theorem 2's other conditions, non-manipulability requires a dictatorship of one of the two individuals. The unanimity rule, while also non-manipulable, violates completeness of collective judgments.

[8] $\vDash$ can be interpreted either as semantic entailment or as syntactic derivability (usually denoted $\vdash$). The two interpretations give rise to semantic or syntactic notions of rationality, respectively.

minimal properties (Dietrich 2007; Dietrich and List 2007a).[9]

The most familiar logical language is *(classical) propositional logic*, containing a given set of *atomic* propositions a, b, c, ..., such as the propositions about the candidate's teaching, research and tenure in the example above, and *compound* propositions with the logical connectives $\neg$ ("not"), $\wedge$ ("and"), $\vee$ ("or"), $\rightarrow$ ("if-then"), $\leftrightarrow$ ("if and only if"), such as the connection rule $c \leftrightarrow (a \wedge b)$ in the tenure example.[10] Examples of valid logical entailments in propositional logic are $\{a, a \rightarrow b\} \vDash b$ ("modus ponens"), $\{a \rightarrow b, \neg b\} \vDash \neg a$ ("modus tollens"), whereas the entailment $\{a \vee b\} \vDash a$ is not valid. Examples of consistent sets are $\{a, a \vee b\}$, $\{\neg a, \neg b, a \rightarrow b\}$, and examples of inconsistent ones are $\{a, \neg a\}$, $\{a, a \rightarrow b, \neg b\}$ and $\{a, b, c \leftrightarrow (a \wedge b), \neg c\}$.

We use classical propositional logic in our examples, but our results also hold for other, more expressive logical languages such as the following:

- *predicate logic*, which includes relation symbols and the quantifiers "there exists ..." and "for all ...";
- *modal logic*, which includes the operators "it's necessary that ..." and "it's possible that ...";
- *deontic logic*, which includes the operators "it's permissible that ..." and "it's obligatory that ...";
- *conditional logic*, which allows the expression of counterfactual or subjunctive conditionals.

Many different propositions that might be considered by a multi-member decision making body (ranging from legislative committees to expert panels) can be formally represented in an appropriate such language. Crucially, a logical language allows us to capture the fact that, in many decision problems, different propositions, such as the reasons for a particular tenure outcome and the resulting outcome itself, are mutually interconnected.

2.2 The agenda

The *agenda* is the set of propositions on which judgments are to be made; it is a non-empty subset $X \subseteq \mathbf{L}$, where X is a union of proposition-negation pairs $\{p, \neg p\}$ (with p not a negated proposition). For simplicity, we assume that double negations cancel each other out, i.e., $\neg\neg p$ stands for p.[11]

Two important examples are *conjunctive* and *disjunctive* agendas in propositional logic. A conjunctive agenda is $X = \{a_1, \ldots, a_k, c, c \leftrightarrow (a_1 \wedge \cdots \wedge a_k)\}^{+neg}$, where $a_1, \ldots, a_k$ are premises ($k \geq 1$), c is a conclusion, and $c \leftrightarrow (a_1 \wedge \cdots \wedge a_k)$ is the connection rule. We write Y^{+neg} as an abbreviation for $\{p, \neg p : p \in Y\}$. To define a disjunctive agenda, we replace $c \leftrightarrow (a_1 \wedge \cdots \wedge a_k)$ with $c \leftrightarrow (a_1 \vee \cdots \vee a_k)$. Conjunctive and disjunctive agendas arise in decision problems in which some outcome (c) is to be decided on the basis of some reasons $(a_1, \ldots, a_k)$. In the tenure example above,

[9] L1 (self-entailment): For all $p \in \mathbf{L}$, $\{p\} \vDash p$. L2 (monotonicity): For all $p \in \mathbf{L}$ and $A \subseteq B \subseteq \mathbf{L}$, if $A \vDash p$ then $B \vDash p$. L3 (completability): $\emptyset$ is consistent, and each consistent set $A \subseteq \mathbf{L}$ has a consistent superset $B \subseteq \mathbf{L}$ containing a member of each pair $p, \neg p \in \mathbf{L}$. L1-L3 are jointly equivalent to three conditions on the consistency notion: each pair $\{p, \neg p\} \subseteq \mathbf{L}$ is inconsistent; if $A \subseteq \mathbf{L}$ is inconsistent, so are its supersets $B \subseteq \mathbf{L}$; and L3 holds. See Dietrich (forthcoming) for details.

[10] $\mathbf{L}$ is the smallest set such that (i) a, b, c, ...$\in \mathbf{L}$ and (ii) if $p, q \in \mathbf{L}$ then $\neg p$,$(p \wedge q)$,$(p \vee q)$,$(p \rightarrow q)$,$(p \leftrightarrow q) \in \mathbf{L}$. We drop brackets when there is no ambiguity. Entailment ($\vDash$) is defined standardly.

[11] Hereafter, when we write $\neg p$ and p is already of the form $\neg q$, we mean q (rather than $\neg\neg q$).

we have a conjunctive agenda with $k = 2$.[12]

Other examples are agendas involving conditionals (in propositional or conditional logic) such as $X = \{a, b, a \rightarrow b\}^{+neg}$. Here proposition a might state some political goal, proposition $a \rightarrow b$ might state what the pursuit of a requires, and proposition b might state the consequence to be drawn. Alternatively, proposition a might be an empirical premise, $a \rightarrow b$ a causal hypothesis, and b the resulting prediction.

Finally, we can also represent standard preference aggregation problems within our model. Here we use a predicate logic with a set of constants K representing options ($|K| \geq 3$) and a two-place predicate R representing preferences, where, for any $x, y \in K$, the proposition xRy is interpreted as "x is preferable to y". Now the *preference agenda* is the set $X = \{xRy : x, y \in K\}^{+neg}$ (Dietrich and List 2007a).[13]

The nature of a judgment aggregation problem depends on what propositions are contained in the agenda and how they are interconnected. Our main characterization theorem holds for any agenda of propositions. Our main impossibility theorem holds for a large class of agendas, defined below. We also discuss applications to the important cases of conjunctive and disjunctive agendas.

2.3 Individual and collective judgments

Each individual i's *judgment set* is a subset $A_i \subseteq X$, where $p \in A_i$ means that individual i accepts proposition p. As the agenda typically contains both atomic propositions and compound ones, our definition of a judgment set captures the fact that an individual makes judgments both on free-standing atomic propositions and on their interconnections; and different individuals may disagree with each other on both kinds of propositions.

A judgment set A_i is *consistent* if it is a consistent set of propositions as defined for the logic; A_i is *complete* if it contains a member of each proposition-negation pair $p, \neg p \in X$. A *profile (of individual judgment sets)* is an n-tuple $(A_1, \ldots, A_n)$.

A *(judgment) aggregation rule* is a function F that assigns to each admissible profile $(A_1, \ldots, A_n)$ a collective judgment set $F(A_1, \ldots, A_n) = A \subseteq X$, where $p \in A$ means that the group accepts proposition p. The set of admissible profiles is called the *domain* of F, denoted $Domain(F)$. Several results below require the following.

Universal Domain. $Domain(F)$ is the set of all possible profiles of consistent and complete individual judgment sets.

2.4 Examples of aggregation rules

We give four important examples of aggregation rules satisfying universal domain, as just introduced. The first two rules are defined for any agenda, the last two only

[12] Although we here interpret connection rules $c \leftrightarrow (a_1 \wedge \cdots \wedge a_k)$ as *material* biimplications, one may prefer to interpret them as *subjunctive* biimplications (in a conditional logic). This changes the logical relations within conjunctive agendas: more judgment sets are consistent, including $\{\neg a_1, ..., \neg a_k, \neg c, \neg(c \leftrightarrow (a_1 \wedge \cdots \wedge a_k))\}$. As a result, our impossibility results (Theorems 2-3 and Corollary 2) do not apply to conjunctive agendas in the revised sense; instead, we obtain stronger possibility results. Analogous remarks hold for disjunctive agendas. See Dietrich (forthcoming).

[13] The entailment relation $\vDash$ in this logical language is defined by $A \vDash p$ if and only if $A \cup Z$ entails p in the standard sense of predicate logic, where Z is the set of rationality conditions on preferences $\{(\forall v) vRv$, $(\forall v_1)(\forall v_2)(\forall v_3)((v_1 Rv_2 \wedge v_2 Rv_3) \rightarrow v_1 Rv_3)$, $(\forall v_1)(\forall v_2)(\neg\ v_1{=}v_2 \rightarrow (v_1 Rv_2 \vee v_2 Rv_1))\}$.

for conjunctive (or disjunctive) agendas (the present definitions are simplified, but a generalization is possible).

Propositionwise majority voting. For each $(A_1, \ldots, A_n) \in Domain(F)$, $F(A_1, \ldots, A_n)$ is the set of all propositions $p \in X$ such that more individuals i have $p \in A_i$ than $p \notin A_i$.

Dictatorship of individual i. For each $(A_1,...,A_n) \in Domain(F)$, $F(A_1,...,A_n)=A_i$.

Premise-based voting. For each $(A_1, \ldots, A_n) \in Domain(F)$, $F(A_1, \ldots, A_n)$ is the set containing

- any premise a_j if and only if more i have $a_j \in A_i$ than $a_j \notin A_i$,
- the connection rule $c \leftrightarrow (a_1 \wedge \cdots \wedge a_k)$,
- the conclusion c if and only if $a_j \in F(A_1, \ldots, A_n)$ for all premises a_j,
- any negated proposition $\neg p$ if and only if $p \notin F(A_1, \ldots, A_n)$.[14]

Here votes are taken only on each premise, and the conclusion is decided by using an exogenously given connection rule.

Conclusion-based voting. For each $(A_1,...,A_n) \in Domain(F)$, $F(A_1,...,A_n)$ is the set containing

- only the conclusion c if more i have $c \in A_i$ than $c \notin A_i$,
- only the negation of the conclusion $\neg c$ otherwise.

Here a vote is taken only on the conclusion, and no collective judgments are made on other propositions.

Dictatorships and premise-based voting always generate consistent and complete collective judgments; propositionwise majority voting sometimes generates inconsistent ones (recall Table 1), and conclusion-based voting always generates incomplete ones (no judgments on the premises).

In debates on the discursive paradox and democratic theory, several arguments have been offered for the superiority of premise-based voting over conclusion-based voting. One such argument draws on a deliberative conception of democracy, which emphasizes that collective decisions on conclusions should follow from collectively decided premises (Pettit 2001; Chapman 2002). A second argument draws on the Condorcet jury theorem. If all the propositions are factually true or false and each individual has a probability greater than 1/2 of judging each premise correctly, then, under certain probabilistic independence assumptions, premise-based voting has a higher probability of producing a correct collective judgment on the conclusion than conclusion-based voting (Grofman 1985; Bovens and Rabinowicz 2006; List 2005, 2006). Here we show that, with regard to strategic manipulability, premise-based voting performs worse than conclusion-based voting.

3 Non-manipulability

When can an aggregation rule be manipulated by strategic voting? We first introduce a new condition of non-manipulability, not yet game-theoretic. Below we prove that,

[14] For a disjunctive agenda, replace "$c \leftrightarrow (a_1 \wedge \cdots \wedge a_k)$" with "$c \leftrightarrow (a_1 \vee \cdots \vee a_k)$" and "for all premises a_j" with "for some premise a_j".

under certain motivational assumptions about the individuals, our non-manipulability condition is equivalent to a game-theoretic strategy-proofness condition. We also notice that non-manipulability and strategy-proofness may sometimes come apart.

3.1 An example

To give a simple example, we use the language of *incentives* to manipulate, although our subsequent formal analysis focuses on underlying *opportunities* for manipulation; we return to incentives formally in Section 4. Recall the profile in Table 1. Suppose, for the moment, that the three committee members each care only about reaching a collective judgment on the conclusion (c) that agrees with their own individual judgments on the conclusion, and that they do not care about the collective judgments on the premises. What matters to them is the final tenure decision, not the underlying reasons; they are "outcome-oriented", as defined precisely later.

Suppose first the committee uses conclusion-based voting; a vote is taken only on c. Then, clearly, no committee member has an incentive to express an untruthful judgment on c. Individual 1, who wants the committee to accept c, has no incentive to vote against c. Individuals 2 and 3, who want the committee to reject c, have no incentive to vote in favour of c.

But suppose now the committee uses premise-based voting; votes are taken on a and b. What are the members' incentives? Individual 1, who wants the committee to accept c, has no incentive to vote against a or b. But at least one of individuals 2 or 3 has an incentive to vote untruthfully. Specifically, if individuals 1 and 2 vote truthfully, then individual 3 has an incentive to vote untruthfully; and if individuals 1 and 3 vote truthfully, then individual 2 has such an incentive.

To illustrate, assume that individual 2 votes truthfully for a and against b. Then the committee accepts a, regardless of individual 3's vote. So, if individual 3 votes truthfully for b, then the committee accepts b and hence c. But if she votes untruthfully against b, then the committee rejects b and hence c. As individual 3 wants the committee to reject c, she has an incentive to vote untruthfully on b. (In summary, if individual judgments are as in Table 1, voting untruthfully against both a and b weakly dominates voting truthfully for individuals 2 and 3.) Ferejohn (2003) has made this observation informally.

3.2 A non-manipulability condition

To formalize these observations, some definitions are needed. We say that one judgment set, A, *agrees* with another, A^*, on a proposition $p \in X$ if either both or none of A and A^* contains p; A *disagrees* with A^* on p otherwise. Two profiles are *i-variants* of each other if they coincide for all individuals except possibly i.

An aggregation rule F is *manipulable* at the profile $(A_1, \ldots, A_n) \in Domain(F)$ by individual i on proposition $p \in X$ if A_i disagrees with $F(A_1, \ldots, A_n)$ on p, but A_i agrees with $F(A_1, \ldots, A_i^*, \ldots, A_n)$ on p for some i-variant $(A_1, \ldots, A_i^*, \ldots, A_n) \in Domain(F)$.

For example, at the profile in Table 1, premise-based voting is manipulable by individual 3 on c (by submitting $A_3^* = \{\neg a, \neg b, c \leftrightarrow (a \wedge b), \neg c\}$ instead of $A_3 = \{\neg a, b, c \leftrightarrow (a \wedge b), \neg c\}$) and also by individual 2 on c (by submitting $A_2^* = \{\neg a, \neg b, c \leftrightarrow (a \wedge b), \neg c\}$ instead of $A_2 = \{a, \neg b, c \leftrightarrow (a \wedge b), \neg c\}$).

Manipulability thus defined is the existence of an *opportunity* for some individual(s) to manipulate the collective judgment(s) on some proposition(s) by expressing untruthful individual judgments (perhaps on other propositions). The question of when such *opportunities* for manipulation translate into *incentives* for manipulation is a separate question. Whether a rational individual will act on a particular opportunity for manipulation depends on the individual's precise motivation and particularly on how much he or she cares about the various propositions involved in a possible act of manipulation. To illustrate, in our example above, we have assumed that individuals care only about the final tenure decision, implying that they do indeed have incentives to act on their opportunities for manipulation. We discuss this issue in detail when we introduce preferences over judgment sets below.

Our definition of manipulability leads to a corresponding definition of non-manipulability. Let $Y \subseteq X$.

Non-manipulability on Y. F is not manipulable at any profile by any individual on any proposition in Y. Equivalently, for any individual i, profile $(A_1, \ldots, A_n) \in Domain(F)$ and proposition $p \in Y$, if A_i disagrees with $F(A_1, \ldots, A_n)$ on p, then A_i still disagrees with $F(A_1, \ldots, A_i^*, \ldots, A_n)$ on p for every i-variant $(A_1, \ldots, A_i^*, \ldots, A_n) \in Domain(F)$.

This definition specifies a family of non-manipulability conditions, one for each $Y \subseteq X$. Non-manipulability on Y requires the absence of opportunities for manipulation on the subset Y of the agenda. If $Y_1 \subseteq Y_2$, then non-manipulability on Y_2 implies non-manipulability on Y_1. If we refer just to "non-manipulability", without adding "on Y", then we mean the default case $Y = X$.

3.3 A characterization result

When is a judgment aggregation rule non-manipulable? We now characterize the class of non-manipulable aggregation rules in terms of an independence condition and a monotonicity condition. Let $Y \subseteq X$.

Independence on Y. For any proposition $p \in Y$ and profiles $(A_1, \ldots, A_n)$, $(A_1^*, \ldots, A_n^*) \in Domain(F)$, if [for all individuals i, $p \in A_i$ if and only if $p \in A_i^*$] then [$p \in F(A_1, \ldots, A_n)$ if and only if $p \in F(A_1^*, \ldots, A_n^*)$].

Monotonicity on Y. For any proposition $p \in Y$, individual i and pair of i-variants $(A_1, \ldots, A_n), (A_1, \ldots, A_i^*, \ldots, A_n) \in Domain(F)$ with $p \notin A_i$ and $p \in A_i^*$, [$p \in F(A_1, \ldots, A_n)$ implies $p \in F(A_1, \ldots, A_i^*, \ldots, A_n)$].

Weak Monotonicity on Y. For any proposition $p \in Y$, individual i and judgment sets $A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_n$, if there exists a pair of i-variants $(A_1, \ldots, A_n)$, $(A_1, \ldots, A_i^*, \ldots, A_n) \in Domain(F)$ with $p \notin A_i$ and $p \in A_i^*$, then for *some* such pair [$p \in F(A_1, \ldots, A_n)$ implies $p \in F(A_1, \ldots, A_i^*, \ldots, A_n)$].

Informally, independence on Y states that the collective judgment on each proposition in Y depends only on individual judgments *on that proposition* and not on individual judgments *on other propositions*. Monotonicity (respectively, weak monotonicity) on Y states that an additional individual's support for some proposition in Y

never (respectively, not always) reverses the collective acceptance of that proposition (other individuals' judgments remaining fixed).

Again, we have defined families of conditions. If we refer just to "independence" or "(weak) monotonicity", without adding "on Y", then we mean the default case $Y = X$.

Theorem 1 *Let X be any agenda. For each $Y \subseteq X$, if F satisfies universal domain, the following conditions are equivalent:*

(i) F is non-manipulable on Y;

(ii) F is independent on Y and monotonic on Y;

(iii) F is independent on Y and weakly monotonic on Y.

Without a domain assumption (e.g., for a subdomain of the universal domain), (ii) and (iii) are equivalent, and each implies (i).[15]

No assumption on the consistency or completeness of collective judgments is needed. The result can be seen as a preference-free analogue in judgment aggregation of a classic characterization of strategy-proof preference aggregation rules by Barberà et al. (1993).

In the case of a conjunctive (or disjunctive) agenda, conclusion-based voting is independent and monotonic, hence non-manipulable; premise-based voting is not independent, hence manipulable. But on the set of premises $Y = \{a_1, \ldots, a_k\}^{+neg}$ premise-based voting *is* independent and monotonic (as premise-based voting on those premises is simply equivalent to propositionwise majority voting), and hence it is non-manipulable on Y.

3.4 An impossibility result

Ideally, we want to achieve non-manipulability *simpliciter* and not just on some subset of the agenda. Conclusion-based voting is non-manipulable in this strong sense, but generates incomplete collective judgments. Are there any non-manipulable aggregation rules that generate consistent and complete collective judgments? We now show that, for a general class of agendas, including the agenda in the tenure example above, all non-manipulable aggregation rules satisfying some mild conditions are dictatorial.

To define this class of agendas, we define the notion of *path-connectedness*, a variant of the notion of *total-blockedness* introduced by Nehring and Puppe (2002) (originally in the model of "property spaces").[16] Informally, an agenda of propositions under consideration is *path-connected* if any two propositions in the agenda are logically connected with each other, either directly or indirectly, via a sequence of (conditional) logical entailments.

Formally, proposition p *conditionally entails* proposition q if $\{p, \neg q\} \cup Y$ is inconsistent for some $Y \subseteq X$ consistent with p and with $\neg q$. An agenda X is *path-connected* if, for all contingent[17] propositions $p, q \in X$, there is a sequence $p_1, p_2, ..., p_k \in X$ (of

[15] Under universal domain, (i), (ii) and (iii) are also equivalent to the conjunction of independence on Y and judgment-set-wise monotonicity on Y, which requires that, for all individuals i and all i-variants $(A_1, \ldots, A_n), (A_1, \ldots, A_i^*, \ldots, A_n) \in \mathit{Domain}(F)$, if $A_i^* = F(A_1, ..., A_n)$ then $F(A_1, ..., A_i^*, ..., A_n) \cap Y = F(A_1, ..., A_n) \cap Y$.

[16] For a compact logic, path-connectedness is equivalent to total blockedness; in the general case, path-connectedness is weaker.

[17] We call a proposition $p \in \mathbf{L}$ *contingent* if both $\{p\}$ and $\{\neg p\}$ are consistent.

length $k \geq 1$) with $p = p_1$ and $q = p_k$ such that p_1 conditionally entails p_2, p_2 conditionally entails p_3, ..., p_{k-1} conditionally entails p_k. The class of path-connected agendas includes conjunctive and disjunctive agendas (see the Appendix) and the preference agenda (Nehring 2003; Dietrich and List 2007a), which can be used to represent Condorcet-Arrow preference aggregation problems.

Consider the following conditions on an aggregation rule in addition to universal domain.

Collective Rationality. For any profile $(A_1, \ldots, A_n) \in \mathit{Domain}(F)$, $F(A_1, \ldots, A_n)$ is consistent and complete.[18]

Responsiveness. For any contingent proposition $p \in X$, there exist two profiles $(A_1, \ldots, A_n)$, $(A_1^*, \ldots, A_n^*) \in \mathit{Domain}(F)$ such that $p \in F(A_1, \ldots, A_n)$ and $p \notin F(A_1^*, \ldots, A_n^*)$.

Theorem 2 *For a path-connected agenda X (e.g., a conjunctive, disjunctive or preference agenda), an aggregation rule F satisfies universal domain, collective rationality, responsiveness and non-manipulability if and only if F is a dictatorship of some individual.*

For the important case of compact logical languages, this result also follows from Theorem 1 above and Nehring and Puppe's (2002) characterization of monotonic and independent aggregation rules for totally blocked agendas.[19] Theorem 2 is the judgment aggregation analogue of the Gibbard-Satterthwaite theorem on preference aggregation, which shows that dictatorships are the only strategy-proof social choice functions that satisfy universal domain, have three or more options in their range and always produce a determinate winner (Gibbard 1973; Satterthwaite 1975). Below we restate Theorem 2 using a game-theoretic strategy-proofness condition.

In the special case of the preference agenda, however, there is an interesting disanalogy between Theorem 2 and the Gibbard-Satterthwaite theorem. As a collectively rational judgment aggregation rule for the preference agenda represents an Arrowian social welfare function, Theorem 2 establishes an impossibility result on the non-manipulability of social welfare functions (generating orderings as in Arrow's framework) as opposed to social choice functions (generating winning options as in the Gibbard-Satterthwaite framework); for a related result, see Bossert and Storcken (1992).

If the agenda is not path-connected, then there may exist non-dictatorial aggregation rules satisfying all of Theorem 2's conditions; examples of such agendas are not only trivial agendas (containing a single proposition-negation pair or several logically independent such pairs), but also agendas involving only conditionals, including the simple example $X = \{a, b, a \rightarrow b\}^{+neg}$ (Dietrich forthcoming).

By contrast, for *atomically closed* or *atomic* agendas, special cases of path-connected agendas with very rich logical connections, an even stronger impossibility result holds,

[18] Although completeness is conventionally called a rationality requirement, one may consider consistency more important. But if the agenda includes all those propositions on which collective judgments are (practically) required, completeness seems reasonable. Below we discuss relaxing it.

[19] Nehring and Puppe's result implies that the theorem's agenda assumption is maximally weak.

in which Theorem 2's responsiveness condition is significantly weakened.[20]

Weak Responsiveness. The aggregation rule is non-constant. Equivalently, there exist two profiles $(A_1, \ldots, A_n)$, $(A_1^*, \ldots, A_n^*) \in Domain(F)$ such that $F(A_1, \ldots, A_n) \neq F(A_1^*, \ldots, A_n^*)$.

Theorem 3 *For an atomically closed or atomic agenda X, an aggregation rule F satisfies universal domain, collective rationality, weak responsiveness and non-manipulability if and only if F is a dictatorship of some individual.*

Given Theorem 1 above, this result follows immediately from theorems by Pauly and van Hees (2006) (for atomically closed agendas) and Dietrich (2006) (for atomic ones).

3.5 Avoiding the impossibility result

To find non-manipulable and non-dictatorial aggregation rules, we must relax at least one condition in Theorems 2 or 3. Non-responsive rules are usually unattractive. Permitting inconsistent collective judgments also seems unattractive. But the following may sometimes be defensible.

Incompleteness. For a conjunctive or disjunctive agenda, conclusion-based voting is non-manipulable. It generates incomplete collective judgments and is only weakly responsive; this may be acceptable when no collective judgments on the premises are required. More generally, *propositionwise supermajority rules* – requiring a supermajority of a particular size (or even unanimity) for the acceptance of a proposition – are consistent and non-manipulable (by Theorem 1), again at the expense of violating completeness as neither member of a pair $p, \neg p \in X$ might obtain the required supermajority. For a finite agenda (or compact logical languages), a supermajority rule requiring at least m votes for the acceptance of any proposition guarantees collective consistency if and only if $m > n - n/z$, where z is the size of the largest minimal inconsistent set $Z \subseteq X$ (Dietrich and List 2007b; List 2004).

Domain restriction. By suitably restricting the domain of propositionwise majority voting, this rule becomes consistent; it is also non-manipulable as it is independent and monotonic. This result holds, for example, for the domain of all profiles of consistent and complete individual judgment sets satisfying the structure condition of *unidimensional alignment* (List 2003).[21] Informally, unidimensional alignment requires that the individuals can be aligned from left to right (under any interpretation of "left" and "right") such that, for each proposition on the agenda, the individuals accepting the proposition are either exclusively to the left, or exclusively to the right, of those rejecting it. This structure condition captures a shared unidimensional

[20] Agenda X is *atomically closed* if (i) X belongs to classical propositional logic, (ii) if an atomic proposition a occurs in some $p \in X$ then $a \in X$, and (iii) for any atomic propositions $a, b \in X$, we have $a \wedge b, a \wedge \neg b, \neg a \wedge b, \neg a \wedge \neg b \in X$ (Pauly and van Hees 2006). X is *atomic* if $\{\neg p : p$ is an atom of $X\}$ is inconsistent, where $p \in X$ is an *atom of* X if p is consistent but inconsistent with some member of each pair $q, \neg q \in X$ (Dietrich 2006). In both cases, X must contain two (or more) contingent propositions p and q, with p not equivalent to q or $\neg q$.

[21] For a related result on preference aggregation, see Saporiti and Thomé (2005).

conceptualization of the decision problem by the decision-makers. In debates on deliberative democracy, it is sometimes hypothesized that group deliberation may reduce disagreement so as to bring about such a shared unidimensional conceptualization (Miller 1992; Dryzek and List 2003), sometimes also described as a "meta-consensus" (List 2002a).

4 Strategy-proofness

Non-manipulability is not yet a game-theoretic concept. We now define strategy-proofness, a game-theoretic concept that depends on individual preferences (over judgment sets held by the group). We identify assumptions on individual preferences that render strategy-proofness equivalent to non-manipulability and discuss the plausibility of these assumptions.

4.1 Preference relations over judgment sets

We interpret a judgment aggregation problem as a game with n players (the individuals).[22] The game form is given by the aggregation rule: each individual's possible actions are the different judgment sets the individual can submit to the aggregation rule (which may or may not coincide with the individual's true judgment set); the outcomes are the collective judgment sets generated by the aggregation rule.

To specify the game fully, we assume that each individual, in addition to holding a true judgment set A_i, also has a preference relation $\succsim_i$ over all possible outcomes of the game, i.e., over all possible collective judgment sets of the form $A \subseteq X$. For any two judgment sets, $A, B \subseteq X$, $A \succsim_i B$ means that individual i weakly prefers the group to endorse A as the collective judgment set rather than B. We assume that $\succsim_i$ is reflexive and transitive, but do not require it to be complete.[23] Individuals need not be able to rank all pairs of judgment sets relative to each other; in principle, our model allows studying a further relaxation of these conditions.

What preferences over collective judgment sets can we expect an individual i to hold when i's judgment set is A_i? The answer is not straightforward, and it may even be difficult to say *anything* about i's preferences on the basis of A_i alone. To illustrate this, consider first a single proposition p, say, "CO_2 emissions lead to global warming". If individual i judges that p (i.e., $p \in A_i$), it does not necessarily follow that i wants the group to judge that p. Just imagine that i owns an oil company which benefits from low taxes on CO_2 emissions, and that taxes are increased if and only if the group judges that p. In general, accepting p and wanting the group to accept p are conceptually distinct (though the literature is often unclear about this distinction). Whether acceptance and desire of group acceptance happen to coincide in a particular case is an empirical question.[24] There are important situations in which

[22] For an earlier version of this game-theoretic interpretation of judgment aggregation, the notion of closeness-respecting preferences over judgment sets, and a sufficient condition for strategy-proofness (in a sequential context), see List (2002b, 2004).

[23] $\succsim_i$ is: *reflexive* if, for any A, $A \succsim_i A$; *transitive* if, for any A, B, C, $A \succsim_i B$ and $B \succsim_i C$ implies $A \succsim_i C$; *complete* if, for any distinct A, B, $A \succsim_i B$ or $B \succsim_i A$.

[24] This argument identifies accepting with believing, thus interpreting judgment sets as (binary) belief sets, and judgment aggregation as the aggregation of (binary) belief sets into group belief sets. Although this interpretation is standard, other interpretations are possible. If accepting means

the two may indeed be reasonably expected to coincide. An important example is that of *epistemically motivated* individuals: here each individual prefers group judgments that she considers closer to the truth, where she may consider her own judgments as the truth. A *non-epistemically motivated* individual prefers judgment sets for reasons other than the truth, for example because she personally benefits from group actions resulting from the collective endorsement of some judgment sets rather than others.[25]

We now give examples of possible assumptions (empirical claims) on how the individuals' preferences are related to their judgment sets. Which of these assumptions is correct depends on the group of individuals and the aggregation problem in question. Different assumptions capture different motivations of the individuals, as illustrated above. Specifically, the assumption of "unrestricted" preferences captures the case where an individual's preferences are not in any systematic way linked to her judgments; the assumption of "top-respecting" preferences and the stronger one of "closeness-respecting" preferences capture situations in which agents would like group judgments to agree with their own judgments. We use a function C that assigns to each possible judgment set A_i a non-empty set $C(A_i)$ of (reflexive and transitive) preference relations that are considered "compatible" with A_i (i.e., possible given A_i). Our examples of preference assumptions can be stated formally as follows (in increasing order of strength).

Unrestricted preferences. For each A_i, $C(A_i)$ is the set of all preference relations $\succsim$ (regardless of A_i).

Top-respecting preferences. For each A_i, $C(A_i)$ is the set of all preference relations $\succsim$ for which A_i is a most preferred judgment set, i.e., $C(A_i) = \{\succsim\text{: } A_i \succsim B \text{ for all judgment sets } B\}$.

To define "closeness-respecting" preferences, we say that a judgment set B is *at least as close* to A_i on some $Y \subseteq X$ as another judgment set B^* if, for all propositions $p \in Y$, if B^* agrees with A_i on p, then B also agrees with A_i on p. For example, $\{\neg a, b, c \leftrightarrow (a \wedge b), \neg c\}$ is at least as close to $\{a, b, c \leftrightarrow (a \wedge b), c\}$ on X as $\{\neg a, \neg b, c \leftrightarrow (a \wedge b), \neg c\}$,[26] whereas $\{\neg a, b, c \leftrightarrow (a \wedge b), \neg c\}$ and $\{a, \neg b, c \leftrightarrow (a \wedge b), \neg c\}$ are unranked in terms of relative closeness to $\{a, b, c \leftrightarrow (a \wedge b), c\}$ on X. We say that a preference relation $\succsim$ *respects closeness* to A_i on Y if, for any two judgment sets B and B^*, if B is at least as close to A_i as B^* on Y, then $B \succsim B^*$.

Closeness-respecting preferences on Y (for some $Y \subseteq X$). For each A_i, $C(A_i)$ is the set of all preference relations $\succsim$ that respect closeness to A_i on Y, and we write $C = C_Y$.

---

desiring, judgment aggregation is the aggregation of (binary) desire sets into group desire sets. It is then more plausible that i wants the group to accept (desire) the propositions that i accepts (desires).

[25] Even non-epistemically motivated individuals may sometimes prefer group judgments that match their own individual judgments. Suppose each individual is motivated by her desires over outcomes of group actions, which depend on the state of the world. Suppose, further, all individuals hold the same desires over outcomes but different beliefs about the state of the world, and each individual is convinced that her own beliefs are true and that their collective acceptance would lead to the desired outcomes. Such individuals may want the group judgments to match their individual judgments, but mainly to satisfy their desires over outcomes rather than to bring about true group beliefs.

[26] In fact, it is "closer", where "closer than" is the strong component of "at least as close as".

In the important case $Y = X$, we drop the reference "on Y" and speak of closeness-respecting preferences *simpliciter*. One element of $C_X(A_i)$ is the (complete) preference relation induced by the Hamming distance to A_i.[27] Below we analyse the important cases of "reason-oriented" and "outcome-oriented" preferences, where Y is given by particular subsets of X. Generally, if $Y_1 \subseteq Y_2$, then, for all A_i, $C_{Y_1}(A_i) \subseteq C_{Y_2}(A_i)$.

4.2 A strategy-proofness condition

Given a specification of the function C, an aggregation rule is strategy-proof for C if, for any profile, any individual and any preference relation compatible with the individual's judgment set (according to C), the individual (weakly) prefers the outcome of expressing her judgment set truthfully to any outcome that would result from misrepresenting her judgment set.

Strategy-proofness for C. For any individual i, profile $(A_1, \ldots, A_n) \in Domain(F)$ and preference relation $\succsim_i \in C(A_i)$, $F(A_1, \ldots, A_n) \succsim_i F(A_1, \ldots, A_i^*, \ldots, A_n)$ for every i-variant $(A_1 \ldots, A_i^*, \ldots, A_n) \in Domain(F)$.[28]

If the aggregation rule F has the universal domain, then strategy-proofness implies that truthfulness is a weakly dominant strategy for every individual.[29] Our definition of strategy-proofness (generalizing List 2002b, 2004) is similar to Gibbard's (1973) and Satterthwaite's (1975) classical one and related to other definitions of strategy-proofness in the literature on preference aggregation (particularly, for C_X, those by Barberà et al. 1993, 1997 and Nehring and Puppe 2002, employing the notion of generalized single-peaked preferences).

As in the case of non-manipulability above, we have defined a family of strategy-proofness conditions, one for each specification of C. This means that different motivational assumptions about the individuals lead to different strategy-proofness conditions. If individuals have very restrictive preferences over possible judgment sets, then strategy-proofness is easier to achieve than if their preferences are largely unrestricted. Formally, if two functions C_1 and C_2 are such that $C_1 \subseteq C_2$ (i.e., for each A_i, $C_1(A_i) \subseteq C_2(A_i)$), then strategy-proofness for C_1 is less demanding than (i.e., implied by) strategy-proofness for C_2. The more preference relations are compatible with each individual judgment set, the more demanding is the corresponding requirement of strategy-proofness.

[27] The Hamming distance between two judgment sets B and B^* is $d(B, B^*) := |\{p \in X : B \text{ and } B^* \text{ disagree on } p\}|$. The preference relation $\succeq$ induced by Hamming distance to A_i is defined, for any B, B^*, by $[B \succeq B^*$ if and only if $d(B, A_i) \leq d(B^*, A_i)]$. For the preference agenda, a preference relation $\succeq$ over judgment sets (each representing a preference ordering over the option set K) represents a meta-preference over preference orderings. Bossert and Storcken (1992) use the Kemeny distance between preference orderings to obtain such a meta-preference. For related work on distances between preferences and theories, see Baigent (1987) and Schulte (2005), respectively.

[28] Our definition of strategy-proofness can be generalized by admitting a different function C_i for each individual i. This removes a homogeneity assumption, whereby, if individuals i and j hold the same judgment set $A_i = A_j$, then their preference relations fall into the same set $C(A_i) = C(A_j)$. The homogeneity assumption is undemanding when $C(A_i)$ is large.

[29] This interpretation of strategy-proofness holds for product domains. For certain subdomains of the universal domain (i.e., non-product domains), we do not have a strictly well defined game, but our definition of strategy-proofness remains applicable and can be reinterpreted as one of "conditional strategy-proofness" for non-product domains, as discussed by Saporiti and Thomé (2005).

4.3 The equivalence of strategy-proofness and non-manipulability

What is the logical relation between non-manipulability as defined above and strategy-proofness? We show that, if preferences are closeness-respecting (on some $Y \subseteq X$), then an equivalence between these two concepts arises. Let X be any agenda.

Theorem 4 *For each $Y \subseteq X$, F is strategy-proof for C_Y if and only if F is non-manipulable on Y.*

In other words, for any subset Y of the agenda X (including the case $Y = X$), strategy-proofness of an aggregation rule for closeness-respecting preferences on Y is equivalent to non-manipulability on the propositions in Y. In particular, strategy-proofness for closeness-respecting preferences *simpliciter* is equivalent to non-manipulability *simpliciter*. This also implies that, for unrestricted or top-respecting preferences, strategy-proofness is more demanding than our default condition of non-manipulability, whereas, for closeness-respecting preferences on some $Y \subsetneq X$, it is less demanding.

Given the equivalence result of Theorem 4, we can now state corollaries of Theorems 1 and 2 above for strategy-proofness:[30]

Corollary 1 *For each $Y \subseteq X$, if F satisfies universal domain, the following conditions are equivalent:*

(i) F is strategy-proof for C_Y;

(ii) F is independent on Y and monotonic on Y;

(iii) F is independent on Y and weakly monotonic on Y.

Without a domain assumption (e.g., for a subdomain of the universal domain), (ii) and (iii) are equivalent, and each implies (i).

Corollary 2 *For a path-connected agenda X (e.g., a conjunctive, disjunctive or preference agenda), an aggregation rule F satisfies universal domain, collective rationality, responsiveness and strategy-proofness for C_X if and only if F is a dictatorship of some individual.*

Corollary 2 is a judgment aggregation analogue of Nehring and Puppe's (2002) characterization of strategy-proof social choice functions in the model of "property spaces".[31] The negative part of corollary 2 (i.e., if an aggregation rule satisfies the conditions, then it is a dictatorship) holds not only for closeness-respecting preferences (C_X) but for any preference specification C at least as broad as C_X, i.e., $C_X \subseteq C$, as strategy-proofness for C then implies strategy-proofness for C_X. The positive part of corollary 2 (i.e., if an aggregation rule is a dictatorship, then it satisfies the conditions) holds for any preference specification C allowing only top-respecting preferences, i.e., for any C such that, if $\succsim \in C(A_i)$, then $A_i \succsim B$ for all judgment sets B; otherwise a dictatorship, although non-manipulable, is not strategy-proof (to see this point, recall the example of the oil company in Section 4.1).

[30] Our remarks on Theorems 1 and 2 above also apply to Corollaries 1 and 2.

[31] For compact logics, it follows from their result via Corollary 1. As noted, a disanalogy lies in the aggregation rule's different informational input. In Barberà et al. (1993, 1997) and Nehring and Puppe (2002), each individual submits a preference relation, here a single judgment set. Under some conditions, judgment sets can be associated with peaks of preference relations.

In summary, if the individuals' preferences over judgment sets are unrestricted, top-respecting or closeness-respecting, we obtain a negative result. Moreover, in analogy with Theorem 3 above, for atomically closed or atomic agendas, we get an impossibility result even if we weaken responsiveness to the requirement of a non-constant aggregation rule.

5 Outcome- and reason-oriented preferences

As we have introduced families of strategy-proofness and non-manipulability conditions, it is interesting to consider some less demanding conditions within these families. If we demand strategy-proofness for $C = C_X$, equivalent to non-manipulability *simpliciter*, this precludes all incentives for manipulation, where individuals have closeness-respecting preferences. But individual preferences may sometimes fall into a more restricted set: they may be closeness-respecting on some subset $Y \subsetneq X$, in which case it is sufficient to require strategy-proofness for C_Y. As an illustration, we now apply these ideas to the case of a conjunctive (analogously disjunctive) agenda.

5.1 Definitions

Let X be a conjunctive (or disjunctive) agenda. Two important cases of closeness-respecting preferences on Y are the following.

Outcome-oriented preferences. $C = C_{Y_{outcome}}$, where $Y_{outcome} = \{c\}^{+neg}$.

Reason-oriented preferences. $C = C_{Y_{reason}}$, where $Y_{reason} = \{a_1, ..., a_k\}^{+neg}$.

An individual with outcome-oriented preferences cares only about achieving a collective judgment on the conclusion that matches her own judgment, regardless of the premises. Such preferences make sense if only the conclusion but not the premises have consequences the individual cares about. An individual with reason-oriented preferences cares only about achieving collective judgments on the premises that match her own judgments, regardless of the conclusion. Such preferences make sense if the individual gives primary importance to the reasons given in support of outcomes, rather than the outcomes themselves, or if the group's judgments on the premises have important consequences themselves that the individual cares about (such as setting precedents for future decisions). Proponents of a deliberative conception of democracy often argue that the motivational assumption of reason-oriented preferences is appropriate in deliberative settings (for a discussion, see Elster 1986; Goodin 1986). Economists, by contrast, assume that in many settings outcome-oriented preferences are the more accurate motivational assumption. Ultimately, it is an empirical question what preferences are triggered by various settings.

To illustrate, consider premise-based voting and the profile in Table 1. Individual 3's judgment set is $A_3 = \{\neg a, b, \neg c, r\}$, where $r = c \leftrightarrow (a \wedge b)$. If all individuals are truthful, the collective judgment set is $A = \{a, b, c, r\}$. If individual 3 untruthfully submits $A_3^* = \{\neg a, \neg b, \neg c, r\}$ and individuals 1 and 2 are truthful, the collective judgment set is $A^* = \{a, \neg b, \neg c, r\}$. Now A^* is closer to A_3 than A on $Y_{outcome} =$

$\{c\}^{+neg}$, whereas A is closer to A_3 than A^* on $Y_{reason} = \{a, b\}^{+neg}$. So, under outcome-oriented preferences, individual 3 (at least weakly) prefers A^* to A, whereas, under reason-oriented preferences, individual 3 (at least weakly) prefers A to A^*.

5.2 The strategy-proofness of premise-based voting for reason-oriented preferences

As shown above, conclusion-based voting is strategy-proof for C_X and hence also for $C_{Y_{reason}}$ and $C_{Y_{outcome}}$. Premise-based voting is not strategy-proof for C_X and neither for $C_{Y_{outcome}}$, as can easily be seen from our first example of manipulation. But the following holds.

Proposition 1 *For a conjunctive or disjunctive agenda X, premise-based voting is strategy-proof for $C_{Y_{reason}}$.*

This result is interesting from a deliberative democracy perspective. If individuals have reason-oriented preferences in deliberative settings, as sometimes argued by proponents of a deliberative conception of democracy, then premise-based voting is strategy-proof in such settings. But if individuals have outcome-oriented preferences, then the aggregation rule advocated by deliberative democrats is vulnerable to strategic manipulation, posing a challenge to the deliberative democrats' view that truthfulness can easily be achieved under their preferred aggregation rule.

5.3 The strategic equivalence of premise- and conclusion-based voting for outcome-oriented preferences

Surprisingly, if individuals have outcome-oriented preferences, then premise- and conclusion-based voting are strategically equivalent in the following sense. For any profile, there exists, for each of the two rules, a (weakly) dominant-strategy equilibrium leading to the same collective judgment on the conclusion. To state this result formally, some definitions are needed.

Under an aggregation rule F, for individual i with preference ordering $\succsim_i$, submitting the judgment set B_i (which may or may not coincide with individual i's true judgment set A_i) is a *weakly dominant strategy* if, for every profile $(B_1, \ldots, B_i, \ldots, B_n) \in Domain(F)$, $F(B_1, \ldots, B_i, \ldots, B_n) \succsim_i F(B_1, \ldots, B_i^*, \ldots, B_n)$ for every i-variant $(B_1, \ldots, B_i^*, \ldots, B_n) \in Domain(F)$.

Two aggregation rules F and G with identical domain are *strategically equivalent* on $Y \subseteq X$ for C if, for every profile $(A_1, \ldots, A_n) \in Domain(F) = Domain(G)$ and preference relations $\succsim_1 \in C(A_1)$, $\ldots$, $\succsim_n \in C(A_n)$, there exist profiles $(B_1, \ldots, B_n)$, $(C_1, \ldots, C_n) \in Domain(F) = Domain(G)$ such that

(i) for each individual i, submitting B_i is a weakly dominant strategy under rule F and submitting C_i is a weakly dominant strategy under rule G;

(ii) $F(B_1, \ldots, B_n)$ and $G(C_1, \ldots, C_n)$ agree on every proposition $p \in Y$.

Theorem 5 *For a conjunctive or disjunctive agenda X, premise- and conclusion-based voting are strategically equivalent on $Y_{outcome} = \{c\}^{+neg}$ for $C_{Y_{outcome}}$.*

Despite the differences between premise- and conclusion-based voting, if individuals have outcome-oriented preferences and act on appropriate weakly dominant

strategies, the two rules generate identical collective judgments on the conclusion. This is surprising as premise- and conclusion-based voting are regarded in the literature as two diametrically opposed aggregation rules.

6 Concluding remarks

As judgment aggregation problems arise in many real-world decision-making bodies, it is important to understand which judgment aggregation rules are vulnerable to manipulation and which not. We have introduced a non-manipulability condition for judgment aggregation and characterized the class of non-manipulable judgment aggregation rules. Non-manipulability rules out the existence of *opportunities* for manipulation by the untruthful expression of individual judgments. We have then defined a game-theoretic strategy-proofness condition and shown that, under some (but not all) motivational assumptions, it is equivalent to non-manipulability, as defined earlier. For these motivational assumptions, our characterization of non-manipulable aggregation rules has allowed us to characterize all strategy-proof aggregation rules. Strategy-proofness rules out the existence of *incentives* for manipulation. Crucially, if individuals do not generally want the group to make collective judgments that match their own individual judgments, the concepts of non-manipulability and strategy-proofness may come significantly apart.

We have also proved an impossibility result that is the judgment aggregation analogue of the classical Gibbard-Satterthwaite theorem on preference aggregation. For the class of path-connected agendas, including conjunctive, disjunctive and preference agendas, all non-manipulable aggregation rules satisfying some mild conditions are dictatorial. The impossibility result becomes even stronger for agendas with particularly rich logical connections between propositions.

To avoid this impossibility, we have suggested that permitting incomplete collective judgments or domain restrictions are the most promising routes. For example, conclusion-based voting is strategy-proof, but violates completeness. Another way to avoid the impossibility is to relax non-manipulability or strategy-proofness itself. Both conditions fall into more general families of conditions of different strength. Instead of requiring non-manipulability on the entire agenda of propositions, we may require non-manipulability only on some subset of the agenda. Premise-based voting, for example, is non-manipulable on the set of premises, but not non-manipulable *simpliciter*. Whether such a weaker non-manipulability condition is sufficient in practice depends on how worried we are about possible opportunities for manipulation on propositions outside the subset of the agenda for which non-manipulability holds. Likewise, instead of requiring strategy-proofness for a large class of individual preferences over judgment sets, we may require strategy-proofness only for a restricted class of preferences, for example for "outcome-" or "reason-oriented" preferences. Premise-based voting, for example, is strategy-proof for "reason-oriented" preferences. Whether such a weaker strategy-proofness condition is sufficient in practice depends on the motivations of the decision-makers.

Finally, we have shown that, for "outcome-oriented" preferences, premise- and conclusion-based voting are strategically equivalent. They generate the same collective judgment on the conclusion if individuals act on appropriate weakly dominant

strategies.

Our results raise questions about a prominent position in the literature, according to which premise-based voting is superior to conclusion-based voting from a deliberative democracy perspective. We have shown that, with respect to non-manipulability and strategy-proofness, conclusion-based voting outperforms premise-based voting. This result could be generalized beyond conjunctive and disjunctive agendas.

Until now, comparisons between judgment aggregation and preference aggregation have focused mainly on Condorcet's paradox and Arrow's theorem. With this paper, we hope to inspire further research on strategic voting and a game-theoretic perspective in a judgment aggregation context. An important challenge is the development of models of *deliberation* on interconnected propositions – where individuals not only "feed" their judgments into some aggregation rule, but where they deliberate about the propositions prior to making collective judgments – and the study of the strategic aspects of such deliberation. We leave this challenge for further work.

7 References

Barberà, S., F. Gul and E. Stacchetti. 1993. "Generalized Median Voter Schemes and Committees." *Journal of Economic Theory* 61: 262-289.

Barberà, S., J. Massóa and A. Nemeb. 1997. "Voting under constraints." *Journal of Economic Theory* 76(2): 298-321.

Baigent, N. 1987. "Preference Proximity and Anonymous Social Choice." *Quarterly Journal of Economics* 102(1): 161-170.

Bovens, L., and W. Rabinowicz. 2006. "Democratic Answers to Complex Questions - An Epistemic Perspective." *Synthese* 150(1, May): 131-153.

Bossert, W., and T. Storcken. 1992. "Strategy-proofness of social welfare functions: the use of the Kemeny distance between preference orderings." *Social Choice and Welfare* 9: 345-360.

Brams, S. J., D. M. Kilgour and W. S. Zwicker. 1997. "Voting on Referenda: the Separability Problem and Possible Solutions." *Electoral Studies* 16(3): 359-377.

Brams, S.J., D. M. Kilgour and W. S. Zwicker. 1998. "The paradox of multiple elections." *Social Choice and Welfare* 15: 211-236.

Brennan, G. 2001. "Collective Coherence?" *International Review of Law and Economics* 21(2, June): 197-211.

Chapman, B. 1998. "More Easily Done than Said: Rules, Reason and Rational Social Choice." *Oxford Journal of Legal Studies* 18(Summer): 293-330.

Chapman, B. 2002. "Rational Aggregation." *Politics, Philosophy and Economics* 1(October): 337-354.

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

Dietrich, F. 2007. "A generalised model of judgment aggregation." *Social Choice and Welfare* 28(4): 529-565.

Dietrich F. Forthcoming. "The possibility of judgment aggregation on agendas with subjunctive implications." *Journal of Economic Theory.*

Dietrich, F., and C. List. 2007a. "Arrow's theorem in judgment aggregation." *Social Choice and Welfare* 29(1): 19-33.

Dietrich, F., and C. List. 2007b. "Judgment aggregation by quota rules." *Journal of Theoretical Politics* 19(4, in press).

Dokow, E., and R. Holzman. 2005. "Aggregation of binary evaluations." Working paper, Technion Israel Institute of Technology.

Dryzek, J., and C. List. 2003. "Social Choice Theory and Deliberative Democracy: A Reconciliation." *British Journal of Political Science* 33(1, January): 1-28.

Elster, J. 1986. "The Market and the Forum." In *Foundations of Social Choice Theory*, ed. J. Elster and A. Hylland. Cambridge, Cambridge University Press, 103-132.

Ferejohn, J. 2003. "Conversability and Collective Intention." Paper presented at the Common Minds Conference, Australian National University, 24-25 July 2003.

Gärdenfors, P. 2006. "An Arrow-like theorem for voting with logical consequences." *Economics and Philosophy* 22(2): 181-190.

Gibbard, A. 1973. "Manipulation of voting schemes: a general result." *Econometrica* 41(July): 587-601.

Goodin, R. E. 1986. "Laundering preferences." In *Foundations of Social Choice Theory*, ed. J. Elster and A. Hylland. Cambridge, Cambridge University Press, 75-101.

Grofman, B. 1985. "Research Note: The Accuracy of Group Majorities for Disjunctive and Conjunctive Decision Tasks." *Organizational Behavior and Human Decision Processes* 35: 119-123.

van Hees, M. 2007. "The limits of epistemic democracy." *Social Choice and Welfare* 28(4): 649-666.

Kelly, J. S. 1989. "The Ostrogorski Paradox." *Social Choice and Welfare* 6: 71-76.

Konieczny, S., and R. Pino-Perez. 2002. "Merging information under constraints: a logical framework." *Journal of Logic and Computation* 12: 773-808.

Kornhauser, L. A., and L. G. Sager. 1986. "Unpacking the Court." *Yale Law Journal* 96(1, November): 82-117.

List, C. 2002a. "Two Concepts of Agreement." *The Good Society* 11(1): 72-79.

List, C. 2002b. "Discursive Path-Dependencies." Nuffield College Working Paper in Politics 2002-W15 (9 May 2002).

List, C. 2003. "A Possibility Theorem on Aggregation over Multiple Interconnected Propositions." *Mathematical Social Sciences* 45(1, February): 1-13 (with Corrigendum in *Mathematical Social Sciences* 52:109-110).

List, C. 2004. "A Model of Path Dependence in Decisions over Multiple Propositions." *American Political Science Review* 98(3, August): 495-513.

List, C. 2005. "The Probability of Inconsistencies in Complex Collective Decisions." *Social Choice and Welfare* 24(1, February): 3-32.

List, C. 2006. "The Discursive Dilemma and Public Reason." *Ethics* 116(2, January): 362-402.

List, C., and P. Pettit. 2002. "Aggregating Sets of Judgments: An Impossibility Result." *Economics and Philosophy* 18(1, April): 89-110.

List, C., and P. Pettit. 2004. "Aggregating Sets of Judgments: Two Impossibility Results Compared." *Synthese* 140(1-2): 207-235.

Miller, D. 1992. "Deliberative Democracy and Social Choice." *Political Studies* 40(Special Issue), 54-67.

Nehring, K. 2003. "Arrow's theorem as a corollary." *Economics Letters* 80(3, September): 379-382.

Nehring, K., and C. Puppe. 2002. "Strategyproof Social Choice on Single-Peaked Domains: Possibility, Impossibility and the Space Between." Working paper, University of California at Davies.

Nehring, K., and C. Puppe. 2005. "Consistent Judgement Aggregation: A Characterization." Working paper, University of Karlsruhe.

Osherson, D., and M. Vardi. Forthcoming. "Aggregating Disparate Estimates of Chance." *Games and Economic Behavior.*

Pauly, M., and M. van Hees. 2006. "Logical Constraints on Judgment Aggregation." *Journal of Philosophical Logic* 35(6, December): 569-585.

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

Pigozzi, G. 2006. "Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation." *Synthese* 152(2): 285-298.

Saporiti, A., and F. Thomé. 2005. "Strategy-proofness and single-crossing." Working paper, Queen Mary, University of London.

Satterthwaite, M. 1975. "Strategyproofness and Arrow's conditions: existence and correspondences for voting procedures and social welfare functions." *Journal of Economic Theory* 10(April): 187-217.

Schulte, O. 2005. "Minimal belief change, Pareto-optimality and logical consequence." *Economic Theory* 19(1): 105-144.

Sunstein, C. 1994. "Political Conflict and Legal Agreement." Tanner Lectures on Human Values, Harvard.

Taylor, A. D. 2002. "The Manipulability of Voting Systems. American Mathematical Monthy." 109(April): 321-337.

Taylor, A. D. 2005. *Social Choice and the Mathematics of Manipulation.* Cambridge, Cambridge University Press.

Wilson, R. 1975. "On the Theory of Aggregation." *Journal of Economic Theory* 10: 89-99.

A Appendix

Proof of Theorem 1. Let $Y \subseteq X$. We prove first that (ii) and (iii) are equivalent, then that (ii) implies (i), and then that, given universal domain, (i) implies (ii).

(ii) implies (iii). Trivial as monotonicity on Y implies weak monotonicity on Y.

(iii) implies (ii). Suppose F is independent on Y and weakly monotonic on Y. To show monotonicity on Y, note that in the requirement defining weak monotonicity on Y one may, by independence on Y, replace "for some such pair" by "for all such pairs". The modified requirement is equivalent to monotonicity on Y.

(ii) implies (i). Suppose F is independent on Y and monotonic on Y. To show non-manipulability on Y, consider any proposition $p \in Y$, individual i, and profile $(A_1, ..., A_n) \in Domain(F)$, such that $F(A_1, ..., A_n)$ disagrees with A_i on p. Take any i-variant $(A_1, ..., A_i^*, ..., A_n) \in Domain(F)$. We have to show that $F(A_1, ..., A_i^*, ..., A_n)$ still disagrees with A_i on p. Assume first that A_i and A_i^* agree on p. Then in both profiles $(A_1, ..., A_n)$ and $(A_1, ..., A_i^*, ..., A_n)$ exactly the same individuals accept p.

Hence, by independence on Y, $F(A_1, ..., A_i^*, ..., A_n)$ agrees with $F(A_1, ..., A_n)$ on p, hence disagrees with A_i on p. Now assume A_i^* disagrees with A_i on p, i.e., agrees with $F(A_1, ..., A_n)$ on p. Then, by monotonicity on Y, $F(A_1, ..., A_i^*, ..., A_n)$ agrees with $F(A_1, ..., A_n)$ on p, i.e., disagrees with A_i on p.

(i) implies (ii). Now assume universal domain, and let F be non-manipulable on Y. To show monotonicity on Y, consider any proposition $p \in Y$, individual i, and pair of i-variants $(A_1, ..., A_n), (A_1, ..., A_i^*, ..., A_n) \in Domain(F)$ with $p \notin A_i$ and $p \in A_i^*$. If $p \in F(A_1, ..., A_n)$, then A_i disagrees on p with $F(A_1, ..., A_n)$, hence also with $F(A_1, ..., A_i^*, ..., A_n)$ by non-manipulability on Y. So $p \in F(A_1, ..., A_i^*, ..., A_n)$. To show independence on Y, consider any proposition $p \in Y$ and profiles $(A_1, ..., A_n)$, $(A_1^*, ..., A_n^*) \in Domain(F)$ such that, for all individuals i, A_i and A_i^* agree on p. We have to show that $F(A_1, ..., A_n)$ and $F(A_1^*, ..., A_n^*)$ agree on p. Starting with the profile $(A_1, ..., A_n)$, we replace first A_1 by A_1^*, then A_2 by A_2^*, ..., then A_n by A_n^*. By universal domain, each replacement leads to a profile still in $Domain(F)$. We now show that each replacement preserves the collective judgment about p. Assume for contradiction that for individual i replacement of A_i by A_i^* changes the collective judgment about p. Since A_i and A_i^* agree on p but the respective outcomes for A_i and for A_i^* disagree on p, either A_i or A_i^* (but not both) disagrees with the respective outcome. This is a contradiction, since it allows individual i to manipulate: in the first case by submitting A_i^* with genuine judgment set A_i, in the second case by submitting A_i with genuine judgment set A_i^*. Since no replacement has changed the collective judgment about p, it follows that $F(A_1, ..., A_n)$ and $F(A_1^*, ..., A_n^*)$ agree on p, which proves independence on Y. ■

For any propositions p, q, we write $p \vDash^* q$ to mean that p *conditionally entails* q.

Proof that conjunctive and disjunctive agendas are path-connected. Let X be the conjunctive agenda $X = \{a_1, \neg a_1, ..., a_k, \neg a_k, c, \neg c, r, \neg r\}$, where $k \geq 1$ and r is the connection rule $c \leftrightarrow (a_1 \wedge ... \wedge a_k)$. (The proof for a disjunctive agenda is analogous.) We have to show that for any $p, q \in X$ there is a sequence $p = p_1, p_2, ..., p_k = q$ in X $(k \geq 1)$ such that $p_1 \vDash^* p_2, p_2 \vDash^* p_3, ..., p_{k-1} \vDash^* p_k$. To show this, it is sufficient to prove that

$$p \vDash^* q \text{ for any propositions } p, q \in X \text{ of } \textit{different types}, \tag{1}$$

where a proposition is of type 1 if it is a possibly negated premise $(a_1, \neg a_1, ..., a_k, \neg a_k)$, of type 2 if it is the possibly negated conclusion $(c, \neg c)$ and of type 3 if it is the possibly negated connection rule $(r, \neg r)$. The reason is (in short) that, if (1) holds, then, for any $p, q \in X$ of the *same* type, taking any $s \in X$ of a different type, there is by (1) a path connecting p to s and a path connecting s to q; the concatenation of both paths connects p to q, as desired. As $p \vDash^* q$ if and only if $\neg q \vDash^* \neg p$ (use both times the same Y), claim (1) is equivalent to

$$p \vDash^* q \text{ for any propositions } p, q \in X \text{ such that } p \text{ has smaller type than } q. \tag{2}$$

We show (2) by going through the different cases (where $j \in \{1, ..., k\}$):

From type 2 to type 3: we have $c \vDash^* r$ and $\neg c \vDash^* \neg r$ (take $Y = \{a_1, ..., a_k\}$ both times), and $c \vDash^* \neg r$ and $\neg c \vDash^* r$ (take $Y = \{\neg a_1\}$ both times).

From type 1 to type 2: we have $a_j \vDash^* c$ and $\neg a_j \vDash^* \neg c$ (take $Y = \{r, a_1, ..., a_{j-1}, a_{j+1}, ..., a_k\}$ both times), and $a_j \vDash^* \neg c$ and $\neg a_j \vDash^* c$ (take $Y = \{\neg r, a_1, ..., a_{j-1}, a_{j+1}, ..., a_k\}$ both times);

From type 1 to type 3: we have $a_j \vDash^* r$ and $\neg a_j \vDash^* \neg r$ (take $Y = \{c, a_1, ..., a_{j-1}, a_{j+1}, ..., a_k\}$ both times), and $a_j \vDash^* \neg r$ and $\neg a_j \vDash^* r$ (take $Y = \{\neg c, a_1, ..., a_{j-1}, a_{j+1}, ..., a_k\}$ both times). ■

Proof of Theorem 2. Let X be path-connected. If F is dictatorial, it obviously satisfies universal domain, collective rationality, responsiveness and non-manipulability. Now suppose F has all these properties, hence is also independent and monotonic by Theorem 1. We show that F is dictatorial. If X contains no contingent proposition, F is trivially dictatorial (where each individual is a dictator). From now on, suppose X is not of this degenerate type. For any consistent set $Z \subseteq X$, let A_Z be some consistent and complete judgment set such that $Z \subseteq A_Z$ (which exists by L1-L3).

Claim 1. F satisfies the unanimity principle: for any $p \in X$ and any $(A_1, ..., A_n) \in Domain(F)$, if $p \in A_i$ for each i then $p \in F(A_1, ..., A_n)$.

Consider any $p \in X$ and $(A_1, ..., A_n) \in Domain(F)$ such that $p \in A_i$ for every i. Since the sets A_i are consistent, p is consistent. If $\neg p$ is inconsistent (i.e., p is a tautology), $p \in F(A_1, ..., A_n)$ by collective rationality. Now suppose $\neg p$ is consistent. As each of $p, \neg p$ is consistent, p is contingent. So, by responsiveness, there exists a profile $(B_1, ..., B_n) \in Domain(F)$ such that $p \in F(B_1, ..., B_n)$. In $(B_1, ..., B_n)$ we now replace one by one each judgment set B_i by A_i, until we obtain the profile $(A_1, ..., A_n)$. Each replacement preserves the collective acceptance of p, either by monotonicity (if $p \notin B_i$) or by independence (if $p \in B_i$). So $p \in F(A_1, ..., A_n)$, as desired.

Claim 2. F is systematic: there exists a set $\mathcal{W}$ of ("winning") coalitions $C \subseteq N$ such that, for every $(A_1, ..., A_n) \in Domain(F)$, $F(A_1, ..., A_n) = \{p \in X : \{i : p \in A_i\} \in \mathcal{W}\}$.

For each contingent $p \in X$, let $\mathcal{W}_p$ be the set all subsets $C \subseteq N$ such that $p \in F(A_1, ..., A_n)$ for some (hence by independence any) $(A_1, ..., A_n) \in Domain(F)$ with $\{i : p \in A_i\} = C$. Consider any contingent $p, q \in X$. We prove that $\mathcal{W}_p = \mathcal{W}_q$. Suppose $C \in \mathcal{W}_p$, and let us show that $C \in \mathcal{W}_q$; this proves the inclusion $\mathcal{W}_q \subseteq \mathcal{W}_p$, and the converse inclusion can be shown analogously. As X is path-connected, there are $p = p_1, p_2, ..., p_k = q \in X$ with $p_1 \vDash^* p_2$, $p_2 \vDash^* p_3$, ..., $p_{k-1} \vDash^* p_k$. We show by induction that $C \in \mathcal{W}_{p_j}$ for all $j = 1, 2, ..., k$. If $j = 1$ then $C \in C_{p_1}$ by $p_1 = p$. Now let $1 \leq j < k$ and assume $C \in \mathcal{W}_{p_j}$. By $p_j \vDash^* p_{j+1}$, there is a set $Y \subseteq X$ such that $\{p_j\} \cup Y$ and $\{\neg p_{j+1}\} \cup Y$ are each consistent but $\{p_j, p_{j+1}\} \cup Y$ is inconsistent. It follows that each of $\{p_j, p_{j+1}\} \cup Y$ and $\{\neg p_j, \neg p_{j+1}\} \cup Y$ is consistent (using L3 in conjunction with L1,L2). So we may define a profile $(A_1, ..., A_n) \in Domain(F)$ by

$$A_i := \begin{cases} A_{\{p_j, p_{j+1}\} \cup Y} & \text{if } i \in C \\ A_{\{\neg p_j, \neg p_{j+1}\} \cup Y} & \text{if } i \in N \backslash C. \end{cases}$$

Since $Y \subseteq A_i$ for all i, $Y \subseteq F(A_1, ..., A_n)$ by claim 1. Since $\{i : p_j \in A_i\} = C \in \mathcal{W}_{p_j}$, we have $p_j \in F(A_1, ..., A_n)$. So $\{p_j\} \cup Y \subseteq F(A_1, ..., A_n)$. Hence, since $\{p_j, \neg p_{j+1}\} \cup Y$ is inconsistent, $\neg p_{j+1} \notin F(A_1, ..., A_n)$, whence $p_{j+1} \in F(A_1, ..., A_n)$. So, as $\{i : p_{j+1} \in A_i\} = C$, we have $C \in \mathcal{W}_{p_{j+1}}$, as desired.

As $\mathcal{W}_p$ is the same set for each contingent $p \in X$, let $\mathcal{W}$ be this set. To complete the proof of the claim, it is sufficient to show that, for every $(A_1, ..., A_n) \in Domain(F)$

and every $p \in X$, $p \in F(A_1, ..., A_n)$ if and only if $\{i : p \in A_i\} \in \mathcal{W}$. If p is contingent this holds by definition of $\mathcal{W}$; if p is a tautology it holds because $p \in F(A_1, ..., A_n)$ (by collective rationality), $\{i : p \in A_i\} = N$ (by universal domain) and $N \in \mathcal{W}$ (by claim 1); analogously, if p is a contradiction it holds because $p \notin F(A_1, ..., A_n)$, $\{i : p \in A_i\} = \emptyset$ and $\emptyset \notin \mathcal{W}$.

Claim 3. (1) $N \in \mathcal{W}$; (2) for every coalition $C \subseteq N$, $C \in \mathcal{W}$ if and only if $N \backslash C \notin \mathcal{W}$; (3) for every coalitions $C, C^* \subseteq N$, if $C \in \mathcal{W}$ and $C \subseteq C^*$ then $C^* \in \mathcal{W}$.

Part (1) follows from claim 1. Regarding parts (2) and (3), note that, for any $C \subseteq N$, there exists a $p \in X$ and an $(A_1, ..., A_n) \in Domain(F)$ with $\{i : p \in A_i\} = C$; this holds because X contains a contingent proposition p. Part (2) holds because, for any $(A_1, ..., A_n) \in Domain(F)$, each of the sets $A_1, ..., A_n, F(A_1, ..., A_n)$ contains exactly one member of any pair $p, \neg p \in X$, by universal domain and collective rationality. Part (3) follows from a repeated application of monotonicity and universal domain.

Claim 4. There exists an inconsistent set $Y \subseteq X$ with pairwise disjoint subsets Z_1, Z_2, Z_3 such that $(Y \backslash Z_j) \cup Z_j^{\neg}$ is consistent for any $j \in \{1, 2, 3\}$. Here, $Z^{\neg} := \{\neg p : p \in Z\}$ for any $Z \subseteq X$.

By assumption, there exists a contingent $p \in X$; also $\neg p$ is then contingent. So, by path-connectedness, there exist $p = p_1, p_2, ..., p_k = \neg p \in X$ and $Y_1^*, Y_2^*, ..., Y_{k-1}^* \subseteq X$ such that

$$\text{for each } t \in \{1, ..., k-1\},\ \{p_t, \neg p_{t+1}\} \cup Y_t^* \text{ is inconsistent; and} \tag{3}$$

$$\text{for each } t \in \{1, ..., k-1\},\ \{p_t\} \cup Y_t^* \text{ and } \{\neg p_{t+1}\} \cup Y_t^* \text{ are consistent.} \tag{4}$$

From (*3*) and (*4*) it follows (using L3 in conjunction with L1, L2) that

$$\text{for each } t \in \{1, ..., k-1\},\ \{p_t, p_{t+1}\} \cup Y_t^* \text{ and } \{\neg p_t, \neg p_{t+1}\} \cup Y_t^* \text{ are consistent.} \tag{5}$$

We first show that there exists a $t \in \{1, ..., k-1\}$ such that $\{p_t, \neg p_{t+1}\}$ is consistent. Assume for contradiction that each of $\{p_1, \neg p_2\}, ..., \{p_{k-1}, \neg p_k\}$ is inconsistent. So (using L2) each of $\{p_1, \neg p_2\}, \{p_1, p_2, \neg p_3\}, ..., \{p_1, ..., p_{k-1}, \neg p_k\}$ is inconsistent. As $\{p_1\} = \{p\}$ is consistent, either $\{p_1, p_2\}$ or $\{p_1, \neg p_2\}$ is consistent (by L2 and L3); hence, as $\{p_1, \neg p_2\}$ is inconsistent, $\{p_1, p_2\}$ is consistent. So either $\{p_1, p_2, p_3\}$ or $\{p_1, p_2, \neg p_3\}$ is consistent (again by L2 and L3); hence, as $\{p_1, p_2, \neg p_3\}$ is inconsistent, $\{p_1, p_2, p_3\}$ is consistent. Continuing this argument, it follows after $k-1$ steps that $\{p_1, ..., p_k\}$ is consistent. Hence $\{p_1, p_k\}$ is consistent (by L2), i.e., $\{p, \neg p\}$ is consistent, a contradiction (by L1).

We have shown that there is a $t \in \{1, ..., k-1\}$ such that $\{p_t, \neg p_{t+1}\}$ is consistent, whence $Y_t^* \neq \emptyset$ by (3). Define $Y := \{p_t, \neg p_{t+1}\} \cup Y_t^*$, $Z_1 := \{p_t\}$, and $Z_2 := \{\neg p_{t+1}\}$. Since $\{p_t, \neg p_{t+1}\}$ is consistent, $\{p_t, \neg p_{t+1}\} \cup B$ is consistent for some set B that contains q or $\neg q$ (but not both) for each $q \in Y_t^*$ (by L3 together with L1,L2). Note that there exists a $Z_3 \subseteq Y_t^*$ with $B = (Y_t^* \backslash Z_3) \cup Z_3^{\neg}$. This proves the claim, since:

- $Y = \{p_t, \neg p_{t+1}\} \cup Y_t^*$ is inconsistent by (3),
- Z_1, Z_2, Z_3 are pairwise disjoint subsets of Y,
- $(Y \backslash Z_1) \cup Z_1^{\neg} = (Y \backslash \{p_t\}) \cup \{\neg p_t\} = \{\neg p_t, \neg p_{t+1}\} \cup Y_t^*$ is consistent by (4),
- $(Y \backslash Z_2) \cup Z_2^{\neg} = (Y \backslash \{\neg p_{t+1}\}) \cup \{p_{t+1}\} = \{p_t, p_{t+1}\} \cup Y_t^*$ is consistent by (4),
- $(Y \backslash Z_3) \cup Z_3^{\neg} = \{p_t, \neg p_{t+1}\} \cup (Y_t^* \backslash Z_3) \cup Z_3^{\neg} = \{p_t, \neg p_{t+1}\} \cup B$ is consistent.

Claim 5. For any coalitions $C, C^* \subseteq N$, if $C, C^* \in \mathcal{W}$ then $C \cap C^* \in \mathcal{W}$.

Consider any $C, C^* \in \mathcal{W}$, and assume for contradiction that $C_1 := C \cap C^* \notin \mathcal{W}$. Put $C_2 := C^* \backslash C$ and $C_3 := N \backslash C^*$. Let Y, Z_1, Z_2, Z_3 be as in claim 4. Noting that C_1, C_2, C_3 form a partition of N, we define the profile $(A_1, ..., A_n)$ by:

$$A_i := \begin{cases} A_{(Y \backslash Z_1) \cup Z_1^-} & \text{if } i \in C_1 \\ A_{(Y \backslash Z_2) \cup Z_2^-} & \text{if } i \in C_2 \\ A_{(Y \backslash Z_3) \cup Z_3^-} & \text{if } i \in C_3. \end{cases}$$

By $C_1 \notin \mathcal{W}$ and $N \backslash C_1 = C_2 \cup C_3$ we have $C_2 \cup C_3 \in \mathcal{W}$ by claim 3, and so $Z_1 \subseteq F(A_1, ..., A_n)$. By $C \in \mathcal{W}$ and $C \subseteq C_1 \cup C_3$ we have $C_1 \cup C_3 \in \mathcal{W}$ by claim 3, and so $Z_2 \subseteq F(A_1, ..., A_n)$. Further, $Z_3 \subseteq F(A_1, ..., A_n)$ as $C_1 \cup C_2 = C^* \in \mathcal{W}$. Finally, $Y \backslash (Z_1 \cup Z_2 \cup Z_3) \subseteq F(A_1, ..., A_n)$ as $N \in \mathcal{W}$ by claim 3. In summary, we have $Y \subseteq F(A_1, ..., A_n)$, violating consistency.

Claim 6. There is a dictator.

Consider the intersection of all winning coalitions, $\widetilde{C} := \cap_{C \in \mathcal{W}} C$. By claim 5, $\widetilde{C} \in \mathcal{W}$. So $\widetilde{C} \neq \emptyset$, as by claim 3, $\emptyset \notin \mathcal{W}$. Hence there is a $j \in \widetilde{C}$. To show that j is a dictator, consider any $(A_1, ..., A_n) \in Domain(F)$ and $p \in X$, and let us prove that $p \in F(A_1, ..., A_n)$ if and only if $p \in A_j$. If $p \in F(A_1, ..., A_n)$ then $C := \{i : p \in A_i\} \in \mathcal{W}$, whence $j \in C$ (as j belongs to every winning coalition), i.e., $p \in A_j$. Conversely, if $p \notin F(A_1, ..., A_n)$, then $\neg p \in F(A_1, ..., A_n)$; so by an argument analogous to the previous one, $\neg p \in A_j$, whence $p \notin A_j$. ■

Proof of Theorem 4. Let $Y \subseteq X$.

(i) First, assume F is strategy-proof for C_Y. To show non-manipulability on Y, consider any proposition $p \in Y$, individual i, and profile $(A_1, ..., A_n) \in Domain(F)$, such that $F(A_1, ..., A_n)$ disagrees with A_i on p. Let $(A_1, ..., A_i^*, ..., A_n) \in Domain(F)$ be any i-variant. We have to show that $F(A_1, ..., A_i^*, ..., A_n)$ still disagrees with A_i on p. Define a preference relation $\succsim_i$ over judgment sets by [$B \succsim_i B^*$ if and only if A_i agrees on p with B but not with B^*, *or* with both B and B^*, *or* with neither B nor B^*]. ($\succsim_i$ is interpreted as individual i's preference relation in case i cares only about p.) It follows immediately that $\succsim_i$ is reflexive and transitive and respects closeness to A_i on Y, i.e., is a member of $C_Y(A_i)$. So, by strategy-proofness for C_Y, $F(A_1, ..., A_n) \succsim_i F(A_1, ..., A_i^*, ..., A_n)$. Since A_i disagrees with $F(A_1, ..., A_n)$ on p, the definition of $\succsim_i$ implies that A_i still disagrees with $F(A_1, ..., A_i^*, ..., A_n)$ on p.

(ii) Now assume that F is non-manipulable on Y. To show strategy-proofness for C_Y, consider any individual i, profile $(A_1, ..., A_n) \in Domain(F)$, and preference relation $\succsim_i \in C_Y(A_i)$, and let $(A_1, ..., A_i^*, ..., A_n) \in Domain(F)$ be any i-variant. We have to prove that $F(A_1, ..., A_n) \succsim_i F(A_1, ..., A_i^*, ..., A_n)$. By non-manipulability on Y, for every proposition $p \in Y$, if A_i disagrees with $F(A_1, ..., A_n)$ on p, then also with $F(A_1, ..., A_i^*, ..., A_n)$; in other words, if A_i agrees with $F(A_1, ..., A_i^*, ..., A_n)$ on p, then also with $F(A_1, ..., A_n)$. So $F(A_1, ..., A_n)$ is at least as close to A_i on Y as $F(A_1, ..., A_i^*, ..., A_n)$. Hence $F(A_1, ..., A_n) \succsim_i F(A_1, ..., A_i^*, ..., A_n)$, as $\succsim_i \in C_Y(A_i)$. ■

Proof of Proposition 1. We prove this result directly, although it can also be derived from Corollary 1. Let F be premise-based voting. To show that F is strategy-proof for $C_{Y_{reason}}$, consider any individual i, profile $(A_1, \ldots, A_n) \in Domain(F)$, i-variant $(A_1, \ldots, A_i^*, \ldots, A_n) \in Domain(F)$, and preference relation $\succsim_i \in C_{Y_{reason}}(A_i)$. The definition of premise-based voting implies that $F(A_1, \ldots, A_n)$ is at least as close

to A_i as $F(A_1, \ldots, A_i^*, \ldots, A_n)$ on Y_{reason}. So, by $\succsim_i \in C_{Y_{reason}}(A_i)$, we have $F(A_1, \ldots, A_n) \succsim_i F(A_1, \ldots, A_i^*, \ldots, A_n)$. ■

Proof of Theorem 5. Consider the conjunctive agenda (the proof is analogous for disjunctive agendas). Let F and G be premise- and conclusion-based voting, respectively. Take any profile $(A_1, ..., A_n) \in Domain(F) = Domain(G)$ and any preference relations $\succsim_1 \in C_{Y_{outcome}}(A_1)$, ..., $\succsim_n \in C_{Y_{outcome}}(A_n)$. Define $(B_1, \ldots, B_n)$ by

$$B_i = \begin{cases} \{\neg a_1, \ldots, \neg a_k, c \leftrightarrow (a_1 \wedge \ldots \wedge a_k), \neg c\} & \text{if } \neg c \in A_i, \\ \{a_1, \ldots, a_k, c \leftrightarrow (a_1 \wedge \ldots \wedge a_k), c\} & \text{if } c \in A_i. \end{cases}$$

It can easily be seen that, for each i and any pair of i-variants $(D_1, \ldots, B_i, \ldots, D_n)$, $(D_1, \ldots, B_i^*, \ldots, D_n) \in Domain(F)$, $F(D_1, \ldots, B_i, \ldots, D_n)$ is at least as close to A_i on $Y_{outcome}$ $(= \{c, \neg c\})$ as $F(D_1, \ldots, B_i^*, \ldots, D_n)$; so $(D_1, \ldots, B_i, \ldots, D_n) \succsim_i (D_1, \ldots, B_i^*, \ldots, D_n)$ as $\succsim_i \in C_{Y_{outcome}}(A_i)$. Hence, submitting B_i is a weakly dominant strategy for each i under F. Second, let $(C_1, \ldots, C_n)$ be $(A_1, ..., A_n)$ (the truthful profile). Then, for each i, submitting C_i is a weakly dominant strategy under G, as G is strategy-proof. Finally, it can easily be seen that $F(B_1, \ldots, B_n)$ and $G(C_1, \ldots, C_n) = G(A_1, ..., A_n)$ agree on each proposition in $Y_{outcome} = \{c, \neg c\}$. ■

Chapter 5

Agenda manipulation

Paper: Judgment aggregation: (im)possibility theorems, *Journal of Economic Theory* 126(1): 286-298, 2006

Judgment aggregation: (im)possibility theorems

Franz Dietrich[1]

The aggregation of individual judgments over interrelated propositions is a newly arising field of social choice theory. I introduce several independence conditions on judgment aggregation rules, each of which protects against a specific type of manipulation by agenda setters or voters. I derive impossibility theorems whereby these independence conditions are incompatible with certain minimal requirements. Unlike earlier impossibility results, the main result here holds for any (non-trivial) agenda. However, independence conditions arguably undermine the logical structure of judgment aggregation. I therefore suggest restricting independence to "premises", which leads to a generalised premise-based procedure. This procedure is proven to be possible if the premises are logically independent. *Journal of Economic Literature Classification Numbers*: D70, D71, D79

Key words: judgment aggregation, formal logic, collective inconsistency, manipulation, impossibility theorems, premise-based procedure, possibility theorems

1 Introduction

While the more traditional discipline in social choice theory, preference aggregation, aims to merge individual *preference orderings* over a set of *alternatives*, judgment aggregation aims to merge individual *(yes/no-)judgments* over a set of interrelated *propositions* (expressed in formal logic). Suppose for instance that a cabinet has to reach a judgment about the following three propositions. a : "we can afford a budget deficit", b : "spending on education should be raised", and $a \rightarrow b$: "*if* we can afford a budget deficit *then* spending on education should be raised". The cabinet is split into three camps of equal size. Ministers of the first camp accept all three propositions. The two other camps both reject b, but disagree on the *reason* for rejecting b: the second camp accepts a but rejects $a \rightarrow b$, and the third camp accepts $a \rightarrow b$ but rejects a. So, although a 2/3 majority of the ministers rejects b, 2/3 majorities accept each premise a and $a \rightarrow b$. Should the cabinet reject b, or rather accept b on the grounds of accepting both premises of b?

Such collective inconsistencies arise not just for the particular rule of propositionwise majority voting, and not just for the mentioned agenda. List and Pettit [7,8] prove a first formal impossibility theorem for judgment aggregation, recently complemented by Pauly and van Hees' [9] powerful results. List [4,5,6] and Bovens and Rabinowicz [1] derive possibility results. For discussions of judgment aggregation, e.g. Brennan [2] and Chapman [3].

At the heart of the existing impossibility theorems is the requirement of *propositionwise aggregation* or *independence*, an analogue of Arrow's independence of irrelevant alternatives. Is it justified to impose independence on a judgment aggregation rule? I first introduce a family of new independence conditions, and show that each of them protects against a particular type of manipulation. Second, I prove impossibility theorems for these independence

[1] I would like to thank Christian List, Marc Pauly and Martin van Hees for inspiring comments on previous versions of the paper. I also thank the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government, for supporting this research. I have presented this paper at the workshop *Judgment Aggregation and the Discursive Dilemma*, 18-19 June 2004, University of Konstanz.

conditions. One novelty is that the main impossibility theorem applies to *all* (non-trivial) agendas, and hence to a wide range of real situations. Finally, to make premise-based collective decision-making possible, I suggest restricting the independence requirement to a set of "premises", and prove a characterisation theorem for the so-called *premise-based procedure*.

2 The basic model

Let there be a group of individuals, labelled 1, 2, ..., n ($n \geq 2$), having to make collective judgments on a set of propositions X, the *agenda*. Specifically, consider a set of propositional symbols $a, b, c, ...$ (representing non-decomposable sentences such as a and b in the above example), and define the set of *all* propositions, $\mathcal{L}$, as the (smallest) set such that

- $\mathcal{L}$ contains all propositional symbols, called *atomic propositions*;
- if $\mathcal{L}$ contains p and q, then $\mathcal{L}$ also contains $\neg p$ ("not p"), $(p \wedge q)$ ("p and q"), $(p \vee q)$ ("p or q"), $(p \rightarrow q)$ ("p implies q") and $(p \leftrightarrow q)$ ("p if and only if q").

For ease of notation, I drop the external ()-brackets around propositions, e.g. I write $a \wedge (b \rightarrow c)$ for $(a \wedge (b \rightarrow c))$. A *truth-value assignment* is a function assigning the value "true" or "false" to each proposition in $\mathcal{L}$, with the standard consistency properties.[2] A set $A \subseteq \mathcal{L}$ is *(logically) consistent/inconsistent* if there exists a/no truth-value assignment that assigns "true" to each $p \in A$. Finally, for $A \subseteq \mathcal{L}$ and $p \in \mathcal{L}$, A *(logically) entails* p, written $A \vDash p$, if $A \cup \{\neg p\}$ is inconsistent.

Now, the agenda X is a non-empty subset of $\mathcal{L}$, where by assumption X contains no double-negated propositions ($\neg\neg p$), and X consists of proposition-negation pairs in the following sense: if $p \in X$, then also $\sim p \in X$, where

$$\sim p := \begin{cases} \neg p & \text{if } p \text{ is not itself a negated proposition,} \\ q & \text{if } p \text{ is the negated proposition } \neg q. \end{cases}$$

The example had $X = \{a, b, a \rightarrow b,$negations$\}$ ("negations" stands for "$\neg a, \neg b, \neg(a \rightarrow b)$").

A *judgment set* (held by an individual or the collective) is a subset $A \subseteq X$, where $p \in A$ means "acceptance of proposition p". I consider two rationality conditions on judgment sets A: *consistency* (see above) and *completeness* (i.e., for every $p \in X$, $p \in A$ or $\sim p \in A$). (Together they imply List's "deductive closure" condition.) For instance, for the above agenda, the judgment set $A = \emptyset$ is consistent but incomplete, the judgment set $A = \{a, a \rightarrow b, \neg b\}$ is complete but inconsistent, and the judgment set $A = \{a, a \rightarrow b, b\}$ is consistent and complete. Let $\mathbf{A}$ be the set of all consistent and complete judgment sets.

A *profile (of individual judgment sets)* is an n-tuple $(A_1, ..., A_n)$. A *(judgment) aggregation rule* is a function, F, assigning to each admissible profile $(A_1, ..., A_n)$ a (collective) judgment set $F(A_1, ..., A_n) = A \subseteq X$; the set of admissible profiles is called the domain of F, denoted $Dom(F)$. For instance, propositionwise majority voting (with universal domain $\mathbf{A}^n$) is the aggregation rule F such that, for each profile $(A_1, ..., A_n) \in \mathbf{A}^n$, $F(A_1, ..., A_n)$ contains each proposition $p \in X$ if and only if more individuals i have $p \in A_i$ than $p \notin A_i$; as seen above, $F(A_1, ..., A_n)$ may then be inconsistent, hence not in $\mathbf{A}$.

[2] Specifically, for any $p, q \in \mathcal{L}$, $\neg p$ is true if and only if p is false; $p \wedge q$ is true if and only if both p and q are true; $p \vee q$ is true if and only if p or q is true; $p \rightarrow q$ is true if and only if q is true or p is false; $p \leftrightarrow q$ is true if and only if p and q are both true or both false.

3 Collective judgments are sensitive to the agenda choice: examples of agenda manipulation

Collective judgments are highly sensitive to reformulations of the agenda, as some examples will demonstrate. An "agenda manipulation" is the modification of the agenda by the agenda setter in order to affect the collective judgments on certain propositions.

The sensitivity to the agenda choice. Consider again the above cabinet of ministers split into three camps, where a is "we can afford a budget deficit" and b is "spending on education should be raised". Many different specifications of the agenda X are imaginable. Assuming that the collective judgment set A is formed by propositionwise majority voting,

(a) the agenda $X = \{a, b, \text{negations}\}$ leads to $A = \{a, \neg b\}$,

(b) the agenda $X = \{a, a \rightarrow b, \text{negations}\}$ leads to $A = \{a, a \rightarrow b\}$,

(c) the agenda $X = \{a, a \rightarrow b, b, \text{negations}\}$ leads to $A = \{a, a \rightarrow b, \neg b\}$ (collective inconsistency).

While in (a) the collective judgment set contains b, in (b) it logically entails $\neg b$, and in (c) it is inconsistent.

General agenda manipulation. Assume the original (non-manipulated) agenda is that in (a). An agenda setter who thinks spending on education should be raised can reverse the rejection of b by using the agenda in (b) instead.

Logical agenda manipulation. Note that the manipulated agenda in (b) need not settle b: it may lead to a collective judgment set of $\{\neg a, a \rightarrow b\}$, which entails neither b nor $\neg b$, hence entails no decision about spending on education. The agenda setter may not have the power to manipulate the agenda to the extent of possibly not settling b. Then he can achieve acceptance of b by using the agenda $X^* = \{a, a \leftrightarrow b, \text{negations}\}$, which settles b whatever the (complete and consistent) collective judgment set. Formally, I say that b belongs to the *scope* of X^*.

Definition 1 *A set $A \subseteq \mathcal{L}$ "settles" a proposition $p \in \mathcal{L}$ if $A \vDash p$ or $A \vDash \neg p$. The "scope" or "extended agenda" of an agenda X is the set $\overline{X}$ of propositions $p \in \mathcal{L}$ settled by each (consistent and complete) judgment set $A \in \mathbf{A}$.*

For instance, the scope of $X = \{a, b, \text{negations}\}$ contains the propositions $b, a \vee b, a \rightarrow (\neg b)$, etc. In general, how much larger than X is $\overline{X}$? The scope $\overline{X}$ is the (infinite) set of all (arbitrarily complex) propositions constructible from propositions X using logical operations $(\neg, \wedge, \vee, \rightarrow, \leftrightarrow)$, as well as all propositions logically equivalent to such propositions.

I call an agenda manipulation of $X \subseteq \mathcal{L}$ into $X^* \subseteq \mathcal{L}$ *logical* if it preserves the scope, i.e. if $\overline{X} = \overline{X^*}$, or equivalently $X \subseteq \overline{X^*}$ and $X^* \subseteq \overline{X}$. Logical agenda manipulation, which has a wide range of examples[3], might appear to be a mild form of manipulation, as it merely frames the same decision problem in different logical terms: X and X^* are equivalent in that any (complete and consistent) judgment set for X entails one for X^*, and vice versa. Yet X and X^* may reverse collective judgments on certain propositions, as demonstrated above.

[3] For instance: (1) adding or removing propositions settled by the other propositions, e.g. modifying $\{a, b, \text{negations}\}$ into $\{a, b, a \wedge b, \text{negations}\}$, or vice versa; (2) replacing a proposition by one logically equivalent to it or to its negation, unconditionally or given judgments on the other proposition(s), e.g. modifying $\{a, b, \text{negations}\}$ into $\{a, b \leftrightarrow a, \text{negations}\}$; (3) replacing X by its set of (possibly negated) "states of the world" $\{\wedge_{p \in A} p, \neg \wedge_{p \in A} p | A \in \mathbf{A}\}$, e.g. modifying $\{a, b, \text{negations}\}$ into $\{a \wedge b, (\neg a) \wedge b, a \wedge (\neg b), (\neg a) \wedge (\neg b), \text{negations}\}$.

4 Independence conditions to prevent manipulation by agenda setters and voters

Different independence conditions, all of the following general form, may each prevent a specific type of manipulation by agenda setters or voters. Consider any subset $Y \subseteq \overline{X}$.

Independence on Y (I_Y). For every proposition $p \in Y$ and every two profiles $(A_1, ..., A_n)$, $(A'_1, ..., A'_n) \in Dom(F)$, if [for every person i, $A_i \vDash p$ if and only if $A'_i \vDash p$] then [$F(A_1, ..., A_n) \vDash p$ if and only if $F(A'_1, ..., A'_n) \vDash p$].

Here, I interpret "$A_i \vDash p$" and "$F(A_1, ..., A_n) \vDash p$" as "acceptance of p", even when this acceptance is not expressed explicitly (i.e. no "$\ni p$") but only entailed logically. Note that $A \vDash p$ is equivalent to $p \in A$ if $p \in X$ and $A \in \mathbf{A}$. If $Y \subseteq Y^*$ ($\subseteq \overline{X}$), then (I_{Y^*}) implies (I_Y).

Condition (I_Y) prescribes propositionwise aggregation for each proposition in Y. To make this precise, following Pauly and van Hees [9] I define a *(propositionwise) decision method* as a mapping $M : \{0,1\}^n \to \{0,1\}$, taking vectors $(t_1, ..., t_n)$ of (individual) truth values to single (collective) truth values $M(t_1, ..., t_n)$ (where 0/1 stands for rejection/acceptance of a given proposition). For instance, the *absolute majority method* M is defined by [$M(t_1, ..., t_n) = 1$ if and only if $t_1 + ... + t_n > n/2$], and the *unanimity method* by [$M(t_1, ..., t_n) = 1$ if and only if $t_1 = ... = t_n = 1$]. I say that F "applies decision method M for p" if, for every profile $(A_1, ..., A_n) \in Dom(F)$, we have $t = M(t_1, ..., t_n)$, where $t_1, ..., t_n$ and t are the individual and collective truth values of p (i.e., t_i is 1 if $A_i \vDash p$ and 0 else, and t is 1 if $F(A_1, ..., A_n) \vDash p$ and 0 else). The following characterisation of (I_Y) is obvious.

Proposition 1 *Let $Y \subseteq \overline{X}$. Then F is independent on Y (I_Y) if and only if, for each proposition $p \in Y$, F applies some decision method M_p for p.*

Preventing agenda manipulation. Consider the following special cases of (I_Y).

Definition 2 *Independence on Y (I_Y) is called "independence" if $Y = X$, "strong independence" if $Y = \overline{X}$, and "independence on states of the world" if $Y = \widetilde{X} := \{\wedge_{p \in A} p : A \in \mathbf{A}\}$.*

Independence (I_X) is equivalent to Pauly and van Hees' independence condition if all (individual or collective) judgment sets belong to $\mathbf{A}$, as required in all present and previous impossibility theorems.[4] I call $\wedge_{p \in A} p$ ($A \in \mathbf{A}$) a *state of the world* since it is the conjunction of all propositions of a complete and consistent judgment set.[5] States of the world are maximally fine-grained descriptions of the world (relative to X). For instance, if the agenda is $X = \{a, b, \text{negations}\}$ then $\widetilde{X} = \{a \wedge b, (\neg a) \wedge b, a \wedge (\neg b), (\neg a) \wedge (\neg b)\}$.

To state the merits of these conditions, I say that an agenda manipulation "*reverses*" the decision about a proposition p if the old agenda leads to a consistent collective judgment set entailing p and the new agenda leads to one entailing $\neg p$, or vice versa.

[4] Specifically, Pauly and van Hees require that, for every $p \in Y$ and $(A_1, ..., A_n), (A'_1, ..., A'_n) \in Dom(F)$, if [for every person i, $p \in A_i$ if and only if $p \in A'_i$] then [$p \in F(A_1, ..., A_n)$ if and only if $p \in F(A'_1, ..., A'_n)$]. This condition is equivalent to (I_X) if all judgment sets accepted or generated by F are in $\mathbf{A}$, because [$p \in A$ if and only if $A \vDash p$] for all $p \in X$ and $A \in \mathbf{A}$.

[5] For *infinite* X, the conjunction $\wedge_{p \in A} p$ is one over an infinite set of propositions, hence not part of the language, so not part of the scope $\overline{X}$. However, as each judgment set in $\mathbf{A}$ settles each $\wedge_{p \in A} p$, $A \in \mathbf{A}$, states of the world are part of the scope formed in an extended language that allows conjunctions over infinite sets of propositions of the cardinality (size) of X (e.g. countably infinite conjunctions if X is countably infinite). So condition ($\mathrm{I}_{\widetilde{X}}$) may be considered even for infinite agendas X.

Claim A. By imposing independence, the decisions on propositions $p \in X$ cannot be reversed by adding or removing propositions in X *other than* p.
Claim B. By imposing independence on states of the world, the decisions on propositions $p \in \overline{X}$ cannot be reversed by *logical* agenda manipulation.
Claim C. By imposing strong independence, the decisions on propositions $p \in \overline{X}$ cannot be reversed by *any* form of agenda manipulation.
Claim D. If F violates independence (resp. strong independence), then for some profile in $Dom(F)$ the decision on some proposition $p \in X$ (resp. $p \in \overline{X}$) can be reversed by an agenda manipulation of the type in claim A (resp. C).

These claims rest on the following assumptions:

(1) For any agenda, each individual i holds a consistent and complete judgment set, and i's judgment sets for two agendas are consistent with each other.

(2) For any agenda, collective judgment sets have to be consistent and complete.

(3) For any agendas X and X^* with corresponding aggregation rules F resp. F^* on which (I_Y) resp. (I_{Y^*}) is imposed, and each proposition $p \in Y \cap Y^*$, F and F^* apply the *same* decision method M_p for p. (*Interpretation*: M_p is chosen independently of the other propositions in the agenda, e.g. M_p is prescribed by law or is "intrinsically adequate" for p).

Proof of claim A [assuming (1)-(3)]. Suppose independence is imposed. Let $p \in X$ and consider a (manipulated) agenda X^* with $p \in X$. For the two agendas, by (1) the individual truth values of p stay the same, and by (I_X)/(I_{X^*}) and (3) the decision method M_p applied for p stays the same. Hence the collective truth value of p stays the same. ■

Proof of claim B [assuming (1)-(3)]. Suppose independence on states of the world is imposed. Let $p \in \overline{X}$ and consider a (manipulated) agenda X^* with $\overline{X} = \overline{X^*}$. For simplicity, assume X and X^* are both finite (but the proof could be generalised). Then $\widetilde{X}$ and $\widetilde{X^*}$ each contains, up to logical equivalence, all atoms (i.e. maximally consistent members) of $\overline{X} = \overline{X^*}$. Let r be any atom of $\overline{X} = \overline{X^*}$. For the two agendas, by (1) the individual truth values of r stay the same, and by ($\mathrm{I}_{\widetilde{X}}$)/($\mathrm{I}_{\widetilde{X^*}}$) and (3) the decision method applied for r stays the same. So the collective truth value of r stays the same. Since p is equivalent to a disjunction of atoms r of $\overline{X} = \overline{X^*}$, the collective truth value of p follows from those of the atoms r of $\overline{X} = \overline{X^*}$ (by using (2)). So the collective truth value of p stays the same. ■

Proof of claim C [assuming (1)-(3) and the monotonicity condition (4) below]. Now impose strong independence. Let $p \in \overline{X}$ and consider *any* (manipulated) agenda X^*. First let $p \in \overline{X^*}$. Then for both agendas, by (1) the individual truth values of p stay the same, and by ($\mathrm{I}_{\overline{X}}$)/($\mathrm{I}_{\overline{X^*}}$) and (3) the same decision method M_p is applied for p. So the collective truth value of p stays the same. Now let $p \notin \overline{X^*}$. Suppose the agendas X and X^* result in the collective judgment sets A resp. A^*. To show that the collective judgment on p is not reversed, it is (by (2)) sufficient to show that $A^* \vDash p$ implies $A \vDash p$, and $A^* \vDash \neg p$ implies $A \vDash \neg p$. I only show the former, as the proof of the latter is analogous. So, let $A^* \vDash p$. It is plausible that decision methods are chosen as monotonic both in truth values and in propositions:

(4) If decision method M_q is applied for q by all aggregation rules on which (I_Y) is imposed for some Y containing q, then, for fixed q, [$t_i \leq t_i^*$ for all i implies $M_q(t_1, ..., t_n) \leq M_q(t_1^*, ..., t_n^*)$], and, for fixed $t_1, ..., t_n$, [$q^* \vDash q$ implies $M_{q^*}(t_1, ..., t_n) \leq M_q(t_1, ..., t_n)$].

Take any $p^* \in \overline{X^*}$ with $A^* \vDash p^*$ and $p^* \vDash p$ (e.g. $p^* = \wedge_{q \in A^*} q$). For agendas X (X^*), let M_p (M_{p^*}) be the decision method applied for p (p^*), and t_i (t_i^*) i's truth values of p

(p^*). By $p^* \vDash p$ and (1), we have $t_i^* \leq t_i$ for all i. Since also $p^* \vDash p$, by (4) $M_{p^*}(t_1^*, ..., t_n^*) \leq M_p(t_1, ...t_{,n})$. By $A^* \vDash p^*$ we have $M_{p^*}(t_1^*, ..., t_n^*) = 1$, so $M_p(t_1, ..., t_n) = 1$, so $A \vDash p$. ■

Proof of claim D. [assuming (1),(2)]. Suppose F violates (I_X) (the proof for $(\mathrm{I}_{\overline{X}})$ is analogous). So there are two profiles in $Dom(F)$ with identical individual but opposed collective judgments about some $p \in X$. So, using the agenda $X^* := \{p, \sim p\}$ instead of X reverses the collective judgment for one of the two mentioned profiles. ■

Preventing manipulation by voters. Assume that it is desirable that no person i can, by submitting a false judgment set, reverse in his/her favour the collective judgment about any given proposition in Y $(\subseteq \overline{X})$. Generalising Dietrich and List's[6] definition of "strategy-proofness on Y" to subsets $Y \subseteq \overline{X}$ (rather than $Y \subseteq X$), [7] one may easily prove a result analogous to their Theorem 1:

• If F is independent on Y and monotonic on Y then F is strategy-proof on Y, and the converse implication also holds in case F has universal domain.

(Monotonicity on Y and universal domain are defined below). So, independence on Y (I_Y) is crucial for strategy-proofness on Y: (I_Y) is together with monotonicity on Y sufficient, and under universal domain also necessary for strategy-proofness on Y.

5 Impossibility theorems for judgment aggregation

I now prove that each independence condition is incompatible with seemingly minimal requirements on F. However, the impossibility for independence (I_X) holds only for special agendas.

First, individual judgments are left unrestricted subject to the rationality constraint of consistency and completeness, and collective judgments have to be equally rational:

Universal Domain (U). The domain of F, $Dom(F)$, is the set $\mathbf{A}^n = \mathbf{A} \times ... \times \mathbf{A}$ of all logically possible profiles of complete and consistent individual judgment sets.

Collective Rationality (C). For any profile $(A_1, ..., A_n) \in Dom(F)$, $F(A_1, ..., A_n) \in \mathbf{A}$.

Recently inspired by Pauly and van Hees' [9] findings, I realised that a unanimity principle (as in Arrow's Theorem) is not necessary for my theorem; I can replace it by:

Weak Responsiveness (R). There exist two profiles $(A_1, ..., A_n), (A'_1, ..., A'_n) \in Dom(F)$ such that $F(A_1, ..., A_n) \neq F(A'_1, ..., A'_n)$.

Propositions p, q are "*in trivial dependence*" if p is logically equivalent to q or to $\neg q$, *or* p or q is a tautology or a contradiction. An aggregation rule F with universal domain is *dictatorial* if for some person j (a "dictator") $F(A_1, ..., A_n) = A_j$ for all profiles $(A_1, ..., A_n) \in \mathbf{A}^n$.

Theorem 1 *If X contains at least two propositions (not in trivial dependence), then an aggregation rule F is independent on states of the world and weakly responsive (and satisfies universal domain and collective rationality) if and only if F is dictatorial.*

[6] F. Dietrich and C. List, Strategy-Proof Judgment Aggregation, unpublished paper, Konstanz Univ., 2004.

[7] More precisely, I call F *strategy-proof on* Y $(\subseteq \overline{X})$ if, for every person i, profile $(A_1, \ldots, A_n) \in Dom(F)$ and proposition $p \in Y$, if A_i disagrees with $F(A_1, \ldots, A_n)$ on p (i.e. $A_i \vDash p$ if and only if $F(A_1, \ldots, A_n) \nvDash p$), then A_i still disagrees with $F(A_1, \ldots, A_i^*, \ldots, A_n)$ on p for every i-variant $(A_1, \ldots, A_i^*, \ldots, A_n) \in Dom(F)$. A game-theoretic justification for this definition may be given along the lines of Dietrich and List's analysis.

As independence on states of the world implies strong independence, we have:

Corollary 1 *If X contains at least two propositions (not in trivial dependence), then an aggregation rule F is strongly independent and weakly responsive (and satisfies universal domain and collective rationality) if and only if F is dictatorial.*

So, for non-trivial agendas, every aggregation rule must of necessity *either* be dictatorial, *or* be vulnerable to manipulation (see Section 4), *or* always generate the same judgment set, *or* sometimes generate no or an inconsistent or incomplete judgment set.

The proof of Theorem 1 relies on three lemmata, to be proven first.

Lemma 1 *Assume (U) and (C). Then ($I_{\widetilde{X}}$) holds if and only if, for every $A \in \mathbf{A}$ and $(A_1, ..., A_n), (A'_1, ..., A'_n) \in Dom(F)$, if [for every person i, $A_i = A$ if and only if $A'_i = A$] then [$F(A_1, ..., A_n) = A$ if and only if $F(A'_1, ..., A'_n) = A$].*

Proof. Obvious, as a judgment set in $\mathbf{A}$ entails $\wedge_{p \in A} p$ ($\in \widetilde{X}$) just in case it equals A. ■

Judgment-Set Monotonicity (JM). For any person j and any j-variants $(A_1, ..., A, ..., A_n)$, $(A_1, ..., A', ..., A_n) \in Dom(F)$, if $F(A_1, ..., A, ..., A_n) = A'$ then $F(A_1, ..., A', ..., A_n) = A'$.

Lemma 2 *Let X contain at least two propositions (not in trivial dependence). If F satisfies (U), (C) and ($I_{\widetilde{X}}$), then F satisfies (JM).*

Proof. Let X be as specified, and suppose (U), (C) and ($\mathrm{I}_{\widetilde{X}}$). To show (JM), let j be a person and $(..., A, ...), (..., A', ...) \in Dom(F)$ be j-variants, where "..." denotes the other persons' votes. Assume for contradiction that $F(..., A, ...) = A'$ but $F(..., A', ...) \neq A'$. In $(..., A, ...)$ and $(..., A', ...)$ exactly the same persons endorse each $A'' \in \mathbf{A} \backslash \{A, A'\}$; hence, as $F(..., A, ...) \neq A''$, we have $F(..., A', ...) \neq A''$ by Lemma 1, so $F(..., A', ...) \in \{A, A'\}$, hence $F(..., A', ...) = A$. By $|\mathbf{A}| \geq 3$ there exists an $A'' \in \mathbf{A} \backslash \{A, A'\}$. Consider the new j-variant $(..., A'', ...)$. I apply twice Lemma 1, with contradictory implications: as $F(..., A, ...) = A'$ and as in $(..., A, ...)$ and $(..., A'', ...)$ exactly the same persons endorse A' (in neither profile person j), $F(..., A'', ...) = A'$; but, as $F(..., A', ...) = A$ and as in $(..., A', ...)$ and $(..., A'', ...)$, exactly the same persons endorse A (in neither profile person j), $F(..., A'', ...) = A$. ■

Judgment-Set Unanimity Principle (JUP). $F(A, ..., A) = A$ for all $(A, ..., A) \in Dom(F)$.

Lemma 3 *Let X contain at least two propositions (not in trivial dependence). If F satisfies (U), (C), ($I_{\widetilde{X}}$) and (R), then F satisfies (JUP).*

Proof. Let X be as specified, and assume (U), (C), ($\mathrm{I}_{\widetilde{X}}$) and (R). To show (JUP), consider any $A \in \mathbf{A}$, and suppose for contradiction that $F(A, ..., A) \neq A$. I show that $F(A'_1, ..., A'_n) = F(A, ..., A)$ for *all* $(A'_1, ..., A'_n) \in \mathbf{A}^n$, violating (R). Take any $(A'_1, ..., A'_n) \in \mathbf{A}^n$ and write $A' := F(A'_1, ..., A'_n)$. By (JM) (see Lemma 2), if the votes $A'_1, ..., A'_n$ are replaced one by one by A', the decision remains A', and so $F(A', ..., A') = A'$. In $(A', ..., A')$ and $(A, ..., A)$ exactly the same persons (namely nobody) endorse each $A'' \in \mathbf{A} \backslash \{A, A'\}$; hence, as $F(A', ..., A') \neq A''$, we have $F(A, ..., A) \neq A''$ (see Lemma 1). So $F(A, ..., A) \in \{A, A'\}$. As $F(A, ..., A) \neq A$, we have $F(A, ..., A) = A'$, i.e. $F(A, ..., A) = F(A'_1, ..., A'_n)$, as claimed. ■

Proof of Theorem 1. Let X be as specified. If F is dictatorial then F obviously satisfies all of (U), (C), ($\mathrm{I}_{\widetilde{X}}$) and (R). Now I assume (U), (C), ($\mathrm{I}_{\widetilde{X}}$) and (R), and show that there is a dictator. By Lemmata 2 and 3 we have (JM) and (JUP).

1. *A simple algorithm.* As $|X| \geq 3$, there exist three distinct $A, A', A'' \in \mathbf{A}$. By (JUP), $F(A, ..., A) = A$. Modify $(A, ..., A)$ step by step as follows. Starting with person $i = 1$, (i) substitute i's vote A by A'. If the collective outcome is not anymore A, stop here. Otherwise, (ii) substitute i's vote A' by A'', which by Lemma 1 leaves the outcome again at A, and do the same substitution procedure with person $i + 1$ (unless $i = n$). There exists a person j for whom the vote substitution in (i) alters the outcome (thus terminating the algorithm), since otherwise one would end up with $F(A'', ..., A'') = A$, violating (JUP).

2. *j is a dictator for A'.* I write profiles by underlining j's vote. In the profiles before and after replacing j's vote, $(A'', ..., A'', \underline{A}, A, ..., A)$ and $(A'', ..., A'', \underline{A'}, A, ..., A)$, exactly the same persons endorse each $A^* \in \mathbf{A}\backslash\{A, A'\}$; hence, as $F(A'', ..., A'', \underline{A}, A, ..., A) \neq A^*$, we have $F(A'', ..., A'', \underline{A'}, A, ..., A) \neq A^*$ (see Lemma 1). So $F(A'', ..., A'', \underline{A'}, A, ..., A) \in \{A, A'\}$. As $F(A'', ..., A'', \underline{A'}, A, ..., A) \neq A$, we have $F(A'', ..., A'', \underline{A'}, A, ..., A) = A'$, although here j is the only person to vote A'. To show that j is a dictator for A', consider *any* profile $(A_1, ..., A_{j-1}, A', A_{j+1}, ..., A_n)$ in which j votes A'. The one-by-one substitution in $(A'', ..., A'', \underline{A'}, A, ..., A)$ of the votes of persons $i \neq j$ by their respective votes in $(A_1, ..., A_{j-1}, \underline{A'}, A_{j+1}, ..., A_n)$ leaves the outcome at A', by (JM) if $A_i = A'$ and by Lemma 1 if $A_i \neq A'$. So $F(A_1, ..., A_{j-1}, \underline{A'}, A_{j+1}, ..., A_n) = A'$.

3. *There is a dictator.* Repeating this argument with different triples $A, A', A'' \in \mathbf{A}$ shows that there is a dictator for *every* judgment set $A' \in \mathbf{A}$. But these dictators for particular judgment sets must all be the same person (consider profiles in which different judgment sets are voted by their respective dictators), who is hence a dictator simpliciter. ■

Theorem 1 also implies an impossibility result for independence (I_X). The reason is that ($\mathrm{I}_{\widetilde{X}}$) implies ($\mathrm{I}_X$) if the agenda X is *atomic*, i.e. if each consistent proposition in X is equivalent to a disjunction of atoms of X; here, an *atom (of X)* (not an "atomic proposition") is a maximally consistent member p of X, i.e. p is consistent and, for every $q \in X$, $p \vDash q$ or $p \vDash \neg q$. Equivalently, X is atomic if its set of atoms is exhaustive, i.e., for every truth-value assignment, X contains at least one true atom. Basic logic yields examples of atomic agendas X (where I denote by X_0 the set of atomic propositions occurring in proposition(s) in X):

(a) agendas X with finite X_0 for which $p, q \in X$ implies $p \wedge q \in X$ (or for which $p, q \in X$ implies $p \vee q \in X$, or for which $p, q \in X$ implies $p \rightarrow q \in X$);

(b) agendas X with finite X_0 and identical to their scope ($X = \overline{X}$);

(c) agendas $X = \{p, \sim p : p \in Y\}$, where Y consists of mutually exclusive and exhaustive propositions, e.g. $Y = \{a \wedge b, \neg a \wedge b, a \wedge \neg b, \neg a \wedge \neg b\}$.

Corollary 2 *If X is atomic and contains at least two propositions (not in trivial dependence), then an aggregation rule F is independent and weakly responsive (and satisfies universal domain and collective rationality) if and only if F is dictatorial.*

Proof. Let X be atomic. I have to show that (I_X) implies ($\mathrm{I}_{\widetilde{X}}$). This holds if every state of the world $q \in \widetilde{X}$ is logically equivalent to some atom r of X. Consider any $q = \wedge_{p \in A} p \in \widetilde{X}$ ($A \in \mathbf{A}$). Let B be the set of all atoms of X consistent with q. B is non-empty, since otherwise $q \vDash \neg r$ for all atoms r, and there would be a truth-value assignment (namely one that makes q true) making all atoms false. Let $r \in B$. I show that r is equivalent to q. A does not contain $\neg r$ (by consistency with r), hence contains r (by completeness). So

$q = \wedge_{p \in A} p \vDash r$. Also, $r \vDash q$. Otherwise r would be consistent with $\neg q$, hence with $\neg p$ for some $p \in A$, so that $r \vDash \neg p$ for this p (since q is an atom), and hence $p \vDash \neg r$, in contradiction with $\wedge_{p \in A} p \vDash r$. ■

So, coming from a somewhat different angle, Corollary 2 is an analogous result to Pauly and van Hees' [9] Theorem 3, except that their agenda is not assumed atomic but *atomically closed*, i.e. (i) if $p \in X$ and a is an atomic proposition occurring in p then $a \in X$, and (ii) if $p, q \in X$ are two literals (i.e. possibly negated atomic propositions) then $p \wedge q \in X$. (I drop their third condition, "if $a \in X$ is atomic then $\neg a \in X$", since I already assume X to contain proposition-negation pairs.) Let me combine both results in a single more general impossibility theorem. I call an agenda X *rich* if it is atomically closed *or* atomic, and contains at least two propositions (not in trivial dependence).

Theorem 2 *For a rich agenda X, an aggregation rule F is independent and weakly responsive (and satisfies universal domain and collective rationality) if and only if F is dictatorial.*

Incidentally, Theorems 1 and 2 have an interesting corollary on how independence (I_X) and independence on states of the world ($\mathrm{I}_{\widetilde{X}}$) are logically related – of which I otherwise have little intuition except that both are of course weaker than strong independence ($\mathrm{I}_{\overline{X}}$).

Corollary 3 *If (U) and (C) hold, ($I_{\widetilde{X}}$) implies (I_X) and both are equivalent for rich X.*

Proof. Let X contain at least two propositions not in trivial dependence (otherwise the claim is trivial since both ($\mathrm{I}_{\widetilde{X}}$) and ($\mathrm{I}_X$) hold). If F satisfies ($\mathrm{I}_{\widetilde{X}}$), by Theorem 1 F is dictatorial or not weakly responsive, hence satisfies (I_X). Conversely, if F satisfies (I_X) and X is rich, by Theorem 2 F is dictatorial or not weakly responsive, hence satisfies ($\mathrm{I}_{\widetilde{X}}$). ■

6 A possibility theorem on premise-based decision-making

Despite their merits in preventing manipulation, there are good reasons to reject the independence conditions (I_X)/($\mathrm{I}_{\widetilde{X}}$)/($\mathrm{I}_{\overline{X}}$). For they undermine premise-based reasoning on the collective level, i.e. the "collectivization of reason" (Pettit [10]). For instance, (I_X) prevents the collective from accepting b *because* it accepts the premises a and $a \rightarrow b$, and from accepting c *because* it accepts the premises a, b, and $c \leftrightarrow (a \& b)$ (all propositions in X). I therefore suggest imposing instead *independence on premises*, which allows judgments about "conclusions" to be derived from judgments about "premises".

The so-called *premise-based procedure* is usually defined only in the context of the *discursive dilemma* or *doctrinal paradox* (e.g. Pettit [10]). To generalise this procedure, suppose there is a set $P \subseteq X$ of propositions considered as *premises*, where P consists of proposition-negation pairs, i.e. $p \in P$ implies $\sim p \in P$. (P is related to Osherson's "basis".[8])

Definition 3 *The "premise-based procedure (for set of premises P)" is the aggregation rule F with universal domain such that, for each $(A_1, ..., A_n) \in \mathbf{A}^n$, $F(A_1, ..., A_n) = \{p \in X : P^* \vDash p\}$, where $P^* := \{p \in P : n_p > n_{\sim p}$, or $[n_p = n_{\sim p}$ and p is a negated proposition$]\}$ with n_p denoting the number of persons i with $p \in A_i$.*

[8] D. Osherson, Notes on Aggregating Belief, unpublished paper, Princeton University, 2004.

So the premise-based procedure first votes on premises, and then forms the deductive closure in X. To break potential ties in the case of even group size n, by convention $\neg q$ wins over q whenever there is a tie between $q, \neg q \in P$. (List's [6] *priority-to-the-past* rule is another generalisation of the premise-based procedure.)

I now prove, in short, that premise-based decision-making is possible if the system of premises is logically independent. Consider the following conditions (where $Y \subseteq \overline{X}$).

Anonymity (A). For every two profiles $(A_1, ..., A_n), (A_{\pi(1)}, ..., A_{\pi(n)}) \in Dom(F)$, where $\pi : \{1, ..., n\} \to \{1, ..., n\}$ is any permutation of the individuals, $F(A_1, ..., A_n) = F(A_{\pi(1)}, ..., A_{\pi(n)})$.

Monotonicity on Y ($\mathbf{M}_Y$). For each proposition $p \in Y$, individual i and i-variants $(A_1, ..., A_n), (A_1, \ldots, A_i^*, \ldots, A_n) \in Dom(F)$ with $A_i \not\models p$ and $A_i^* \models p$, if $F(A_1, ..., A_n) \models p$ then $F(A_1, \ldots, A_i^*, \ldots, A_n) \models p$.

Systematicity on Y ($\mathbf{S}_Y$). For every two propositions $p, p' \in Y$ and every two profiles $(A_1, ..., A_n), (A'_1, ..., A'_n) \in Dom(F)$, if [for every person i, $A_i \models p$ if and only if $A'_i \models p'$], then [$F(A_1, ..., A_n) \models p$ if and only if $F(A'_1, ..., A'_n) \models p'$].

(S_Y) generalises List and Pettit's [7] *systematicity*, and implies (I_Y) (take $p = p'$). It requires not only propositionwise aggregation on Y (like (I_Y)) but also the use of the *same* decision method for each $p \in Y$. More precisely, one easily proves the following:

Proposition 2 *Let $Y \subseteq \overline{X}$. F is systematic on Y (S_Y) if and only if F applies an identical decision method M for each proposition $p \in Y$.*

Definition 4 *Condition (I_P)/(S_P)/(M_P) is called "independence/systematicity/monotonicity on premises". The system of premises P is "(logically) independent" if every subset $A \subseteq P$ that contains exactly one member of each pair $p, \neg p \in P$ is consistent. The "scope of P" is the set $\overline{P}$ of all propositions $p \in \mathcal{L}$ settled by any $A \subseteq P$ that is consistent and complete in P (i.e. P contains a member of each pair $p, \neg p \in P$).*

For instance, P is independent if it consists of atomic propositions (and their negations).

Theorem 3 *Assume the system of premises P is logically independent. Then*

(i) the premise-based procedure generates consistent judgment sets;

(ii) if $X \subseteq \overline{P}$ (so $\overline{X} = \overline{P}$), the premise-based procedure satisfies collective rationality, and if also n is odd it is the only aggregation rule that is systematic on premises, monotonic on premises and anonymous and satisfies universal domain and collective rationality.

Here, "$X \subseteq \overline{P}$" means that the premises do not underdetermine the judgments to be made. If X is the agenda of the discursive dilemma, $\{a, b, c, c \leftrightarrow (a \wedge b)$,negations$\}$, then $P := \{a, b, c \leftrightarrow (a \wedge b)$,negations$\}$ not only is logically independent, but also satisfies $X \subseteq \overline{P}$.

Proof. Assume P is logically independent, and let F be the premise-based procedure.

(i) For each $(A_1, ..., A_n) \in \mathbf{A}^n$, the set $P^* \subseteq P$ (see Definition 3) is consistent since P^* contains exactly one member of each pair $p, \neg p \in P$ and P is logically independent. Hence $F(A_1, ..., A_n) = \{p \in X : P^* \models p\}$ is consistent.

(ii) Assume $X \subseteq \overline{P}$. For each $(A_1, ..., A_n) \in \mathbf{A}^n$, the set $P^* \subseteq P$ is consistent and complete in P, as seen in (i). So, as $X \subseteq \overline{P}$, P^* settles each $p \in X$. Hence $F(A_1, ..., A_n) = \{p \in X :$

$P^* \vDash p\}$ is consistent and complete. So F satisfies (C). Now let n be odd. F satisfies (S_P) (as n is odd), (M_P), (A), (U) and (C). Conversely, assume F^* satisfies all these conditions. I show that $F^* = F$. By (S_P) and Proposition 2, F^* applies some identical decision method M for each premise $p \in P$. By (A), $M(t_1, ..., t_n)$ depends only on the *number* of persons i with $t_i = 1$, i.e. there exists a function $g : \{0, ..., n\} \to \{0, 1\}$ such that, for all $(A_1, ..., A_n) \in \mathbf{A}^n$ and $p \in P$, $[p \in F^*(A_1, ..., A_n)$ if and only if $g(|N_p|) = 1]$, where $N_p := \{i : p \in A_i\}$. By (M_P) and (U), $g(k) \leq g(k + 1)$ for all $k \in \{0, ..., n - 1\}$. Hence, by induction, (a) $k < l$ *implies* $g(k) \leq g(l)$, *for all* $k, l \in \{0, ..., n\}$. As by (C) exactly one of each pair $p, \neg p \in P$ is collectively accepted, we have $g(|N_p|) + g(|N_{\neg p}|) = 1$ for all $(A_1, ..., A_n) \in \mathbf{A}^n$, and so (b) $g(k) + g(n - k) = 1$ *for all* $k \in \{0, ..., n\}$. For, as $(A_1, ..., A_n)$ runs through $\mathbf{A}^n$, $|N_p|$ runs through $\{0, ..., n\}$ and always takes the value $n - |N_p|$. Of course, the only solution of (a) and (b) (for odd n) is given by $g(k) = 0$ for $0 \leq k < n/2$ and $g(k) = 1$ for $n/2 < k \leq n$. So F^* applies, like F, propositionwise majority voting for each premise $p \in P$. Hence, for all $(A_1, ..., A_n) \in \mathbf{A}^n$, $F^*(A_1, ..., A_n) \cap P = F(A_1, ..., A_n) \cap P =: A^*$. As F^* satisfies collective rationality, A^* is consistent and complete in P. So, by $X \subseteq \overline{P}$, A^* settles each $p \in X$. Hence, again by collective rationality of F^*, $F^*(A_1, ..., A_n) = \{p \in X : A^* \vDash p\}$, and so $F^* = F$. ■

7 Brief summary

Independence conditions are crucial to protect against manipulation both by agenda setters and by voters. In particular, independence on states of the world protects against logical agenda manipulation, strong independence protects against general agenda manipulation, and independence on Y ($\subseteq \overline{X}$) together with monotonicity on Y guarantees strategyproofness on Y. However, different impossibility theorems establish that these independence conditions cannot be fulfilled together with the minimal conditions of weak responsiveness and non-dictatorship (and universal domain and collective rationality). Unlike earlier impossibility theorems by List and Pettit and by Pauly and van Hees, my main impossibility result is valid for any agenda (with at least two propositions not in trivial dependence).

However, even ignoring impossibility results, independence requirements are inherently problematic as they undermine premise-driven collective judgment formation. I therefore suggested imposing merely *independence on premises*. This allows for the *premise-based procedure*, which was shown to generate consistent collective judgment sets provided that the system of premises is logically independent. This leaves open the practically important question of how to determine a system of premises – one of many future challenges.

8 References

1. L. Bovens, W. Rabinowicz, Democratic Answers to Complex Questions - an Epistemic Perspective, Synthese, forthcoming.
2. G. Brennan, Collective Coherence? Int. Rev. Law Econ. 21(2) (2001), 197-211.
3. B. Chapman, Rational Aggregation, Polit. Philos. Econ. 1(3) (2002), 337-354.
4. C. List, A Possibility Theorem on Decisions over Multiple Propositions, Math. Soc. Sci. 45 (1) (2003), 1-13.
5. C. List, The Probability of Inconsistencies in Complex Collective Decisions, Soc. Choice Welfare, forthcoming.

6. C. List, A Model of Path-Dependence in Decisions over Multiple Propositions, Amer. Polit. Sci. Rev., forthcoming.
7. C. List, P. Pettit, Aggregating Sets of Judgments: an Impossibility Result, Economics and Philosophy 18 (2002), 89-110.
8. C. List, P. Pettit, Aggregating Sets of Judgments: two Impossibility Results Compared, Synthese, forthcoming.
9. M. Pauly, M. van Hees, Logical Constraints on Judgment Aggregation, Journal of Philosophical Logic, forthcoming.
10. P. Pettit, Deliberative Democracy and the discursive dilemma, Philosophical Issues (supplement 1 of Nous) 11 (2001), 268-95.

Chapter 6

Aggregating conditional judgments

Paper: The possibility of judgment aggregation on agendas with subjunctive implications, *Journal of Economic Theory*, forthcoming

The possibility of judgment aggregation on agendas with subjunctive implications

Franz Dietrich[1]

Abstract. The new field of judgment aggregation aims to find collective judgments on logically interconnected propositions. Recent impossibility results establish limitations on the possibility to vote independently on the propositions. I show that, fortunately, the impossibility results do not apply to a wide class of realistic agendas once propositions like "if a then b" are adequately modelled, namely as subjunctive implications rather than material implications. For these agendas, consistent and complete collective judgments can be reached through appropriate quota rules (which decide propositions using acceptance thresholds). I characterise the class of these quota rules. I also prove an abstract result that characterises consistent aggregation for arbitrary agendas in a general logic.

Key words: judgment aggregation, subjunctive implication, material implication, characterisation of possibility agendas

JEL Classification Numbers: D70, D71, D79

1 Introduction

In judgment aggregation, the objects of the group decision are not as usual (mutually exclusive) *alternatives*, but *propositions* representing interrelated (yes/no) questions the group faces. To ensure that these interrelations are well-defined, propositions are statements in a formal logic. As a simple example, suppose the three-member board of a central bank disagrees on which of the following propositions hold.

a :	GDP growth will pick up.
b :	Inflation will pick up.
$a \to b$:	*If* GDP growth will pick up *then* inflation will pick up.

Reaching collective beliefs is non-trivial. In Table 1, each board member holds consistent (yes/no) beliefs but the propositionwise majority beliefs are inconsistent. To achieve consistent collective judgments, the group cannot use majority voting. What procedure should the group use instead? A wide-spread view is that, in this as in most other judgment aggregation problems, we must give up aggregating propositionwise[2], for instance in favour of a premise-base rule (as discussed below). I

[1]This paper was presented at the *Risk, Uncertainty and Decision* seminar (MSE, Paris, April 2005), the *Aggregation of Opinions* workshop (Yale Law School, September 2006), and the 8th *Augustus de Morgan* workshop (King's College London, November 2006). Benjamin Polak's extensive comments and suggestions have benefited the paper on substantive and presentational levels. I am also grateful for helpful comments by two referees and by Christian List and Philippe Mongin.

[2]That is, aggregating by voting independently on the propositions: the collective judgment on any proposition p depends only on how the individuals judge p, not on how they judge other propositions. This property is usually called "independence".

	a	$a \to b$	b
1/3 of the board	Yes	Yes	Yes
1/3 of the board	No	Yes	No
1/3 of the board	Yes	No	No
Collective under majority rule	Yes	Yes	No
Collective under premise-based rule	Yes	Yes	Yes
Collective under the (below-defined) quota rule	No	No	No

Table 1: A simple judgment aggregation problem and three aggregation rules

show that this conclusion is often an artifact of an inappropriate way to model implications like $a \to b$. In many judgment aggregation problems, a more appropriate *subjunctive* interpretation of implications changes the logical relations between propositions in such a way that we can aggregate on a propositionwise basis without creating collective inconsistencies. Indeed, we can use *quota rules*: here, separate anonymous votes are taken on each proposition using (proposition-specific) acceptance thresholds. Suppose for instance that the following thresholds are used: b is accepted if and only if a majority accepts b, and a $p \in \{a, a \to b\}$ is accepted if and only if at least 3/4 of people accept p. Then, in the situation of Table 1, a, $a \to b$ and b are all rejected, i.e. the outcome is $\{\neg a, \neg(a \to b), \neg b\}$.

The problem is that this outcome, although intuitively perfectly consistent, is declared *in*consistent in classical logic, because classical logic defines $\neg(a \to b)$ as equivalent to $a \wedge \neg b$ ("a and not-b"), by interpreting "$\to$" as a *material* rather than a *subjunctive* implication. Is this equivalence plausible in our example? Intuitively, $a \wedge \neg b$ does indeed entail $\neg(a \to b)$, but $\neg(a \to b)$ does not entail $a \wedge \neg b$ because $\neg(a \to b)$ does not intend to say anything about whether a and b are *actually* true or false: rather it intends to say that b *would* be false in the *hypothetical* (hence possibly counterfactual) case of a's truth. Indeed, a person who believes that it is false that a pick up in GDP growth leads to a pick up in inflation may or may not believe that GDP growth or inflation will *actually* pick up; what he believes is rather that inflation will not pick up in the *hypothetical* case(s) that GDP growth will pick up.

In real-life judgment aggregation problems, implication statements usually have a subjunctive meaning. It is important not to misrepresent this meaning using material implications and classical logic, because this creates unnatural logical connections and artificial impossibilities of aggregation. The above quota rule, for instance, guarantees collective consistency (not just for the profile in Table 1) if the implication "$a \to b$" is subjunctive, but not if it is material. More generally, I establish the existence of quota rules with consistent outcomes for a large class of realistic agendas: the so-called *implication agendas*, which contain (bi-)implications and atomic propositions. This possibility is created by interpreting (bi-)implications subjunctively; it disappears if we instead use classical logic, i.e. interpret (bi-)implications materially. At first sight, this positive finding seems in conflict with the recent surge of impossibility results on propositionwise aggregation (see below). In fact, these results presuppose logical interconnections between propositions that are stronger than (or different to) those which I obtain here under the subjunctive interpretation of (bi-)implications. In various results, I derive the (necessary and sufficient) conditions that the acceptance thresholds of quota rules must satisfy in order to guarantee consistent outcomes.

These results are applications of an abstract characterisation result, Theorem 3, which is valid for arbitrary agendas in a general logic. It also generalises the "intersection property" result by Nehring and Puppe [18, 19] (but not that by Dietrich and List [6]).

Although I show that collective consistency is often achievable by aggregating propositionwise (using quota rules), I do not wish to generally advocate propositionwise aggregation. In particular, one may reject propositionwise aggregation rules by arguing that they neglect *relevant information*: in order to decide on b it is arguably not just relevant how people judge b but also *why* they do so, i.e. how they judge b's "premises" a and $a \rightarrow b$. This naturally leads to the popular *premise-based* rule: here, only a and $a \rightarrow b$ – the "premises" – are decided through (majority) votes, while b – the "conclusion" – is accepted if and only if a and $a \rightarrow b$ have been accepted; so that, in the situation of Table 1, a and $a \rightarrow b$, and hence b, are accepted.

Despite the mentioned objection, propositionwise aggregation rules are superior from a manipulation angle: non-propositionwise aggregation rules can be manipulated by agenda setters (Dietrich [2]) and by voters (Dietrich and List [5]).[3]

In general, the judgment aggregation problem – deciding which propositions to accept based on which ones the individuals accept – and its formal results are open to different interpretations of "accepting" and different sorts of propositions. This paper's examples and discussion focus on the case that "accepting" means "believing",[4] and mostly on the case that the propositions have a *descriptive* content (like "GDP growth will pick up"), although Section 4 touches on *normative* propositions (like "peace is better than war").[5]

In the literature, judgment aggregation is discussed on a less formal basis in law (e.g. Kornhauser and Sager [12], Chapman [1]) and political philosophy (e.g. Pettit [22]), and is formalised in List and Pettit [15] who use classical propositional logic. Also the related *belief merging* literature in artificial intelligence uses classical propositional logic to represent propositions (e.g. Konieczny and Pino-Perez [11] and Pigozzi [23]). A series of results establish, for different agendas, the impossibility of aggregating on a propositionwise basis in accordance with collective consistency and different other conditions (e.g. List and Pettit [15], Pauly and van Hees [21], Dietrich [2, 4], Gärdenfors [10], Nehring and Puppe [20], van Hees [26], Dietrich and List [7], Dokow and Holzman [9] and Mongin [17]). Further impossibilities (with minimal agenda conditions) follow from Nehring and Puppe's [18, 19] results on strategy-proof social choice. To achieve possibility, propositionwise aggregation is given up in favour of *distance-based* aggregation by Pigozzi [23] (drawing on Konieczny and Pino-Perez [11]), of *sequential* aggregation by List [14] and Dietrich and List [6], and of aggregating *relevant information* by Dietrich [4].

[3] Consider for instance premise-based voting in Table 1. The agenda setter may reverse the outcome on b by replacing the premises a and $a \rightarrow b$ by other premises a' and $a' \rightarrow b$. Voter 2 or 3 can reverse the outcome on b by pretending to reject *both* premises a and $a \rightarrow b$.

[4] Judgment aggregation is the aggregation of belief sets if "accepting" means "believing", the aggregation of desire sets if "accepting" means "desiring", the aggregation of moral judgment sets if "accepting" means "considering as morally good", etc.

[5] By considering beliefs on possibly normative propositions, judgment aggregation uses a broader "belief" notion than is common in economics, where beliefs usually apply to descriptive facts only. For instance, standard preference aggregation problems can be modelled as judgment aggregation problems by interpreting preferences as *beliefs* of normative propositions like "x is better than y" (see Dietrich and List [7]; also List and Pettit [16]).

Section 7 uses Dietrich's [3] judgment aggregation model in general logics, and the other sections use for the first time possible-worlds semantics.

2 Definitions

We consider a group or persons $N = \{1, 2, \ldots, n\}$ $(n \geq 2)$, who need collective judgments on a set of propositions expressed in formal logic.

The language. Following Dietrich's [3] general logics model, a language is given by a non-empty set $\mathbf{L}$ of sentences (called *propositions*) closed under negation, i.e. $p \in \mathbf{L}$ implies $\neg p \in \mathbf{L}$. (Interesting languages of course have also other connectives than negation $\neg$). Logical interconnections are captured either by an *entailment relation* $\vDash$ (telling for which $A \subseteq \mathbf{L}$ and $p \in \mathbf{L}$ we have $A \vDash p$) or, equivalently, by a *consistency* notion (telling which sets $A \subseteq \mathbf{L}$ are consistent).[6] The two notions are interdefinable: a set $A \subseteq \mathbf{L}$ is inconsistent if and only $A \vDash p$ and $A \vDash \neg p$ for some $p \in \mathbf{L}$; and an entailment $A \vDash p$ holds if and only if $A \cup \{\neg p\}$ is inconsistent.[7] The precise nature of logical interconnections is addressed later. A proposition $p \in \mathbf{L}$ is a *contradiction* if $\{p\}$ is inconsistent, and a *tautology* if $\{\neg p\}$ is inconsistent.

All following sections except Section 7 consider a *particular* language: $\mathbf{L}$ is the set of propositions constructible using $\neg$ ("not"), $\wedge$ ("and") and $\rightarrow$ ("if-then") from a set $\mathcal{A} \neq \emptyset$ of non-decomposable symbols, called *atomic* propositions (and representing simple statements like "inflation will pick up"). So $\mathbf{L}$ is the smallest set such that (i) $\mathcal{A} \subseteq \mathbf{L}$ and (ii) $p, q \in \mathbf{L}$ implies $\neg p \in \mathbf{L}$, $(p \wedge q) \in \mathbf{L}$ and $(p \rightarrow q) \in \mathbf{L}$. The critical question, treated in the next section, is how (not) to define the logical interconnections on $\mathbf{L}$: while some entailments like $a, b \vDash a \wedge b$ and $a, a \rightarrow b \vDash b$ are not controversial, others are. Notationally, I drop brackets when there is no ambiguity, e.g. $c \rightarrow (a \wedge b)$ stands for $(c \rightarrow (a \wedge b))$. Further, $p \vee q$ ("p or q") stands for $\neg(\neg p \wedge \neg q)$, and $p \leftrightarrow q$ ("p if and only if q") stands for $(p \rightarrow q) \wedge (q \rightarrow p)$. For any conjunction $p = a_1 \wedge ... \wedge a_k$ of one or more atomic propositions $a_1, ..., a_k$ (called the *conjuncts* of p), let $\mathrm{C}(p) := \{a_1, ..., a_k\}$ (e.g. $\mathrm{C}(a) = \{a\}$ and $\mathrm{C}(a \wedge b) = \mathrm{C}(b \wedge a) = \{a, b\}$).

In judgment aggregation, the term "connection rule" commonly refers to implicational statements like "*if* GDP growth continues *and* interest rates stay below X *then* inflation will rise". I now formalise this terminology. If each of p and q is a conjunction of one or more atomic propositions,

- $p \rightarrow q$ is a *uni-directional connection rule*, called *non-degenerate* if $\mathrm{C}(q) \backslash \mathrm{C}(p) \neq \emptyset$, i.e. if $p \rightarrow q$ is not a tautology (under the classical or the non-classical entailment relation discussed later);
- $p \leftrightarrow q$ is a *bi-directional connection rule*, called *non-degenerate* if $\mathrm{C}(q) \backslash \mathrm{C}(p) \neq \emptyset$ and $\mathrm{C}(p) \backslash \mathrm{C}(q) \neq \emptyset$, i.e. if neither $p \rightarrow q$ nor $q \rightarrow p$ is a tautology.

A uni- or bi-directional connection rule is simply called a *connection rule*.

[6] For the two approaches, see Dietrich [3]. Logical interconnections can be interpreted either *semantically* or *syntactically* (in the latter case, the symbol "$\vdash$" is more common than "$\vDash$"). In the (classical or non-classical) logics considered in Sections 3-6, I define interconnections semantically (but there are equivalent syntactic definitions). Dropping brackets, I often write $p_1, ..., p_k \vDash p$ for $\{p_1, ..., p_k\} \vDash p$.

[7] The latter equivalence supposes that the logic is not paraconsistent. All logics considered in this paper are of this kind.

The agenda. The *agenda* is the set of propositions on which decisions are needed. Formally, it is a non-empty set $X \subseteq \mathbf{L}$ of the form $X = \{p, \neg p : p \in X^+\}$ for some set X^+ containing no negated proposition $\neg q$. In the introductory example, $X^+ = \{a, b, a \to b\}$. Notationally, double-negations cancel each other out: if $p \in X$ is a negated proposition $\neg q$ then hereafter when I write "$\neg p$" I mean q rather than $\neg\neg q$. (This ensures that $\neg p \in X$.)

An agenda$X$ (in the language $\mathbf{L}$ just defined) is an *implication* agenda if X^+ consists of non-degenerate connection rules and the atomic propositions occurring in them; it is called *simple* if all its connection rules are uni-directional ones $p \to q$ in which p and q are atomic propositions.

Many standard examples of judgment aggregation problems can be modelled with implication agendas. The atomic propositions represent (controversial) issues, and connection rules represent (controversial) links between issues. Any accepted connection rule establishes a constraint on how to decide the issues.

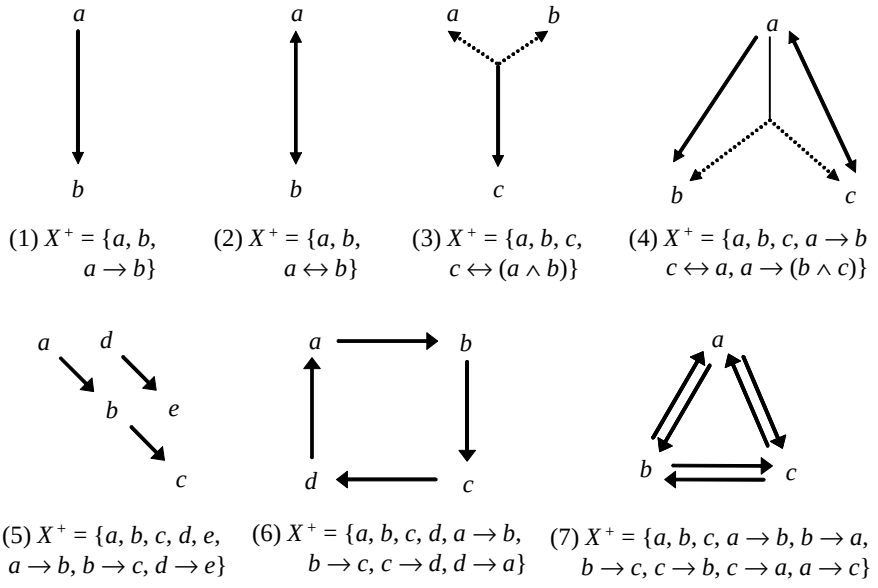

Figure 1: Seven implication agendas, represented as networks.

Implication agendas can always be represented graphically as networks over its atomic propositions.[8] Figure 1 shows seven implication agendas, of which (1) and (5)-(7) are simple. Agenda (1) represents our central bank example. An environmental expert commission might face the agenda (2), where a is "global warming will continue" and b is "the ozone hole exceeds size X". The judges of a legal court face, in the decision problem from which judgment aggregation originated, an agenda of type (3), where a is "the defendant has broken the contract", b is "the contract is legally valid", c is "the defendant is liable", and $c \leftrightarrow (a \wedge b)$ is a claim on what constitutes (necessary and sufficient) conditions of liability.[9] A company board trying to predict

[8] Nodes contain atomic propositions. Arrows represent connection rules: bi-directional arrows indicate bi-implications, and bifurcations indicate conjunctions of more than one atomic proposition.

[9] A *doctrinal paradox* arises if there is a majority for a, a majority for b, a unanimity for $c \leftrightarrow (a \wedge b)$,

the price policy of three rival firms A-C might face the agenda (3) or (4) or (7), where a is "Firm A will raise prices", b is "Firm B will raise prices", and c is "Firm C will raise prices". The three agendas differ in the type of connections between a, b, c deemed possible.

In Section 4, I discuss two types of decision problems captured by implication agendas: reaching judgments on facts and their causal relations, and reaching judgments on hypotheses and their justificational/evidential relations.

But not all realistic judgment aggregation problems are formalisable by implication agendas. Some judgment aggregation problems involve a generalised kind of implication agenda, obtained by generalising the definition of connection rules so as to include (bi-)implications between propositions p and q other than conjunctions of atomic propositions.[10] More radical departures from implication agendas include: (i) the agenda given by $X^+ = \{a, b, a \wedge b\}$, which contains no connection rule but the Boolean expression $a \wedge b$; (ii) the agenda representing a preference aggregation problem, which contains propositions of the form xRy from a predicate logic (see Section 7); (iii) agendas where X^+ contains only atomic propositions, between which certain connection rules are imposed exogenously (rather than subjected to a decision).

Judgment sets. A *judgment set* (held by a person or the group) is a subset $A \subseteq X$; $p \in A$ stands for "the person/group accepts proposition p". A judgment set A can be more or less rational. Ideally, it should be both *complete*, i.e. contain at least one member of each pair $p, \neg p \in X$, and *(logically) consistent*. A is *weakly consistent* if A does not contain a pair $p, \neg p \in X$ (i.e., intuitively, if A is not "obviously inconsistent"). For agenda (1) in Figure 1, $\{a, a \rightarrow b, \neg b\}$ is complete, weakly consistent, but not consistent because A entails b (in fact, $\{a, a \rightarrow b\}$ entails b) and A entails $\neg b$ (in fact, contains $\neg b$). So to say, "weak consistency" means not to *contain* a contradiction $p, \neg p$, and consistency means not to *entail* one.

Aggregation rules. A *profile* is an n-tuple $(A_1, \ldots, A_n)$ of (individual) judgment sets $A_i \subseteq X$. A *(judgment) aggregation rule* is a function F that maps each profile $(A_1, \ldots, A_n)$ from a given domain of profiles to a (group) judgment set $F(A_1, \ldots, A_n) = A \subseteq X$. The domain of F is *universal* if it consists of all profiles of complete and consistent judgment sets. F is *complete/consistent/weakly consistent* if F generates a complete/consistent/weakly consistent judgment set for each profile in its domain. On the universal domain, *majority rule* (given by $F(A_1, ..., A_n) = \{p \in X :$ more persons i have $p \in A_i$ than $p \notin A_i\}$) is weakly consistent, and a *dictatorial rule* (given by $F(A_1, ..., A_n) = A_j$ for a fixed j) is even consistent. We will focus on *quota rules* thus defined. To each family $(m_p)_{p \in X^+}$ of numbers in $\{1, ..., n\}$, the *quota rule with thresholds* $(m_p)_{p \in X^+}$ is the aggregation rule with universal domain given by

$$F_{(m_p)_{p \in X^+}}(A_1, ..., A_n) = \{p \in X : \text{at least } m_p \text{ persons } i \text{ have } p \in A_i\},$$

but a majorty for $\neg c$.

[10] If, for instance, p and q were allowed to be *dis*junctions of atomic propositions then $a \rightarrow (b \vee c)$ would count as a connection rule, so that $X^+ = \{a, b, c, (a \vee b) \rightarrow c\}$ would define an implication agenda of the so-generalised kind. Generalised implication agendas may well be relevant as groups may need to make up their mind on generalised types of connection rules. The possibility of consistent aggregation by quota rules may disappear for such agendas. So our subjunctive reading of (bi-)implications (see Section 3) is not a general recipe for possibility in judgment aggregation.

where $m_{\neg p} := n - m_p + 1$ for all $p \in X^+$ to ensure exactly one member of each pair $p, \neg p \in X$ is accepted, i.e. that quota rules are complete and weakly consistent.

So each family of thresholds $(m_p)_{p \in X^+}$ in $\{1, ..., n\}$ generates a quota rule. As one easily checks, an aggregation rule is a quota rule if and only if it has universal domain and is complete, weakly consistent, independent, anonymous, monotonic, and responsive.[11] The important property missing here is consistency. We will investigate if and how the thresholds can be chosen so as to achieve consistency. The properties of independence and monotonicity are equivalent to *strategy-proofness* if each individual i holds *epistemic* preferences, i.e. would like the group to hold beliefs close to A_i, the set of propositions i considers true.[12]

3 A non-classical logic

How should we define the logical interconnections within the language **L** specified in Section 2? Although classical logic gets some entailments right (like $a, a \rightarrow b \vDash b$), its treatment of connection rules is inappropriate, or so I will argue.

Requirements on the representation of connection rules. To reflect the intended meaning of connection rules such as $a \rightarrow b, c \leftrightarrow a, a \rightarrow (b \wedge c)$, the logic should respect the following conditions.

(a) The *acceptance* of a connection rule r establishes exactly the intended logical constraints on atomic propositions, i.e. r is consistent with the "right" sets of atomic and negated atomic propositions. For instance, $a \rightarrow b$ is inconsistent with $\{a, \neg b\}$ but consistent with each of $\{a, b\}$, $\{\neg a, b\}$, $\{\neg a, \neg b\}$.

(b) The *negation* of a (non-degenerate) connection rule r does *not* constrain atomic propositions, i.e. $\neg r$ is consistent with *each* (consistent) set of atomic and negated atomic propositions. For instance, $\neg(a \rightarrow b)$ is consistent with each of $\{a, b\}$, $\{a, \neg b\}$, $\{\neg a, b\}$, $\{\neg a, \neg b\}$.

To illustrate (b), consider again the central bank example, where a is "GDP growth will pick up" and b is "inflation will pick up". Consider a board member who believes that $\neg(a \rightarrow b)$, i.e. that rising GDP does *not* imply rising inflation. This

[11] *Independence*: for all $p \in X$ and all admissible profiles $(A_1, ..., A_n)$, $(A_1^*, ..., A_n^*)$, if $\{i : p \in A_i\} = \{i : p \in A_i^*\}$ then $p \in F(A_1, ..., A_n) \Leftrightarrow p \in F(A_1^*, ..., A_n^*)$. *Anonymity*: $F(A_1, ..., A_n) = F(A_{\pi(1)}, ..., A_{\pi(n)})$ for all admissible profiles $(A_1, ..., A_n)$, $(A_{\pi(1)}, ..., A_{\pi(n)})$, where $\pi : N \mapsto N$ is any permutation. *Monotonicity*: for all individuals i and admissible profiles $(A_1, ..., A_n)$, $(A_1, ..., A_i^*, ..., A_n)$ differing only in i's judgment set, if $F(A_1, ..., A_n) = A_i^*$ then $F(A_1, ..., A_i^*, ..., A_n) = A_i^*$. *Responsiveness*: for all $p \in X$ (such that neither p nor $\neg p$ is a tautology) there are admissible profiles $(A_1, ..., A_n)$, $(A_1^*, ..., A_n^*)$ with $p \in F(A_1, ..., A_n)$ and $p \notin F(A_1^*, ..., A_n^*)$. Clearly, quota rules satisfy all seven axioms. Conversely, independence and anonymity imply that the group judgment on any given $p \in X$ depends only on the number $n_p := |\{i : p \in A_i\}|$. This dependence is positive by monotonicity, hence described by an acceptance threshold $m_p \in \{0, 1, ..., n+1\}$. If p and $\neg p$ are not tautologies, m_p is by responsiveness not 0 and not $n+1$, i.e. $m_p \in \{1, ..., n\}$; and $m_{\neg p} = n - m_p + 1$ by completeness and weak consistency. If p or $\neg p$ is a tautology, we may assume w.l.o.g. that, again, $m_p \in \{1, ..., n\}$ and $m_{\neg p} = n - m_p + 1$.

[12] That is, i weakly prefers the group to hold judgment set A over judgment set B if for all $p \in X$ on which A_i agrees with B, A_i also agrees with A. This condition only partly fixes i's preferences, but it for instance implies that i most prefers i's own judgment set A_i. See Dietrich and List [5].

belief is intuitively perfectly consistent with any beliefs on a and b, i.e. on whether GDP will grow and whether inflation will rise.

The failure of the material implication. Material (bi-)implications (used in classical logic) satisfy (a) but not (b). Consider $a \to b$. Interpreted materially, $a \to b$ is equivalent to $\neg a \vee b$ (not-a or b), and $\neg(a \to b)$ to $a \wedge \neg b$ (a and not-b); so:

- (a) holds because $a \to b$ is inconsistent with $\{a, \neg b\}$ (as desired) and consistent with each of $\{a, b\}, \{\neg a, b\}, \{\neg a, \neg b\}$ (as desired);
- (b) is violated because $\neg(a \to b)$, far from imposing no constraints, is inconsistent with all sets containing $\neg a$ or containing b.

It is well-known that the material interpretation misrepresents the intended meaning of most conditional statements in common language. The (in common language clearly false) statement "if the sun stops shining then we burn" is *true* materially because the sun does *not* stop shining. The material interpretation clashes with intuition because, in common language, "if a then b" is not a statement about the *actual* world, but about whether b holds in hypothetical world(s) where a holds, e.g. worlds where the sun stops shining. "If a then b" thus means "if a *were* true ceteris paribus, then b *would be* true", not "a *is* false or b *is* true".

A conditional logic. A *subjunctive* reading of "$\to$", where the truth value of $a \to b$ depends on b's truth value in possibly non-actual worlds, has been formalised using *possible-worlds semantics*, and more specifically using *conditional logic* which originated from Stalnaker [25] and D. Lewis [13] and is now well-established in non-classical logic. I use a standard version of conditional logic, sometimes denoted C^+ (other versions could also be used). For further reference, e.g. Priest [24].

For comparison, recall that in *classical* logic (not in C^+) $A \subseteq \mathbf{L}$ entails $p \in \mathbf{L}$ if and only if every classical interpretation that makes all $q \in A$ true makes p true, where a *classical interpretation* is simply a ("truth") function $v : \mathbf{L} \to \{T, F\}$ that assigns to each proposition a truth value such that, for all $p, q \in \mathbf{L}$,

- $v(\neg p) = T$ if and only if $v(p) = F$,
- $v(p \wedge q) = T$ if and only if $v(p) = T$ and $v(q) = T$,
- $v(p \to q) = T$ if and only if $v(p) = F$ or $v(q) = T$ (material implication).

This leads to counter-intuitive entailments like $\neg a \vDash a \to b$ and $b \vDash a \to b$, the so-called paradoxes of material implication. In response, the notion of an "interpretation" must be redefined. A C^+*-interpretation* consists of

- a non-empty set W of *(possible) worlds* w;
- for every proposition $p \in \mathbf{L}$ a function $f_p : W \to \mathcal{P}(W)$ ($f_p(w)$ contains the worlds to which "if p were true" refers, i.e. the worlds "similar" to w and with true p);
- for every world $w \in W$ a ("truth") function $v_w : \mathbf{L} \to \{T, F\}$ (that tells what propositions hold in w).

But not *any* such triple $(W, (f_p), (v_w)) \equiv (W, (f_p)_{p \in \mathbf{L}}, (v_w)_{w \in W})$ may reasonably count as an interpretation: indeed, the meaning of the functions f_p and v_w suggests

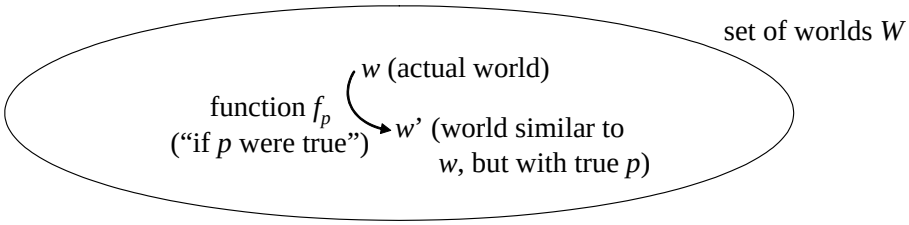

Figure 2: Referring to a non-actual world, in a C^+-interpretation

requiring additional properties. Specifically, such a triple $(W, (f_p), (v_w))$ is defined as a *(C^+-)interpretation* if, for all worlds $w \in W$ and all propositions $p, q \in \mathbf{L}$,

- $v_w(\neg p) = T$ if and only if $v_w(p) = F$ (like in classical logic),
- $v_w(p \wedge q) = T$ if and only if $v_w(p) = T$ and $v_w(q) = T$ (like in classical logic),
- $v_w(p \rightarrow q) = T$ if and only if $v_{w'}(q) = T$ for all worlds $w' \in f_p(w)$ (subjunctive implication),
- if $w' \in f_p(w)$ then $v_{w'}(p) = T$ (i.e. p holds in the worlds to which "if p were true" refers),
- if $v_w(p) = T$ then $w \in f_p(w)$ (i.e. if p already holds in w then "if p were true" refers to w).

The truth condition for $p \rightarrow q$ (third bullet point) captures the intuitive meaning of implications. "If the sun stops shining then we burn" is false in our world: we do not burn in worlds similar to ours but without the sun shining.

By definition, $A \subseteq \mathbf{L}$ *(C^+-)entails* $p \in \mathbf{L}$ ($A \vDash p$) if, for all interpretations $(W, (f_p), (v_w))$ and all worlds $w \in W$, if all $q \in A$ hold in w then p holds in w (i.e. p holds "whenever" all $q \in A$ hold). For instance, $a, b, (a \wedge b) \rightarrow c \vDash c$, but $\neg a \nvDash a \rightarrow b$ and $b \nvDash a \rightarrow b$ (so C^+ does not suffer the paradoxes of material implication). Recall that $A \subseteq \mathbf{L}$ is consistent if and only if there is no $p \in \mathbf{L}$ with $A \vDash p$ and $A \vDash \neg p$. So

- A is (C^+-)consistent if and only if there is an interpretation $(W, (f_p), (v_w))$ and a world $w \in W$ in which all $q \in A$ hold (i.e. all $q \in A$ "can" hold simultaneously).

So $\{a, \neg a\}$ is inconsistent: if a holds in a word w, $\neg a$ is false in w. And $\{a, \neg(a \rightarrow b), b\}$ is consistent (but classically inconsistent): let a and b both hold in w and let $f_a(w)$ contain a world w' in which b is false.[13]

4 Simple implication agendas

Given the logic C^+, which quota rules are consistent? I first give an answer for *simple* implication agendas.

[13] This is an example of why C^+ meets our requirement (b) on the treatment of connection rules. To verify (b) in general, apply Lemma 8 to sets A consisting of negated non-degenerate connection rules and of atomic or negated atomic propositions (and note that (20) does not hold since $A \cap \mathcal{R} = \emptyset$).

Theorem 1 *A quota rule $F_{(m_p)_{p\in X^+}}$ for a simple implication agenda X is consistent if and only if*

$$m_b \leq m_a + m_{a\to b} - n \text{ for all } a \to b \in X. \tag{1}$$

So consistent quota rules do exist: putting $m_p = n$ for all $p \in X^+$ validates (1). But this extreme quota rule is far from the only consistent quota rule; for instance, (1) holds if all atomic propositions $a \in X$ get the same threshold m_a (so all issues are treated symmetrically) and all connection rules $a \to b \in X$ get the unanimity threshold $m_{a\to b} = n$ (so links between issues are very hard to accept).

Some consequences of (1) can be expressed in terms of the network structure of the (simple) implication agenda X (see Figure 1 for examples of network structures). The nodes are the atomic propositions in X, and if $a \to b \in X$ then a is a *parent* of b and b a *child* of a. The notions of *ancestor* and *descendant* follow by transitive closure. By (1), $m_a \geq m_b$ if a is a parent, or more generally an ancestor, of b. In particular, $m_a = m_b$ if a and b are in a cycle, i.e. are ancestors of each other. In short, thresholds of atomic propositions weakly decrease along (descending) paths, and are constant within cycles. Cycles severely restrict the thresholds not only of its member propositions but also of the connection rules $a \to b$ linking them: we must have $m_{a\to b} = n$, as is seen by setting $m_a = m_b$ in (1).

The picture changes radically if we misrepresent the decision problem by using classical logic: then there exists at most one and typically no consistent quota rule $F_{(m_p)_{p\in X^+}}$, where "consistent" now means *classically* consistent and the (universal) domain of $F_{(m_p)_{p\in X^+}}$ now consists of the profiles of complete and *classically* consistent judgment sets.[14] More precisely, the classical counterpart of Theorem 1 is the following result.

Theorem 1* *Defining logical interconnections using classical logic, a quota rule $F_{(m_p)_{p\in X^+}}$ for a simple implication agenda X is consistent if and only if*

$$m_a = n \text{ and } m_{a\to b} = m_b = 1 \text{ for all } a \to b \in X. \tag{2}$$

So there is *no* classically consistent quota rule if X contains a "chain" $a \to b, b \to c$; and there is a *single* (unnatural) one otherwise.

Each of the two theorems can be proven in two steps: step 1 identifies possible types/sources of inconsistency, and step 2 shows that (1) (respectively, (2)) is necessary and sufficient to prevent these types of inconsistency.

More precisely, Theorem 1 follows from the following two lemmas (steps) by noting that any collective judgment set $A \subseteq X$ generated by a quota rule satisfies:

$$A \text{ contains exactly one member of each pair } p, \neg p \in X. \tag{3}$$

Lemma 1 *For a simple implication agenda X, a set $A \subseteq X$ satisfying (3) is consistent if and only if it contains no triple $a, a \to b, \neg b \in X$.*

[14] Within a simple implication agenda X, the classical logical interconnections are stronger than the C^+ ones: all classically consistent sets $A \subseteq X$ are C^+-consistent but not vice versa (see Lemmas 1, 2). So a consistent quoto rule's (universal) domain and co-domain shrink by moving to classical logic.

Lemma 2 *For a simple implication agenda X, a quota rule $F_{(m_p)_{p\in X^+}}$ never accepts any triple $a, a \to b, \neg b \in X$ if and only if (1) holds.*[15]

Analogously, Theorem 1* follows from the following two lemmas (steps).

Lemma 1* *For a simple implication agenda X, a set $A \subseteq X$ satisfying (3) is classically consistent if and only if it contains no triple $a, a \to b, \neg b \in X$ or pair $b, \neg(a \to b) \in X$ or pair $\neg a, \neg(a \to b) \in X$.*

Lemma 2* *For a simple implication agenda with the logical interconnections of classical logic, a quota rule $F_{(m_p)_{p\in X^+}}$ never accepts any triple $a, a \to b, \neg b \in X$ or pair $b, \neg(a \to b) \in X$ or pair $\neg a, \neg(a \to b) \in X$ if and only if (2) holds.*

Lemmas 1 and 1* highlight the difference between non-classical and classical logic: the latter creates two additional types of inconsistency (in simple implication agendas). These additional inconsistencies are artificial; e.g. b is intuitively consistent with $\neg(a \to b)$. By Lemma 2, (1) is necessary and sufficient to exclude all *non-classical* inconsistencies $a, a \to b, \neg b \in X$. But (1) does nothing to prevent the artificial classical inconsistencies.[16] To prevent also these, (1) must be strengthened to (2) by Lemma 2*.

I first prove Lemmas 1 and 1*, in reverse order to start simple.

Proof Lemma 1.* Let X and A be as specified. Clearly, if A contains a triple $a, a \to b, \neg b$ or pair $b, \neg(a \to b)$ or pair $\neg a, \neg(a \to b)$, then A is classically inconsistent. Now suppose A does not contain such a triple or pair. To show A's classical consistency, I define a classical interpretation $v : \mathbf{L} \to \{T, F\}$ that affirms all $p \in A$. Define v by the condition that the only true atomic propositions are those in A. Then all atomic or negated atomic members of A are true. Further, every $a \to b \in A$ is true: as A does not contain the triple $a, a \to b, \neg b$, A either contains b, in which case b is true, hence $a \to b$ is true; or A contains $\neg a$, in which case a is false, hence $a \to b$ is true. Finally, every $\neg(a \to b) \in A$ is true: as A contains neither the pair $b, \neg(a \to b)$ nor the pair $\neg a, \neg(a \to b)$, A contains neither b nor $\neg a$, so that b is false and a true, and hence $a \to b$ is false, i.e. $\neg(a \to b)$ is true.

Proof of Lemma 1. Let X and $A \subseteq X$ be as specified. If $A \subseteq X$ contains a triple $a, a \to b, \neg b$, A is of course (C^+-)inconsistent. Now assume A contains no triple $a, a \to b, \neg b$. To show that A is consistent (though perhaps classically inconsistent), I specify a C^+-interpretation $(W, (f_p), (v_w))$ with a world $\bar{w} \in W$ in which all $p \in A$ hold. Let W contain:

(a) a world $\bar{w}$, in which an atomic proposition a holds if and only if $a \in A$;
(b) for every atomic proposition a, a world w_a ($\neq \bar{w}$) such that
- $f_a(\bar{w}) = \{w_a\}$ if $a \notin A$ and $f_a(\bar{w}) = \{\bar{w}, w_a\}$ if $a \in A$ (so "if a were true" refers to w_a, and as required by the notion of a C^+-interpretation also to the actual world $\bar{w}$ if a holds there, i.e. if $a \in A$);

[15] "Never" of course means "for no profile in the (universal) domain of $F_{(m_p)_{p\in X^+}}$".

[16] For instance, the pair $b, \neg(a \to b) \in X$ is collectively accepted if $m_b < m_{a\to b}$ (which (1) allows) and if m_b persons accept the pair $b, a \to b$ and all others accept the pair $\neg b, \neg(a \to b)$.

- in w_a exactly those atomic propositions b are false for which $\neg(a \to b) \in A$.

We have to convince ourselves that all $p \in A$ hold in $\bar{w}$. All atomic or negated atomic $p \in A$ hold in $\bar{w}$ by (a). Also any negated implication $\neg(a \to b) \in A$ holds in $\bar{w}$: by (b), "if a were true" refers to w_a, in which b is false; whence in $\bar{w}$ $a \to b$ is false, i.e. $\neg(a \to b)$ true. Finally, suppose $a \to b \in A$. I have to show that b holds in all worlds $w \in f_a(\bar{w})$. There are two cases.

Case 1: $a \in A$. Then $f_a(\bar{w}) = \{\bar{w}, w_a\}$ by (b). First, b holds in $\bar{w}$: otherwise $b \notin A$ (by (a)), so that A would contain the triple $a, a \to b, \neg b$, a contradiction. Second, b holds in w_a: otherwise $\neg(a \to b) \in A$ (by (b)), contradicting $a \to b \in A$.

Case 2: $a \notin A$. Then $f_a(\bar{w}) = \{w_a\}$; and b holds in w_a, as just mentioned. ■

I now show Lemmas 2 and 2*, completing the proof of Theorems 1 and 1*.

Proof of Lemma 2. Let $F_{(m_p)_{p \in X^+}}$ be a quota rule for a simple implication agenda X. Take a given triple $a, a \to b, \neg b \in X$. I consider all profiles for which a and $a \to b$ are collectively accepted, and I show that b is collectively accepted (i.e. $\neg b$ rejected) for all such profiles *if and only if* $m_b \leq m_a + m_{a \to b} - n$.

Note first that in all such profiles at least m_a people accept a and at least $m_{a \to b}$ people accept $a \to b$; hence the number of people accepting both these propositions (hence also b) is at least $m_a + m_{a \to b} - n$ (in fact, at least $\max\{m_a + m_{a \to b} - n, 0\}$). Thus, if $m_b \leq m_a + m_{a \to b} - n$, b is in all such profiles accepted by at least m_b people, hence collectively accepted.

For the converse, note that among such profiles there is one such that exactly $\max\{m_a + m_{a \to b} - n, 0\}$ people accept both a and $a \to b$ (hence b) and such that no one else accepts b. If $m_b > m_a + m_{a \to b} - n$, then in this profile less than m_b people accept b, so that b is collectively rejected. ■

Proof of Lemma 2.* Let $F_{(m_p)_{p \in X^+}}$ be a quota rule for a simple implication agenda X, with the *classical* logical interconnections. Lemma 2 (and its proof) also holds under classical logic; so $F_{(m_p)_{p \in X^+}}$ never accepts any triple $a, a \to b, \neg b \in X$ if and only if (1) holds. Further, a given pair $b, \neg(a \to b) \in X$ is never accepted if and only if $m_b \geq m_{a \to b}$: necessity of $m_b \geq m_{a \to b}$ follows from footnote 16, and sufficiency holds because, as b classically entails $a \to b$, $a \to b$ is (in any profile) accepted by at least as many people as b. By an analogous argument, a given pair $\neg a, \neg(a \to b) \in X$ is never accepted if and only if $m_{a \to b} \leq n - m_a + 1$ $(= m_{\neg a})$. In summary, we thus have three inequalities for every $a \to b \in X$: that in (1), $m_b \geq m_{a \to b}$, and $m_{a \to b} \leq n - m_a + 1$. Together these inequalities are equivalent to the condition in (2), as is easily checked. ■

Constructing consistent quota rules. I now discuss how to choose thresholds $(m_p)_{p \in X^+}$ that satisfy (1), for a simple implication agenda X. The notions of a child/parent and a descendant/ancestor are defined above. A *path* is a sequence $(a_1, a_2, ..., a_k)$ in X $(k \geq 2)$ in which each a_j is a parent of a_{j+1} $(j < k)$. X is *acyclic* if it has no cycle, i.e. no path $(a_1, ..., a_k)$ with $a_1 = a_k$. The *depth* of X is $d_X := \sup\{k$: there is a path in X of length $k\}$, and the *level* of an atomic proposition $a \in X$ is $l_a := \sup\{k$: there is a path in X of length k ending with $a\}$, interpreted as 1 if no path ends with a. So $a \in X$ has level 1 if it has no parents, level 2 if it has parents

all of which have level 1, etc. Figure 3 shows an acyclic simple implication agenda with three levels.

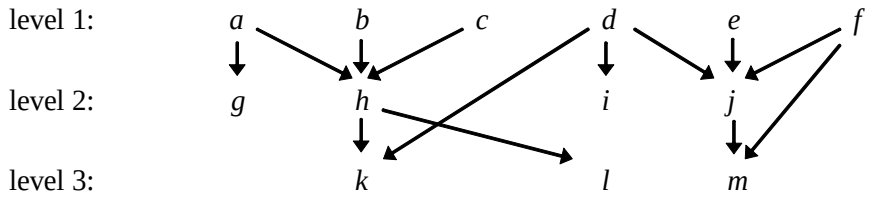

Figure 3: An acyclic simple implication agenda X of depth $d_X = 3$.

How free are we in choosing the thresholds $(m_p)_{p\in X^+}$? Clearly, by (1) the thresholds of atomic propositions must weakly decrease along any path. If X is acyclic and finite (hence of finite depth d_X), $(m_p)_{p\in X^+}$ can be chosen recursively in the following d_X steps.

Step l $(= 1, 2, ..., d_X)$**:** for all $b \in X$ of level l, choose a threshold $m_b \in \{1, ..., n\}$ and thresholds $m_{a\to b} \in \{1, ..., n\}$ for the parents a of b, such that

$$m_b \leq m_a + m_{a\to b} - n \text{ for all parents } a \text{ of } b. \tag{4}$$

But this procedure may involve choosing many thresholds: in Figure 3, those of 13 atomic propositions and 13 implications! To reduce complexity, one might use

- the same threshold $m = m_{a\to b}$ for all connection rules $a \to b \in X$, where m reflects how easily the group imposes constraints between issues,
- the same threshold m_l for all propositions in X with the same level l $(\in \{1, ..., d_X\})$, where m_l reflects how easily the group accepts level l propositions.

I write such a quota rule as $F_{m,m_1,...,m_{d_K}}$. Here, only $d_X + 1$ parameters must be chosen, e.g., in Figure 3, $3 + 1 = 4$ parameters instead of 26. Applied to quota rules of this type, Theorem 1 yields the following characterisation, by a proof left to the reader.

Corollary 1 *For a finite acyclic simple implication agenda X, a quota rule $F_{m,m_1,...,m_{d_K}}$ is consistent if and only if*

$$m_l \leq m_{l-1} + m - n \text{ for all levels } l \in \{2, ..., d_X\}. \tag{5}$$

Consistent quota rules of type $F_{m,m_1,...,m_{d_K}}$ can be constructed as follows.

Step 0: choose $m \in \{1, ..., n\}$ such that (i) $m \geq n - (n-1)/(d_X - 1)$.
Step l $(= 1, 2, ..., d_X)$**:** choose $m_l \in \{1, ..., n\}$ such that (ii) $m_l \geq 1 + (d_X - l)(n - m)$ and (iii) $m_l \leq m_{l-1} + m - n$ if $l > 1$.

The conditions (i)-(iii) follow from Corollary 1: (iii) is obvious, and (i) and (ii) make the choices in future steps possible.[17] For a group of size $n = 10$ and the agenda of Figure 3, a consistent quota rule F_{m,m_1,m_2,m_3} might be chosen as follows.

Step 0: $m = 8$ (note that $8 \geq n - (n-1)/(d_X - 1) = 10 - 9/2 = 5.5$).
Step 1: $m_1 = 8$ (note that $8 \geq 1 + (d_X - 1)(n - m) = 1 + 2 \times 2 = 5$).
Step 2: $m_2 = 6$ (note that $6 \geq 1 + (d_X - 2)(n - m) = 1 + 2 = 3$ and $6 \leq m_1 + m - n = 8 + 8 - 10 = 6$).
Step 3: $m_3 = 4$ (note that $4 \geq 1 + (d_X - 3)(n - m) = 1$ and $4 \leq m_2 + m - n = 6 + 8 - 10 = 4$).

Causal and justificational interpretation. I now offer two interpretations of connection rules, and hence of the kind of decision problems captured by implication agendas. For simplicity, I restrict myself to a *simple* implication agendas X.

First, suppose implications $a \to b \in X$ have a *causal* status: $a \to b$ means that fact a *causes* fact b. So X might contain "*if* the ozone hole has size X *then* global warming will continue" and "*if* global warming will continue *then* species Y will die out". Then X captures a decision problem of forming beliefs about facts and their causal links. A path $(a_1, ..., a_k)$ in X is a *causal chain* (assuming the causal links $a_1 \to a_2, ..., a_{k-1} \to a_k$ hold), and the level of a proposition indicates how "causally fundamental" it is. By an earlier remark, Theorem 1 implies that the acceptance threshold must weakly decrease along any causal chain.

Second, suppose the implications $a \to b \in X$ have a *justificational* (or *evidential* or *indicative*) status: $a \to b$ means that a *indicates* b (a can indicate b without causing b: a wet street indicates rain without causing it). So X captures a decision problem of forming beliefs about claims/statements/hypotheses and their justificational links. Some claims may have a *normative* content, like "a multi-cultural society is desirable" or "option x is better than option y". For instance, an environmental panel might decide on a : "the ozone hole has size larger than X", b : "tax T on kerosine should be introduced", and the justificational link $a \to b$. A path $(a_1, ..., a_k)$ is an "argumentative" chain (assuming the links $a_1 \to a_2, ..., a_{k-1} \to a_k$ hold), and the level of a proposition reflects how "argumentatively fundamental" it is. Often, high level propositions are more concrete and might state that certain collective acts should be taken (a road should be built, a firm downsized, a law amended, etc.), whereas their ancestors describe potential *reasons* or *arguments*, either of a descriptive kind (traffic will increase, demand will fall, etc.) or of a normative kind (multi-culturalism is desirable, etc.). Of course, one may reject a reason $a \in X$, or reject a's status as reason for $b \in X$ (i.e. reject $a \to b$). Again, reasons need at least as high acceptance thresholds as their (argumentative) descendents, e.g. $\frac{3}{4}n$ versus $\frac{1}{2}n$.

5 Other special implication agendas

The difference between using non-classical and using classical logic is now further illustrated by considering two other types of implication agendas X, namely

[17] For instance, without (i) there would be *no* choices of $m_1, ..., m_{d_X}$ satisfying (5).

- *semi-simple* ones: here all connection rules in X are implications $p \to b$ in which b is atomic (such as $(a \wedge c) \to b$ but not $a \to (b \wedge c)$);
- *bi-simple* ones: here all connection rules in X are bi-implications $a \leftrightarrow b$ in which a and b are atomic.

For each of these agenda types, we again perform two steps (analogous to the steps performed for simple implication agendas):

(step 1) we identify possible types/sources of inconsistency in the agenda;
(step 2) we exclude each one by an inequality on thresholds and if possible we simplify the system of inequalities.

This gives a consistency condition for quota rules $F_{(m_p)_{p\in X^+}}$, in analogy to Theorem 1. Again, using instead classical logic leads in step 1 to additional (artificial) types of inconsistency; so that, in analogy to Theorem 1*, at most one (degenerate) quota rule $F_{(m_p)_{p\in X^+}}$ is classically consistent (if X is semi-simple), or even no one (if X is bi-simple). Table 2 summarises the results for each agenda type

Agenda type	A set $A \subseteq X$ satisfying (3) is consistent *iff* it has no subset of type(s)...	$F_{(m_p)_{p\in X^+}}$ is consistent *iff*...
simple	$\{a, a \to b, \neg b\}$ in CL also: $\{b, \neg(a \to b)\}, \{\neg a, \neg(a \to b)\}$	(1) in CL: (2)
semi-simple	$\mathrm{C}(p) \cup \{p \to b, \neg b\}$ in CL also: $\{b, \neg(p \to b)\}, \{\neg a, \neg(p \to b)\}$ $(a \in \mathrm{C}(p))$	(7) in CL: (8)
bi-simple	$\{a, \neg b, a \leftrightarrow b\}, \{\neg a, b, a \leftrightarrow b\}, \{a \leftrightarrow b, \neg(b \leftrightarrow a\}\}$ in CL also: $\{a, b, \neg(a \leftrightarrow b)\}, \{\neg a, \neg b, \neg(a \leftrightarrow b)\}$	(9) in CL: never

Table 2: Three types of implication agendas X, their types of inconsistencies, and their consistent quota rules; in non-classical logic and in classical logic ("CL")

In Table 2, the results for simple X were shown in the last section. Regarding semi- or bi-simple X, we have to adapt Lemmas 1 and 2 (or 1* and 2* for classical logic). Let me briefly indicate how this works. First a general remark. The types of inconsistency in step 1 can be (and are in Table 2) identified with certain inconsistent sets $Y \subseteq X$ (which are *minimal inconsistent*, in fact *irreducible*; see Section 7); and the inequality needed in step 2 to exclude Y's acceptance can always be written as

$$\sum_{p\in Y}(n - m_p) < n, \text{ or equivalently } \sum_{p\in Y} m_p > n(|Y| - 1) \tag{6}$$

(see Lemma 5 in Section 7). Intuitively, (6) requires the propositions in Y to have sufficiently high acceptance thresholds to prevent joint acceptance of all $p \in Y$.

First let X be semi-simple. In step 1 we have to consider not only inconsistent sets of type $\{a, a \to b, \neg b\} \subseteq X$ (as for simple X) but also ones like $\{a, c, (a \wedge c) \to b, \neg b\} \subseteq X$. By adapting Lemma 1 to semi-simple agendas, the inconsistent sets in step 1 turn out to be precisely the sets $\mathrm{C}(p) \cup \{p \to b, \neg b\} \subseteq X$. In the proof of Lemma 1, the C^+-interpretation $(W, (f_p), (v_w))$ should be adapted by letting W contain:

(a) a world $\bar{w}$, in which an atomic proposition a holds iff $a \in A$ (as before)

(b) for any conjunction p of atomic propositions a world w_p ($\neq \bar{w}$) such that

- $f_p(\bar{w}) = \{w_p\}$ if $\mathrm{C}(p) \not\subseteq A$ and $f_p(\bar{w}) = \{\bar{w}, w_p\}$ if $\mathrm{C}(p) \subseteq A$ (so "if p were true" refers to world w_p, and also to the actual world $\bar{w}$ if p holds there);
- in w_p exactly those atomic propositions b are false for which $\neg(p \to b) \in A$.

The rest of Lemma 1 – showing that all $p \in A$ hold in $\bar{w}$ – is easily adapted. Using (6), it then follows that a quota rule $F_{(m_p)_{p\in X^+}}$ is consistent if and only if $\sum_{a\in \mathrm{C}(p)\cup\{p\to b,\neg b\}}(n - m_a) < n$ for all $p \to b \in X$; or equivalently

$$\sum_{a\in \mathrm{C}(p)} (n - m_a) + m_b \leq m_{p\to b} \text{ for all } p \to b \in X. \tag{7}$$

Note that this characterisation indeed reduces to Theorem 1 if X is simple.

By contrast, classical logic leads in step 1 to new inconsistent sets of type $\{b, \neg(p \to b)\} \subseteq X$ and $\{\neg a, \neg(p \to b)\}$ with $a \in \mathrm{C}(p)$, as is seen by adapting Lemma 1* (without even having to redefine the classical interpretation $v : \mathbf{L} \to \{T, F\}$). As a result, a quota rule $F_{(m_p)_{p\in X^+}}$ is classically consistent if and only if

$$m_a = n \text{ and } m_{p\to b} = m_b = 1 \text{ for all } p \to b \in X \text{ and all } a \in \mathrm{C}(p). \tag{8}$$

Again, this characterisation reduces to Theorem 1* if X is simple.

Now let X be bi-simple. In step 1, the sets of type $\{a, \neg b, a \leftrightarrow b\}$ or $\{\neg a, b, a \leftrightarrow b\}$ or $\{a \leftrightarrow b, \neg(b \leftrightarrow a\}\}$ capture all types of (non-classical) inconsistency. This can be shown by again adapting Lemma 1 and its proof; when defining the C^+-interpretation $(W, (f_p), (v_w))$, we simply have to replace the second bullet point of (b) by:

- in the world w_a (to which "if a were true" refers), exactly those atomic propositions b are false for which $\neg(a \leftrightarrow b) \in A$ or $\neg(b \leftrightarrow a) \in A$.

By adapting Lemma 2 and its proof (or by using (6) and that $m_{\neg q} = n - m_q + 1$ for all $q \in X$), a quota rule $F_{(m_p)_{p\in X^+}}$ is seen to be consistent if and only if, for all $a \leftrightarrow b \in X$, $m_b \leq m_a + m_{a\leftrightarrow b} - n$ and $m_a \leq m_b + m_{a\leftrightarrow b} - n$ and if also $b \leftrightarrow a \in X$ $m_{a\leftrightarrow b} \geq m_{b\leftrightarrow a}$; which is equivalent to:

$$m_{a\leftrightarrow b} = n \text{ and } m_a = m_b \text{ for all } a \leftrightarrow b \in X. \tag{9}$$

So all bi-implications need the unanimity threshold, and two atomic propositions ("issues") need the same threshold if they are linked by a bi-implication in X or, more generally, by a path of bi-implications in X.

In contrast, classical logic leads (by adapting Lemma 1*) to the additional types of inconsistency $\{a, b, \neg(a \leftrightarrow b)\}, \{\neg a, \neg b, \neg(a \leftrightarrow b)\}$ (which are artificial since negating $a \leftrightarrow b$ shouldn't constrain a's and b's truth values: it shouldn't establish the constraint $a \leftrightarrow \neg b$). This leads (using (6)) to the additional inequalities

$$m_a + m_b + m_{\neg(a\leftrightarrow b)} > 2n \text{ and } m_{\neg a} + m_{\neg b} + m_{\neg(a\leftrightarrow b)} > 2n \text{ for all } a \leftrightarrow b \in X.$$

In this, we have by (9) $m_{\neg(a\leftrightarrow b)} = 1$, so that $m_a + m_b \geq 2n$ and $m_{\neg a} + m_{\neg b} \geq 2n$, hence $m_a = m_b = m_{\neg a} = m_{\neg b} = n$, a contradiction. So there is *no* classically consistent quota rule.

Transforming implication agendas into semi-simple ones. Semi-simple implication agendas are of special interest. While they exclude from the agenda many connection rules – all uni-directional ones with non-atomic consequent and all bi-directional ones – each connection rule of the excluded type can be rewritten, in logically equivalent terms, as a conjunction of connection rules of the non-excluded type $p \to b$ with atomic b: indeed, each uni-directional connection rule $p \to q$ is equivalent to the conjunction $\wedge_{b \in \mathrm{C}(q) \setminus \mathrm{C}(p)}(p \to b)$, and each bi-directional connection rule $p \leftrightarrow q$ is equivalent to the conjunction $\big(\wedge_{b \in \mathrm{C}(q) \setminus \mathrm{C}(p)}(p \to b)\big) \wedge \big(\wedge_{b \in \mathrm{C}(p) \setminus \mathrm{C}(q)}(q \to b)\big)$. So every implication agenda X can be transformed into a semi-simple one $\tilde{X}$ by replacing every "non-allowed" connection rule $r \in X^+$ by all the "allowed" ones $p \to b$ of which r is a conjunction (up to logical equivalence). For instance, the implication agenda X given by $X^+ = \{a, b, c, c \leftrightarrow (a \wedge b)\}$, which models the judges' decision problem in a law suit (see Section 2), can be transformed into the semi-simple implication agenda $\tilde{X}$ given by $\tilde{X}^+ = \{a, b, c, c \to a, c \to b, (a \wedge b) \to c\}$; under $\tilde{X}$, the judges decide not *en bloc* on $c \leftrightarrow (a \wedge b)$, but separately on whether liability of the defendant implies breach of the contract, whether liability implies validity of the contract, and whether breach of a valid contract implies liability.

Should we conclude from this that all collective decision problems describable by an implication agenda X, like the mentioned one of judges in a law suit, can be remodelled using the corresponding semi-simple implication agenda $\tilde{X}$? And that we could therefore restrict ourselves to the semi-simple case? Not quite, because the change of agenda alters the decision problem. More precisely, it refines (i.e. augments) the decision problem: indeed, from any (complete and consistent) judgment set for $\tilde{X}$ we can always derive a unique one for X, but not vice versa. In the example just given, the judgments on the "new" connection rules $c \to a, c \to b, (a \wedge b) \to c \in \tilde{X}$ together imply a judgment on the "old" one $c \leftrightarrow (a \wedge b) \in X$, but not vice versa because if $c \leftrightarrow (a \wedge b)$ is negated we do not know which one(s) of $c \to a, c \to b, (a \wedge b) \to c$ to negate (we only know that at least one of them must be negated). In summary, it *is* true that the decision problem described by X can be settled by moving to the semi-simple agenda $\tilde{X}$, *but* one thereby settles more and one uses richer in- and output information in the aggregation.

6 General implication agendas

Many implication agendas are of neither of the kinds analysed so far, because they contain connection rules like $a \to (b \wedge c)$ or $(a \wedge b) \leftrightarrow (a \wedge c)$. Which quota rules are consistent for *general* implication agendas (in the non-classical logic C^+)? In principle, the above two-step procedure applies again. But, for an agenda class as rich as this one, a so far neglected question becomes pressing: what is it that makes an inconsistent set $Y \subseteq X$ a "type of inconsistency" (in step 1)? Why for instance did we count sets $\{a, a \to b, \neg b\} \subseteq X$ but not sets $\{a, a \to b, b \to c, \neg c\} \subseteq X$ as types of inconsistency for simple implication agendas X? Surely, the set $\mathcal{Y}$ of all types of inconsistency $Y \subseteq X$ must, to enable step 1, be chosen such that

$$\textit{every inconsistent set } A \subseteq X \textit{ satisfying (3) has a subset in } \mathcal{Y}. \tag{10}$$

But usually many choices of $\mathcal{Y}$ satisfy (10). Intuitively, it is useful to choose $\mathcal{Y}$ small and simple. An always possible – but often unduly large – choice of $\mathcal{Y}$ is to include

in $\mathcal{Y}$ all *minimal inconsistent* sets $Y \subseteq X$.[18] For simple implication agendas X, we chose $\mathcal{Y} = \{\{a, a \to b, \neg b\} : a \to b \in X\}$, although we could have also included minimal inconsistent sets of type $\{a, a \to b, b \to c, \neg c\} \subseteq X$. We were able to exclude such sets (and still satisfy (10)) because such sets are *reducible* in the following sense. For a set $A \subseteq X$ satisfying (3), if $A \supseteq \{a, a \to b, b \to c, \neg c\}$ then, as A contains b or $\neg b$, either $A \supseteq \{b, b \to c, \neg c\}$ or $A \supseteq \{a, a \to b, \neg b\}$, whence A has a subset in $\mathcal{Y}$.

In Section 7, a general method to choose $\mathcal{Y}$ is developed, based on a formalisation of what it means to "reduce" an inconsistent set to a simpler one; $\mathcal{Y}$ then contains *irreducible* sets. Applied to implication agendas, the method yields two kinds of irreducible sets, i.e. two types of inconsistency (as shown in the appendix[19]):

(Ir_+) sets representing an inconsistency between a *non-negated* connection rule and atomic or negated atomic propositions, like $\{a \to (b \wedge c), a, \neg b\}$ or $\{a \leftrightarrow b, \neg a, b\}$;

(Ir_-) sets representing an inconsistency between a *negated* connection rule and non-negated connection rules, like $\{\neg(a \to (b \wedge c)), a \to b, a \to c\}$ or $\{\neg(a \to (b \wedge c \wedge d)), a \to (b \wedge c), a \leftrightarrow d\}$.

In step 2, these irreducible sets yield a system of inequalities whose successive simplification gives the characterisation of Theorem 2 below. This characterisation involves, for every $p \to q \in X$, a particular set $X_{p \to q}$. This set is defined in two steps. First, we form the set

$$X_p := \{s \in \mathbf{L} : p \to s \in X \text{ or } p \leftrightarrow s \in X \text{ or } s \leftrightarrow p \in X\}$$

of all propositions "reachable" from p via (bi-)implications in X. From X_p we then form the set

$$X_{p \to q} := \{S \subseteq X_p : S \text{ is minimal subject to } \mathrm{C}(q) \backslash \mathrm{C}(p) \subseteq \cup_{s \in S} \mathrm{C}(s)\}$$

of all sets $S \subseteq X_p$ that have, and are minimal subject to, this property: each atomic proposition "in" q (but not "in" p) is "in" some $s \in S$. So the sets $S \in X_{p \to q}$ minimally "cover" $\mathrm{C}(q) \backslash \mathrm{C}(p)$.

Evaluating X_p and $X_{p \to q}$ is purely mechanical. As a first example, suppose

$$X^+ = \{a, b, c, a \to b, a \to c, a \to (b \wedge c)\}. \tag{11}$$

Here all three implications have antecedent a, where $X_a = \{b, c, b \wedge c\}$. From X_a we then derive $X_{a \to b}, X_{a \to c}$ and $X_{a \to (b \wedge c)}$. For instance, $X_{a \to b}$ contains $\{b\} \subseteq X_a$ and $\{b \wedge c\} \subseteq X_a$ as both minimally "cover" b, but contains neither $\{c\} \subseteq X_a$ (which fails to "cover" b), nor $\{b, c\} \subseteq X_a$ (which "covers" b *non*-minimally as we can remove c), nor any other set $S \subseteq X_a$. Further, $X_{a \to (b \wedge c)}$ does *not* contain $\{c, b \wedge c\} \subseteq X_a$: although this set "covers" $b \wedge c$ (as all atomic propositions "in" $b \wedge c$ are "in" some $s \in \{c, b \wedge c\}$), it does so non-minimally (as c can be removed); but $X_{a \to (b \wedge c)}$ contains $\{b, c\}$ and $\{b \wedge c\}$ (which "cover" $b \wedge c$ minimally). In summary,

$$X_{a \to b} = \{\{b\}, \{b \wedge c\}\}, X_{a \to c} = \{\{c\}, \{b \wedge c\}\}, X_{a \to (b \wedge c)} = \{\{b, c\}, \{b \wedge c\}\}. \tag{12}$$

[18] Y is *minimal inconsistent* if Y is inconsistent but its proper subsets are consistent.

[19] In fact, each type has two subtypes, one for uni- and one for bi-directional connection rules.

As a second example, suppose

$$X^+ = \{a, b, c, a \to b, a \to (b \wedge c), c \leftrightarrow a\}. \tag{13}$$

The two implications, $a \to b$ and $a \to (b \wedge c)$, both have antecedent a, where $X_a = \{b, c, b \wedge c\}$. From X_a we then derive that:

$$X_{a\to b} = \{\{b\}, \{b \wedge c\}\},\ X_{a\to(b\wedge c)} = \{\{b, c\}, \{b \wedge c\}\}. \tag{14}$$

Sets $X_{p\to q}$ appear in Theorem 2 because they are needed to describe inconsistencies of type (Ir$_-$). Let me give an intuition for why the sets $X_{p\to q}$ relate to inconsistencies of type (Ir$_-$) (details are in the appendix). For the agenda (11), $Y = \{\neg(a \to (b \wedge c)), a \to b, a \to c\}$ is an inconsistency of type (Ir$_-$). Y is inconsistent precisely because the conjuncts of $a \wedge b$ are "covered" by the set of consequents of $a \to b, a \to c \in Y$, i.e. by $\{b, c\}$; in fact, they are so minimally: $\{b, c\} \in X_{a\to(b\wedge c)}$. Another agenda X might have the inconsistency of type (Ir$_-$) $\{\neg(a \to (b \wedge c \wedge d)), a \to (b \wedge c), a \leftrightarrow d\}$. This set is inconsistent precisely because the conjuncts of $b \wedge c \wedge d$ are "covered" by the set of consequents $\{b \wedge c, d\}$; they are so minimally: $\{b \wedge c, d\} \in X_{a\to(b\wedge c\wedge d)}$.

I now state the characterisation result (formally proven in the appendix). As usual, $A\Delta B$ denotes the symmetric difference $(A\backslash B) \cup (B\backslash A)$ of sets A and B.

Theorem 2 *A quota rule $F_{(m_p)_{p\in X^+}}$ for an implication agenda X is consistent if and only if the thresholds satisfy the following:*

(a) *for every $p \to q \in X$,*

$$\sum_{a\in \mathrm{C}(p)} (n - m_a) + \max_{b\in \mathrm{C}(q)\backslash \mathrm{C}(p)} m_b \leq m_{p\to q} \leq n - \max_{S\in X_{p\to q}} \sum_{s\in S: p\to s\in X} (n - m_{p\to s});$$

(b) *for every $p \leftrightarrow q \in X$, (i) $m_{p\leftrightarrow q} = n$, (ii) $m_a = n$ for all $a \in \mathrm{C}(p) \cap \mathrm{C}(q)$, and (iii) m_a is the same for all $a \in \mathrm{C}(p)\Delta\mathrm{C}(q)$ and equals n if $|\mathrm{C}(p)\Delta\mathrm{C}(q)| \geq 3$.*

Theorem 2 characterises consistent quota rules by complicated (in)equalities. A rough interpretation is:

- inconsistencies of type (Ir$_+$) are prevented by the LHS inequalities of (a) and by (b);
- given the LHS inequalities of (a) and (b), inconsistencies of type (Ir$_-$) are prevented by the RHS inequalities in (a).

More detailed clues to understand the conditions (a) and (b) are given at the section end, drawing on the insights gained above on the simple, semi-simple and bi-simple case.

In practice, the system (a)&(b) often simplifies. Part (a) or part (b) drops out if X contains no uni- or no bi-directional connection rules, respectively. If X is simple, semi-simple or bi-simple, (a)&(b) reduces to the conditions derived earlier (namely (1), (7) or (9), respectively).[20] Further, the system (a)&(b) may simplify once the

[20] If X is simple, this is so because (b) drops out and because in (a) the RHS inequality holds trivially (by $X_{p\to q} = \{\{q\}\}$) and the LHS inequality reduces to $n - m_p + m_q \leq m_{p\to q}$.

concrete sets $X_{p\to q}$, $p \to q \in X$, are inserted, possibly resulting in a simpler set of conditions that offers an intuition for the size and structure of the space of possible threshold assignments. The next example demonstrate this.

Example. Consider the agenda in (13). Which thresholds $(m_p)_{p\in X^+}$ guarantee consistency? By Theorem 2, three conditions must hold: one for $a \to b$ (part (a)), one for $a \to (b \wedge c)$ (part (a)), and one for $c \leftrightarrow a$ (part (b)). The three conditions are:

$$\begin{cases} n - m_a + m_b \leq m_{a\to b} \leq n - \max_{S\in X_{a\to b}} \sum_{s\in S: a\to s\in X}(n - m_{a\to s}) \\ n - m_a + \max\{m_b, m_c\} \leq m_{a\to(b\wedge c)} \leq n - \max_{S\in X_{a\to(b\wedge c)}} \sum_{s\in S: a\to s\in X}(n - m_{a\to s}) \\ m_{c\leftrightarrow a} = n \text{ and } m_a = m_c. \end{cases} \tag{15}$$

In this system, the upper bounds of $m_{a\to b}$ and $m_{a\to(b\wedge c)}$ – I call them $B_{a\to b}$ and $B_{a\to(b\wedge c)}$, respectively – should be computed by inserting the sets $X_{a\to b}$ and $X_{a\to(b\wedge c)}$ as given in (14). Then $B_{a\to b}$ and $B_{a\to(b\wedge c)}$ greatly simplify (and turn out to be equal) because each summation "$\sum_{s\in S: a\to s\in X}$" runs over just one term:

$$\begin{aligned} B_{a\to b} &= n - \max\left\{ \sum_{s\in\{b\}: a\to s\in X}(n - m_{a\to s}), \sum_{s\in\{b\wedge c\}: a\to s\in X}(n - m_{a\to s}) \right\} \\ &= n - \max\{n - m_{a\to b}, n - m_{a\to(b\wedge c)}\} = \min\{m_{a\to b}, m_{a\to(b\wedge c)}\}, \\ B_{a\to(b\wedge c)} &= n - \max\left\{ \sum_{s\in\{b,c\}: a\to s\in X}(n - m_{a\to s}), \sum_{s\in\{b\wedge c\}: a\to s\in X}(n - m_{a\to s}) \right\} \\ &= n - \max\{n - m_{a\to b}, n - m_{a\to(b\wedge c)}\} = \min\{m_{a\to b}, m_{a\to(b\wedge c)}\}. \end{aligned}$$

So, in the system (15), the RHS inequalities on the first two lines are jointly equivalent to $m_{a\to b} = m_{a\to(b\wedge c)}$. By $m_a = m_c$, the LHS inequality on the second line implies $n - m_a + \max\{m_b, m_a\} \leq m_{a\to(b\wedge c)}$, and so $\max\{n - m_a + m_b, n\} \leq m_{a\to(b\wedge c)}$; which (by $m_{a\to(b\wedge c)} \leq n$) implies that $m_{a\to(b\wedge c)} = n$, and that $n - m_a + m_b \leq n$, i.e. $m_b \leq m_a$. Using all this, the system (15) is equivalent to:

$$m_b \leq m_a = m_c \text{ and } m_{a\to b} = m_{a\to(b\wedge c)} = m_{c\leftrightarrow a} = n.$$

This is an example of how the presence of a *bi*-directional connection rule r in X can drastically narrow down the possibility space, especially relative to thresholds of r, of atomic propositions "in" r, and of connection rules logically related to r.

I now record two corollaries of Theorem 2. First, a possibility result follows.[21]

Corollary 2 *For an implication agenda X, there exists*

(i) *a consistent quota rule $F_{(m_p)_{p\in X^+}}$ (hence a consistent, complete, independent, anonymous, monotonic and responsive aggregation rule with universal domain);*

[21] See Section 2 and footnote 11 for the conditions listed in part (i). By a different proof, part (i) holds more generally for any agenda X for which each $p \in X^+$ is atomic or a connection rule (where, unlike for implication agendas, the atomic propositions in X^+ may differ from those contained in the connection rules in X^+).

(ii) *a single consistent quota rule* $F_{(m_p)_{p\in X^+}}$ *with identical thresholds* m_p, $p \in X^+$, *namely the quota rule with a unanimity threshold* $m_p = n$ *for all* $p \in X^+$.

Proof. As (ii) implies (i), I only show (ii). Let X be an implication agenda and $F_{(m_p)_{p\in X^+}}$ a quota rule with identical thresholds $m_p = m$ ($\in \{1, ..., n\}$). If $m = n$ then (a)&(b) hold, implying consistency. Conversely, assume consistency. So (a)&(b) hold. X contains a $p \to q$ or a $p \leftrightarrow q$ (otherwise X would be empty, hence not an agenda). In the second case, $m = n$ by (b). In the first case, the LHS inequality in (a) implies $\sum_{a\in \mathrm{C}(p)}(n - m) + m \leq m$, whence again $m = n$. ■

So there is possibility – but how large is it? That is, how much freedom does Theorem 2 leave us in the choice of thresholds? As I now show, *paths* and *cycles* in X impose rather severe restrictions. Extending earlier definitions from simple to general implication agendas, consider the network over the atomic propositions in X, where an atomic proposition $a \in X$ a *parent* of another one $b \in X$ if there is a $p \to q \in X$ or a $p \leftrightarrow q \in X$ or a $q \leftrightarrow p \in X$ such that $a \in \mathrm{C}(p)$ and $b \in \mathrm{C}(q)\backslash\mathrm{C}(p)$. Parenthood yields the notion of an *ancestor* by transitive closure. A *path* is a sequence $(a_1, ..., a_k)$ ($k \geq 2$) where a_j is a parent of a_{j+1} for all $j < k$; it is a *cycle* if $a_1 = a_k$.

Corollary 3 *Let* $F_{(m_p)_{p\in X^+}}$ *be a consistent quota rule for an implication agenda* X.

(i) *If* $a \in X$ *is an ancestor of* $b \in X$ *then* $m_a \geq m_b$.

(ii) *If* $a, b \in X$ *occur in a cycle (i.e. are ancestors of each other) then* $m_a = m_b$ *and* $m_{p\to q} = n$ *for all* $p \to q \in X$ *with* $a \in \mathrm{C}(p)$ *and* $b \in \mathrm{C}(q)\backslash\mathrm{C}(p)$.

Proof. Let X and $F_{(m_p)_{p\in X^+}}$ be as specified.

(i) Let $a \in X$ be a parent of $b \in X$ (obviously it suffices to consider this case). Then $a \in \mathrm{C}(p)$ and $b \in \mathrm{C}(q)\backslash\mathrm{C}(p)$, where $p \to q \in X$ or $p \leftrightarrow q \in X$ or $q \leftrightarrow p \in X$. In the last two cases, (b) implies $m_a \geq m_b$. In the first case, the LHS inequality in (a) implies $(n - m_a) + m_b \leq m_{p\to q}$, so $m_b \leq m_{p\to q} - n + m_a \leq m_a$.

(ii) Let a, b be as specified. By (i) $m_a \leq m_b$ and $m_b \leq m_a$, hence $m_a = m_b$. Now let $p \to q$ be as specified. By the LHS inequality in (a), $(n - m_a) + m_b \leq m_{p\to q}$, hence (by $m_a = m_b$) $m_{p\to q} = n$. ■

An intuition for Theorem 2. Our earlier insights about simple, semi-simple and bi-simple implication agendas offer some clues to understand Theorem 2, more precisely to understand the necessity of (b) and of the LHS of (a). General implication agendas X go in three ways beyond simple ones: (i) implications $p \to q \in X$ may have non-atomic antecedent p; (ii) implications $p \to q \in X$ may have non-atomic consequent q; (iii) X may contain bi-implications $p \leftrightarrow q$.

Here, (i) reminds of semi-simple agendas. And indeed, the LHS of (a), for which (i) is responsible, is closely related to our earlier characterisation (7) of consistent quota rules for semi-simple agendas. To see why, suppose first that in (a) $p \to q$ has atomic consequent q. Then the LHS of (a) coincides with the inequality in (7). Now suppose q is non-atomic. Then $p \to q$ is logically equivalent to the conjunction $\wedge_{b\in \mathrm{C}(q)\backslash\mathrm{C}(p)} p \to b$, and the LHS of (a) is equivalent to applying the inequality in (7) to all implications $p \to b, b \in \mathrm{C}(q)\backslash\mathrm{C}(p)$.

Further, (iii) reminds of bi-simple agendas. Part (b), for which (iii) is responsible, is indeed closely related to our characterisation (9) of consistent quota rules for bi-simple agendas. If in (b) both p and q are atomic, (b) is equivalent to the condition

in (9). If p and/or q is non-atomic, (b) is the right generalisation of (9), as formally shown in the appendix.

Finally, (ii) is the aspect in which general implication agendas go substantially beyond both semi- and bi-simple ones. It is responsible for the (complex) RHS inequalities in (a). These inequalities are needed because (ii) introduces new types of inconsistency like $\{a \to b, a \to c, \neg(a \to (b \wedge c))\}$.

7 An abstract characterisation result

A central issue so far was that each agenda X has its own types/sources of inconsistency $Y \subseteq X$ (e.g. the sets $\{a, a \to b, \neg b\} \subseteq X$ if X is a simple implication agenda). What exactly are "types/sources" of inconsistency? They are *irreducible* sets $Y \subseteq X$, as made precise now. I introduce an abstract *simplicity* relation between inconsistent sets $Y \subseteq X$, which allows one to simplify inconsistent sets, which yields *irreducible* sets. I do this in full generality, i.e. independently of implication agendas and the particular logic C^+. This gives rise to an abstract characterisation result of which all above characterisations are applications. The notion of irreducible sets generalises a special irreducibility notion introduced by Dietrich and List [8]; it also generalises minimal inconsistent sets (which are based on set-inclusion rather than a general simplicity relation), and for this reason the abstract characterisation result generalises the characterisation by the "intersection property" in Nehring and Puppe [18, 19].

To avoid unnecessary restrictions to special judgment aggregation problems, we adopt Dietrich's [3] general logics framework in this section: let $X \subseteq \mathbf{L}$ be an arbitrary agenda of propositions from any formal language $\mathbf{L}$ with well-behaved logical interconnections.[22] Further, let $\mathcal{I}$ be the set of all inconsistent sets $Y \subseteq X$.

Given the intended purpose, I will define irreducibility in such a way that

every inconsistent and complete set $A \subseteq X$ has an irreducible subset. (16)

This property ensures that collective consistency holds if and only if no irreducible set is ever collectively accepted. Property (16) is the analogue of the property (10) underlying the 2-step procedure in earlier sections. Of course, we could achieve (16) by simply defining "irreducible" as "minimal inconsistent", since any inconsistent set $A \subseteq X$ has a minimal inconsistent subset. But this would often create a large number of irreducible sets (hence many redundant inequalities in step 2). The irreducibility notion I introduce depends on a parameter: the *simplicity* notion used. Under a certain (extreme) simplicity notion, "irreducible" will coincide with "minimal inconsistent"; other simplicity notions lead to fewer irreducible sets.

I now define simplicity (from which I later define irreducibility). Suppose we have a notion of simplicity of sets in $\mathcal{I}$ given by a binary relation $<$ on $\mathcal{I}$, where "$Z < Y$"

[22] The well-behavedness can be expressed either in terms of the entailment notion $\models$ (conditions L1-L3 in Dietrich [3]) or in terms of the inconsistency notion (conditions I1-I3 in Dietrich [3]) (assuming that both notions are interdefinable; see Section 2). Stated in terms of the consistency notion, the three conditions are: (I1) sets $\{p, \neg p\} \subseteq \mathbf{L}$ are inconsistent; (I2) subsets of consistent sets are consistent; (I3) the empty set $\emptyset$ is consistent, and each consistent set $A \subseteq \mathbf{L}$ has a consistent superset $B \subseteq \mathbf{L}$ containing a member of each pair $p, \neg p \in \mathbf{L}$. If the agenda X is infinite, I also assume the logic to be compact: every inconsistent set $A \subseteq \mathbf{L}$ has a finite inconsistent subset. All this holds for C^+ and most familiar logics, including propositional and predicate logics, classical and non-classical logics, with the important exception of non-monotonic logics.

is interpreted as "Z is simper than Y". There is much freedom in how to specify $<$ (the goal being to obtain "nice" irreducible sets, as explained later). For instance, we might define $<$ by $Z < Y :\Leftrightarrow |Z| < |Y|$ (i.e. "simpler" means "smaller"), or by $Z < Y :\Leftrightarrow Z \subsetneq_{\text{finite}} Y$ (i.e. "simpler" means to be a proper finite subset). I place only two restrictions on the simplicity notion:

Proper subsets are simpler: for all $Y, Z \in \mathcal{I}$, if $Z \subsetneq_{\text{finite}} Y$ then $Z < Y$.[23]

No infinite simplification chains: $<$ is *well-founded*, i.e. there is no infinite sequence $(Y_k)_{k=1,2,...}$ in $\mathcal{I}$ such that $Y_{k+1} < Y_k$ for all $k = 1, 2, ...$[24]

A *simplicity relation* is a binary relation $<$ on $\mathcal{I}$ with these properties. For instance the two relations $<$ just mentioned are simplicity relations.

Suppose we have chosen a simplicity relation $<$. Then (16) holds for the following reason. Starting from an arbitrary complete set $A \in \mathcal{I}$, one can find a finite sequence of simplifications $A = Y_1 > Y_2 > ...$ such that the simplified sets $Y_1, Y_2, ...$ all remain subsets of A and the sequence terminates with an irreducible set (as defined below). An example is helpful. Suppose $A = Y_1$ is infinite. Then in a first simplification step we can move to a finite inconsistent subset $Y_2 \subseteq A$ (which exists since the logic is compact or X is finite; see footnote 22). To bring the example into familiar terrain, assume that the agenda is an implication agenda and that $Y_2 = \{a, a \to (b_1 \wedge ... \wedge b_5), b_6, (b_1 \wedge .... \wedge b_6) \leftrightarrow c, \neg c\}$. In the next simplification step, there are two cases.

Case 1: If all of $b_1, ..., b_5$ are in A, the inconsistent set $Y_3 := \{b_1, ..., b_6, (b_1 \wedge .... \wedge b_6) \leftrightarrow c, \neg c\}$ is a subset of A.

Case 2: If not all of $b_1, ..., b_5$ are in A, say $b_j \notin A$, then $\neg b_j \in A$ (as A is complete), and so the inconsistent set $Y_3' := \{a, a \to (b_1 \wedge ... \wedge b_5), \neg b_j\}$ is a subset of A.

Of course, the new subset of A (Y_3 or Y_3') is not under all simplicity notions $<$ simpler than Y_2: for instance, we have $Y_3 \not< Y_2$ if $<$ is the "smaller than" relation (i.e. $Z < Y \Leftrightarrow |Z| < |Y|$). There is however an obvious simplicity notion for which both Y_3 and Y_3' *are* simpler than Y_2: they contain fewer connection rules than Y_2 (namely one instead of two). Indeed the application to implication agendas (in the appendix) will use a simplicity relation $<$ that lexicographically prioritises minimising the number of (possibly negated) connection rules over minimising the number of (possibly negated) atomic propositions, thereby ensuring that $Y_3 < Y_2$ and $Y_3' < Y_2$.

The set Y_3 is obtained from Y_2 in a particular manner: I have taken in "new" propositions (namely $b_1, ..., b_5$) each of which is logically entailed by some set of "old" propositions, namely by $V = \{a, a \to (b_1 \wedge ... \wedge b_5)\} \subseteq Y_2$. The fact that each "new" proposition b_j is entailed by a $V \subseteq Y_2$ has the important consequence that a simplification of Y_2 into a subset of A is possible whether or not A contains all "new" propositions: if A does then $Y_3 \subseteq A$, and if A does not contain the "new" proposition

[23] $Z \subsetneq_{\text{finite}} Y$ stands for $Z \subsetneq Y \& |Z| < \infty$. "$\subsetneq_{\text{finite}}$" can be replaced throughout by "$\subsetneq$" if one assumes a finite agenda.

[24] $<$ need not be connected, nor even transitive (if $<$ also satisfies these conditions, $<$ is a well-order). Note that well-foundedness implies asymmetry (i.e. if $Z < Y$ then $Y \not< Z$), hence irreflexivity. Further, given asymmetry and transitivity, $<$ is well-founded if and only if every set $\emptyset \neq \mathcal{J} \subseteq \mathcal{I}$ on which $<$ is connected has a least element (i.e. a $Z \in \mathcal{J}$ with $Z < Y$ for all $Y \in \mathcal{J} \backslash \{Z\}$).

b_j then $V \cup \{\neg b_j\} = Y_3' \subseteq A$. In the latter case, what is it that allows us to simplify Y_2 into $V \cup \{\neg b_j\}$? First, the entailment $V \models b_j$ guarantees us that $V \cup \{\neg b_j\}$ is indeed an inconsistent set, i.e. is in the range $\mathcal{I}$ of the simplicity relation. But this alone does not suffice: $V \cup \{\neg b_j\}$ must actually be simpler than Y_2. In summary, the following properties of the set Y_3 ensure that Y_2 can be simplified (into Y_3 or another set): Y_3 is simpler than Y_2, and moreover each "new" proposition $p \in Y_3 \backslash Y_2$ is entailed by a set of "old" propositions $V \subseteq Y_2$ such that $V \cup \{\neg p\}$ is simpler than Y_2. In this case I call Y_3 a *reduction* of Y_2, as formally defined now.[25]

Definition 1 *Given a simplicity relation* $<$*,*

(i) $Z \in \mathcal{I}$ *is a* ($<$-)reduction of $Y \in \mathcal{I}$ *(and* Y *is* ($<$-)reducible *to* Z*) if* $Z < Y$ *and moreover each* $p \in Z \backslash Y$ *is entailed by some* $V \subseteq Y$ *satisfying* $V \cup \{\neg p\} < Y$*;*

(ii) $Y \in \mathcal{I}$ *is* ($<$-)irreducible *if it has no reduction; let* $\mathcal{IR}_< := \{Y \in \mathcal{I} : Y$ *is* $<$*-irreducible*$\}$ *(the set of of minimal elements of the reduction relation).*

The art is to use a simplicity relation $<$ that allows sufficiently many (and the "right") simplifications so as to give few and elegant irreducible sets (hence a simple characterisation of collective consistency). Let me take up the two example above.

Example 1 (being simpler as being a subset). Let $<$ be $\subsetneq_{\text{finite}}$. Then reduction coincides with simplification: $Z \in \mathcal{I}$ is a reduction of $Y \in \mathcal{I}$ if and only if $Z \subsetneq_{\text{finite}} Y$. So Y is irreducible if and only if Y is a *minimal inconsistent* set, i.e. $\mathcal{IR}_< = \mathcal{MI}$ where

$$\mathcal{MI} := \{Y \in \mathcal{I} : \text{no proper subset of } Y \text{ is in } \mathcal{I}\}.$$

It can be shown that if X is a *simple* implication agenda then the set $\mathcal{IR}_< = \mathcal{MI}$ consists of all sets $Y \subseteq X$ of type $Y = \{p, \neg p\}$ or type

$$Y = \{a_1, a_1 \to a_2, ..., a_{k-1} \to a_k, \neg a_k\} \ (a_1, ..., a_k \text{ pairwise distinct}, k \geq 2). \quad (17)$$

Example 2 (being simpler as being smaller). Let $Z < Y :\Leftrightarrow |Z| < |Y|$. Then reduction is equivalent to Dietrich and List's [8] special reduction notion (see footnote 25). If X is again a simple implication agenda, $\mathcal{IR}_<$ is now much smaller than in Example 1: $\mathcal{IR}_<$ can be shown to consist of all sets $Y \subseteq X$ of type $\{p, \neg p\}$, or of type (17) with $k = 2$ (i.e. of type $\{a, a \to b, \neg b\}$, like in Lemma 1). To see why sets of type (17) are *not* irreducible if $k > 2$, note that such a set Y is reducible for instance to $Z := \{a_{k-1}, a_{k-1} \to a_k, \neg a_k\}$, because $|Z| < |Y|$ and a_{k-1} is entailed by $V := \{a_1, a_1 \to a_2, ..., a_{k-2} \to a_{k-1}\}$ where $|V \cup \{\neg a_{k-1}\}| < |Y|$. As a different agenda, consider a standard strict preference aggregation problem with a set of options $K \neq \emptyset$. This can be represented by the agenda $X_K := \{xPy, \neg xPy : x, y \in K\}$ in a suitable predicate logic with a binary *predicate* P for strict preference, a set of *constants* K for options, and a set of *axioms* containing the rationality conditions on strict linear orders, including for instance the transitivity axiom $(\forall v_1)(\forall v_2)(\forall v_3)((v_1 P v_2 \wedge v_2 P v_3) \to v_1 P v_3)$ (see Dietrich and List [7]; also List and Pettit [16]). Dietrich and List [8] call a set $Y \subseteq X_K$ a "k-cycle" ($k \geq 1$) if it has the form

$$Y = \{x_1 P x_2, x_2 P x_3, ..., x_{k-1} P x_k, x_k P x_1\} \ (x_1, ..., x_k \in K \text{ pairwise distinct}), \quad (18)$$

[25] In the special case that $<$ is defined by $Y < Z :\Leftrightarrow |Y| < |Z|$, this definition of reduction (and irreducibility) becomes equivalent to that introduced for different purposes by Dietrich and List [8]. Proposition 1 generalises one of their results. The present notion of reduction is more flexible and general, as sets may be simplified in other ways than through decreasing their size.

or arises from such a set by replacing one or more of the members xPy by the logically equivalent proposition $\neg yPx$. They show that the irreducible sets are the k-cycles with $k \leq 3$. To see why for $k > 3$ a k-cycle is not irreducible (though minimal inconsistent), note that a set Y of type (18) with $k \geq 4$ is reducible to $Z := \{x_1Px_2, x_2Px_3, x_3Px_1\}$ (a 3-cycle), because $|Z| < |Y|$ and x_3Px_1 is entailed by $V := \{x_3Px_4, x_4Px_5, ..., x_kPx_1\}$ where $|V \cup \{\neg x_3Px_1\}| < |Y|$.

To understand the properties of reduction better, let me record two lemmas.

Lemma 3 *Given any simplicity relation $<$, the reduction relation is itself a simplicity relation, that is:*

(i) *for all $Y, Z \in \mathcal{I}$, if $Z \subsetneq_{\text{finite}} Y$ then Z is a reduction of Y;*

(ii) *there is no infinite sequence $(Y_k)_{k=1,2,...}$ in $\mathcal{I}$ such that Y_{k+1} is a reduction of Y_k for all $k = 1, 2, ...$*

Proof. Both parts follow immediately from the analogous properties of $<$. ■

Lemma 4 (i) *For any simplicity relation $<$, $\mathcal{IR}_< \subseteq \mathcal{MI}$, and if $< = \subsetneq_{\text{finite}}$ then $\mathcal{IR}_< = \mathcal{MI}$.*

(ii) *For any simplicity relations $<$ and $<'$, if $<$ is a subrelation of $<'$ then $\mathcal{IR}_{<'} \subseteq \mathcal{IR}_<$.*

Proof. (i) Let $<$ be a simplicity relation. For all $Y \in \mathcal{I}$, if $Y \notin \mathcal{MI}_<$ then Y has an inconsistent proper subset Z, which we can choose finite by compactness of the logic. By Lemma 3 Y is reducible to Z, whence $Y \notin \mathcal{IR}_<$.

(ii) If $<$ and $<'$ are simplicity relations and $<$ is a subrelation of $<'$, then $<$-reduction is a subrelation of $<'$-reduction, and so $\mathcal{IR}_{<'} \subseteq \mathcal{IR}_<$. ■

Lemma 4 gives a general idea on how the set of irreducible sets $\mathcal{IR}_<$ depends on the simplicity notion $<$ used. The finer $<$ is, i.e. the more simplifications are allowed, the more reductions are allowed, and so the smaller $\mathcal{IR}_<$ is (see part (ii)). The coarsest choice of $<$ is $\subsetneq_{\text{finite}}$; then the only reductions are those to finite proper subsets, and $\mathcal{IR}_<$ is maximal: $\mathcal{IR}_< = \mathcal{MI}$, whereas in general $\mathcal{IR}_< \subseteq \mathcal{MI}$ (see part (i)).

I now prove the central property (16) announced earlier: every inconsistent and complete judgment set $A \subseteq X$ has an irreducible subset (reachable from A via finitely many simplifications).[26]

Proposition 1 *Given any simplicity relation $<$, every inconsistent and complete set $A \subseteq X$ has a subset in $\mathcal{IR}_<$.*

So, by Example 2 above, if X is a simple implication agenda then any inconsistent and complete set $A \subseteq X$ has a subset of type $\{p, \neg p\}$ or $\{a, a \to b, \neg b\}$ (as also shown in Lemma 1); and if X is instead the preference agenda X_K, A has a subset that is a k-cycle with $k \leq 3$ – a well-known result of social choice theory since A corresponds to a connected strict preference relation $\succ$ on K with rationality violation.

[26] The condition "proper subsets are simpler" on the simplicity relation $<$ may be dropped in Proposition 1 but not in Theorem 3.

Proof. Let $<$ and A be as specified. Assume for a contradiction that A has no subset in $\mathcal{IR}_<$. I recursively define a sequence $(Y_k)_{k=1,2,...}$ of inconsistent subsets of A such that $Y_{k+1} < Y_k$ for all k. This contradicts the well-foundedness of $<$.

First, put $Y_1 := A$, which is indeed an inconsistent subset of A.

Second, suppose Y_k is already defined. By assumption, Y_k is reducible, say to $Z \in \mathcal{I}$. First assume $Z \subseteq Y_k$. Letting $Y_{k+1} := Z$, it is true that Y_{k+1} is a subset of A (as $Y_{k+1} \subseteq Y_k \subseteq A$) and that $Y_{k+1} < Y_k$ (as Y_{k+1} is a reduction of Y_k). Now suppose $Z \nsubseteq A$. Then there is a $p \in Z \backslash A$. As Z is a reduction of Y_k, there is a $V \subseteq Y_k$ that entails p with $V \cup \{\neg p\} < Y_k$. Letting $Y_{k+1} := V \cup \{\neg p\}$, we have $Y_{k+1} < Y_k$, and $Y_{k+1} \subseteq A$ because $V \subseteq Y_k \subseteq A$ and because $\neg p \in A$ since $p \notin A$ and A is complete. ■

By Proposition 1, a quota rule $F_{(m_p)_{p \in X^+}}$ is consistent if and only if it never accepts any $Y \in \mathcal{IR}_<$. The following lemma tells which inequality we must impose to achieve this.

Lemma 5 *For every minimal inconsistent (hence every irreducible) set $Y \subseteq X$, a quota rule $F_{(m_p)_{p \in X^+}}$ never accepts all $p \in Y$ if and only if $\sum_{p \in Y}(n - m_p) < n$ (where $m_{\neg p} := n - m_p + 1$ for all $p \in X^+$).*

Proof. Consider a minimal inconsistent $Y \subseteq X$ and a quota rule $F := F_{(m_p)_{p \in X^+}}$.

First assume $\sum_{p \in Y}(n - m_p) \geq n$. Then N can be partitioned into (possibly empty) subgroups $N^p, p \in Y$, of size $|N^p| \leq n - m_p$. Construct a profile $(A_1, ..., A_n)$ of complete and consistent judgment sets such that, for all $p \in Y$, the people in N^p reject just p out of Y, i.e. $A_i \supseteq Y \backslash \{p\}$ for all $i \in N^p$; such A_i's exist as $Y \backslash \{p\}$ is consistent. Then $Y \subseteq F(A_1, ..., A_n)$ (as desired) since the number of people accepting a $p \in Y$ is $n - |N^p| \geq n - (n - m_p) = m_p$.

Conversely, suppose that F has an outcome $F(A_1, ..., A_n) \supseteq Y$. I show that $\sum_{p \in Y}(n - m_p) \geq n$. For all $p \in Y$, put $n_p := |\{i : p \in A_i\}|$; hence $|\{i : p \notin A_i\}| = n - n_p$. So

$$|\{(p, i) \in Y \times N : p \notin A_i\}| = \sum_{p \in Y}(n - n_p).$$

As no A_i contains all $p \in Y$, $|\{(p, i) \in Y \times N : p \notin A_i\}| \geq n$, i.e. $\sum_{p \in Y}(n - n_p) \geq n$. So, as for all $p \in Y$ we have $n_p \geq m_p$ (by $p \in F(A_1, ..., A_n)$), $\sum_{p \in Y}(n - m_p) \geq n$. ■

Proposition 1 and Lemma 5 imply the desired characterisation result.

Theorem 3 *For any simplicity relation $<$, a quota rule $F_{(m_p)_{p \in X^+}}$ is consistent if and only if*

$$\sum_{p \in Y}(n - m_p) < n \text{ for all } Y \in \mathcal{IR}_< \text{ (where } m_{\neg p} := n - m_p + 1 \ \forall p \in X^+).$$

Theorem 3 generalises the anonymous case of the "intersection property" result in Nehring and Puppe [18, 19]. This result makes no reference to a simplicity relation and uses $\mathcal{MI}$ instead of $\mathcal{IR}_<$. Hence it follows from Theorem 3 by choosing $<$ such that $\mathcal{IR}_< = \mathcal{MI}$, i.e. by choosing $<$ as the coarsest simplicity relation $\subsetneq_{\text{finite}}$. A

non-anonymous variant of Theorem 3 can be derived similarly, generalising the non-anonymous intersection property result.[27] One might wonder whether one could also generalise the (anonymous or non-anonymous) intersection property result in Dietrich and List [6] which requires no collective completeness[28], again by using irreducible sets instead of minimal inconsistent sets. No straightforward generalisation works, since the completeness assumption is essential in Proposition 1.

In general, the finer the simplicity relation $<$ is chosen, the smaller $\mathcal{IR}_<$ becomes, and hence the "slimmer" Theorem 3's characterisation becomes since redundant inequalities are avoided. The question of how much smaller than $\mathcal{MI}$ the set $\mathcal{IR}_<$ can get (and hence how much "slimmer" than the intersection property result Theorem 3's characterisation can get) depends on the concrete agenda X. In Example 2 above, $\mathcal{IR}_<$ gets significantly smaller than $\mathcal{MI}$. Note finally that if the inequalities have *no* solution, the agenda has *no* consistent quota rule. This is often so for agendas in classical logic, since here the judgments on atomic propositions fully settle the judgments on compound propositions.

While theoretically elegant, Theorem 3's system of inequalities is abstract. Checking whether it holds requires to know which sets are irreducible. The latter question can even be *non-decidable* in the technical sense: in some logics (such as standard predicate logic), it is non-decidable whether a set of propositions is inconsistent; so derived notions like irreducibility or minimal inconsistency may also be non-decidable.

In view of applications, two corollaries are useful. Call an inconsistent set *trivial* if it contains a pair $p, \neg p$ or contains a contradiction p (like $a \wedge \neg a$). Any trivial $Y \in \mathcal{IR}_<$ has by minimal inconsistency the form $Y = \{p, \neg p\}$ or $Y = \{p\}$. So for trivial $Y \in \mathcal{IR}_<$ the inequality $\sum_{p \in Y}(n - m_p) < n$ holds automatically, whatever the thresholds $(m_p)_{p \in X^+} \in \{1, ..., n\}^{X^+}$. Removing these redundant inequalities, we obtain a slightly slimmer characterisation:

Corollary 4 *Theorem* 3 *still holds if* $\mathcal{IR}_<$ *is replaced by* $\mathcal{IR}^*_< := \{Y \in \mathcal{IR}_< : Y$ *is non-trivial*$\}$.

As an illustration, consider a simple implication agenda X. By Theorem 1, $F_{(m_p)_{p \in X^+}}$ is consistent if and only if

$$m_b \leq m_a + m_{a \to b} - n \text{ for all } a \to b \in X. \tag{19}$$

This characterisation is equivalent to that of Corollary 4 if $<$ is defined by $Z < Y :\Leftrightarrow |Z| < |Y|$: indeed, $\mathcal{IR}^*_< = \{\{a, a \to b, \neg b\} : a \to b \in X\}$ by Example 2 above, so that

[27] If we endow each $p \in X^+$ not with a threshold $m_p \in \{1, ..., n\}$ but, more generally, with a set $\mathcal{C}_p$ of ("winning") coalitions $C \subseteq N$ such that $\emptyset \notin \mathcal{C}_p$, $N \in \mathcal{C}_p$, and $[C \in \mathcal{C}_p \& C \subseteq C^* \subseteq N] \Rightarrow C^* \in \mathcal{C}_p$, we can define an aggregation rule $F_{(\mathcal{C}_p)_{p \in X^+}}$ with universal domain by $F_{(\mathcal{C}_p)_{p \in X^+}}(A_1, ..., A_n) = \{p \in X : \{i \in N : p \in A_i\} \in \mathcal{C}_p\}$ (where $\mathcal{C}_{\neg p} := \{C \subseteq N : N \backslash C \notin \mathcal{C}_p\}$ for all $p \in X^+$). Such a rule $F_{(\mathcal{C}_p)_{p \in X^+}}$ is called a *committee rule*. The quota rules $F_{(m_p)_{p \in X^+}}$ are precisely the anonymous committee rules (where each $p \in X^+$ has set of winning coalitions $\mathcal{C}_p = \{C \subseteq N : |C| \geq m_p\}$). The analogue of Theorem 3 is: for any simplicty relation $<$, a committee rule $F_{(\mathcal{C}_p)_{p \in X^+}}$ is consistent if and only if $\cap_{p \in Y} C_p \neq \emptyset$ for all $Y \in \mathcal{IR}_<$ and all winning coalitions $C_p \in \mathcal{C}_p$, $p \in Y$. This becomes the non-anonymous intersection property result if $\mathcal{IR}_< = \mathcal{MI}$, i.e. if we choose $< := \subseteq_{\text{finite}}$.

[28] More precisely, it does not require for propositions $p \in X$ that $m_{\neg p} = n - m_p + 1$ (or, in the non-anonymous case discussed in footnote 27, that the coalitions winning for $\neg p$ be the coalitions whose complements are not winning for p).

the inequalities $\sum_{p \in Y}(n - m_p) < n$, $Y \in \mathcal{IR}^*_<$, are equivalent to the inequalities (19). If $<$ is alternatively defined as $\subsetneq_{\text{finite}}$, then by Example 1 above $\mathcal{IR}^*_<$ consists of all sets of type (17); thus $\mathcal{IR}^*_<$ is now much larger, and the resulting characterisation of Corollary 4 contains redundant inequalities.

Determining the set $\mathcal{IR}_<$ (or $\mathcal{IR}^*_<$) is often hard, e.g. for general implication agendas. Determining a *superset* of it can be simpler – and it suffices by the next corollary, obtained by combining Corollary 4 with Theorem 3, the latter applied with $< = \subsetneq_{\text{finite}}$, i.e. with $\mathcal{IR}_< = \mathcal{MI}$.

Corollary 5 *Theorem* 3 *still holds if* $\mathcal{IR}_<$ *is replaced by any* $\mathcal{Y}$ *with* $\mathcal{IR}^*_< \subseteq \mathcal{Y} \subseteq \mathcal{MI}$.

So, to find out for a concrete agenda which quota rules are consistent, it suffices to define a suitable simplicity relation $<$ and determine *some* set $\mathcal{Y}$ with $\mathcal{IR}^*_< \subseteq \mathcal{Y} \subseteq \mathcal{MI}$. Precisely this is done for implication agendas in the appendix.

8 Conclusion

Connection rules, of the uni-directional kind $p \rightarrow q$ or bi-directional kind $p \leftrightarrow q$, are at the heart of judgment aggregation. They express links that may be accepted or rejected, for instance causal links between facts or justificational links between claims. Once we interpret these (bi-)implications subjunctively, we can generate consistent and complete collective judgment sets by taking independent and anonymous votes on the propositions, provided that we use appropriate acceptance thresholds (see Theorems 1 and 2 and Table 2). This possibility result holds for judgment aggregation problems on so-called *implication* agendas.

The results on implication agendas are applications of an abstract result, Theorem 3, which applies to arbitrary agendas in a general logic: it characterises consistent aggregation in terms of so-called *irreducible* sets (which generalise minimal inconsistent sets[29]). It would be interesting to apply this result to classes of agendas other than implication agendas, in order to gain new insights on (im)possibilities of propositionwise voting. However, at least as important as this would be to develop a systematic understanding of *non-propositionwise* judgment aggregation rules. Though often mentioned, this route is largely unexplored.

9 References

[1] B. Chapman, Rational aggregation, Polit. Philos. Econ. 1(3) (2002), 337-354.

[2] F. Dietrich, Judgment aggregation: (im)possibility theorems, J. Econ. Theory 126(1) (2006), 286-298.

[3] F. Dietrich, A generalised model of judgment aggregation, Soc. Choice Welfare 28 (2007), 529-65.

[4] F. Dietrich, Aggregation theory and the relevance of some issues to others, Meteor Research Memorandum 07/002 (2006).

[29] Because minimal inconsistent sets are defined relative to set-inclusion whereas irreducible sets are defined relative to a general *simplicity* relation that need not be set-inclusion

[5] F. Dietrich, C. List, Strategy-proof judgment aggregation, Econ. Philos. 23 (2007), 269-300.
[6] F. Dietrich, C. List, Judgment aggregation by quota rules, J. Theoretical Politics 19(4) (2007), 391-424.
[7] F. Dietrich, C. List, Arrow's theorem in judgment aggregation, Soc. Choice Welfare 29 (2007), 19-33.
[8] F. Dietrich, C. List, Judgment aggregation on restricted domains, Meteor Research Memorandum 06/033 (2006).
[9] E. Dokow, R. Holzman, Aggregation of binary evaluations, working paper (2005), Technion Israel Institute of Technology.
[10] P. Gärdenfors, An Arrow-like theorem for voting with logical consequences, Econ. Philos. 22(2) (2006), 181-190.
[11] S. Konieczny, R. Pino-Perez, Merging information under constraints: a logical framework, Journal of Logic and Computation 12(5) (2002), 773-808.
[12] L.A. Kornhauser, L.G. Sager, Unpacking the Court. Yale Law Journal 96(1) (1986), 82-117.
[13] D. Lewis, Counterfactuals, Oxford: Basil Blackwell, 1973.
[14] C. List, A model of path dependence in decisions over multiple propositions, Amer. Polit. Sci. Rev. 98(3) (2004), 495-513.
[15] C. List, P. Pettit, Aggregating sets of judgments: an impossibility result, Econ. Philos. 18 (2002), 89-110.
[16] C. List, P. Pettit, Aggregating sets of judgments: two impossibility results compared, Synthese 140(1-2) (2004), 207-235.
[17] P. Mongin, Factoring out the impossibility of logical aggregation, J. Econ. Theory, forthcoming.
[18] K. Nehring, C. Puppe, Strategyproof social choice on single-peaked domains: possibility, impossibility and the space between, working paper (2002), Karlsruhe University.
[19] K. Nehring, C. Puppe, Consistent judgment aggregation: the truth-functional case, Social Choice and Welfare, forthcoming.
[20] K. Nehring, C. Puppe, The structure of strategy-proof social choice, part II: non-dictatorship, anonymity and neutrality, working paper (2005), Karlsruhe University.
[21] M. Pauly, M. van Hees, Logical constraints on judgment aggregation, J Philos Logic 35 (2006), 569-585.
[22] P. Pettit, Deliberative democracy and the discursive dilemma, Philosophical Issues 11 (2001), 268-299.
[23] G. Pigozzi, Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation, Synthese 152(2) (2006), 285-298.
[24] G. Priest, An introduction to non-classical logic, Cambridge University Press, 2001
[25] R. Stalnaker, A theory of conditionals, in N. Rescher (Ed.) Studies in Logical Theory, Blackwell: Oxford, 1986.
[26] M. van Hees, The limits of epistemic democracy, Soc. Choice Welfare 28 (2007), 649-66.

A Proof of Theorem 2 from Theorem 3

We consider an arbitrary implication agenda X ($\subseteq \mathbf{L}$). The language $\mathbf{L}$ (defined in Section 2) is endowed with the non-classical notions of entailment and (in)consistency defined in Section 3 (using C^+-interpretations). Recall that $\mathcal{A}$ ($\subseteq \mathbf{L}$) is the set of atomic propositions. Denote the set of all connection rules by $\mathcal{R}$ ($= \{a \to b, (a \wedge b) \leftrightarrow c, ...\}$). For all $S \subseteq \mathbf{L}$ let $S^\neg := \{\neg p : p \in S\}$ and $\bar{S} := S \cup S^\neg$. We wish to apply Theorem 3 to X – but with which simplicity relation $<$? Defining $<$ as $\subsetneq_{\text{finite}}$ gives a very complicated set $\mathcal{IR}_< = \mathcal{MI}$ (containing for instance sets like $Y = \{a, a \to b, a', a' \to b', (b \wedge b') \to (a \wedge c), \neg c\}$). Even the finer simplicity relation given by $Z < Y :\Leftrightarrow |Z| < |Y|$, while suitable for *simple* implication agendas (see the end of Section 7), is inappropriate in general since, as indicated in Section 7, we would like to simplify sets like $\{a, a \to (b_1 \wedge b_2 \wedge b_3), (b_1 \wedge b_2 \wedge b_3) \to c, \neg c\}$ into $\{b_1, b_2, b_3, (b_1 \wedge b_2 \wedge b_3) \to c, \neg c\}$ on the grounds that the latter set contains fewer connection rules despite of having more elements overall. The following lexicographic simplicity notion allows us to perform such simplifications and to get a grip on $\mathcal{IR}_<$. For all inconsistent sets $Z, Y \subseteq X$,

$$Z < Y :\Leftrightarrow (|Z \cap \bar{\mathcal{R}}|, |Z \cap \bar{\mathcal{A}}|) \text{ is lexicographically smaller than } (|Y \cap \bar{\mathcal{R}}|, |Y \cap \bar{\mathcal{A}}|),$$

i.e. $|Z \cap \bar{\mathcal{R}}| < |Y \cap \bar{\mathcal{R}}|$ or $|Z \cap \bar{\mathcal{R}}| = |Y \cap \bar{\mathcal{R}}| \& |Z \cap \bar{\mathcal{A}}| < |Z \cap \bar{\mathcal{A}}|$. For instance, $\{a, \neg b\} < \{a \to b\}$ as $(0, 2)$ is lexicographically smaller than $(1, 0)$.

The following is easily shown (using that the "lexicographically smaller than" relation is well-founded).

Lemma 6 *The above relation $<$ is a simplicity relation.*[30]

To identify the $<$-irreducible sets, we first need to understand better which entailments and inconsistencies hold within implication agendas; hence the next two technical lemmas. Generalising Section 6's notation "X_p", I put, for all $p \in \mathbf{L}$ and all $R \subseteq \mathbf{L}$,

$$R_p := \{s \in \mathbf{L} : p \to s \in R \text{ or } p \leftrightarrow s \in R \text{ or } s \leftrightarrow p \in R\},$$

the set of propositions "reachable" from p via (bi-)implications in R. I first establish a plausible fact about entailments between connection rules: namely, for instance, that $R = \{p \to b, p \to (c \wedge d)\} \vDash p \to (b \wedge c)$ because each conjunct of $b \wedge c$ (i.e. b and c) is a conjunct of some $s \in R_p = \{b, c \wedge d\}$.

Lemma 7 *For all $R \subseteq \mathcal{R}$ and $p \to q \in \mathcal{R}$,*

$$R \vDash p \to q \Leftrightarrow \mathrm{C}(q) \backslash \mathrm{C}(p) \subseteq \bigcup_{s \in R_p} \mathrm{C}(s).$$

Note that this characterisation of $R \vDash p \to q$ implies one of $R \vDash p \leftrightarrow q$ (for $R \subseteq \mathcal{R}$ and $p \leftrightarrow q \in \mathcal{R}$), since $R \vDash p \leftrightarrow q$ if and only if $R \vDash p \to q$ and $R \vDash q \to p$.

Proof. Let $R \subseteq \mathcal{R}$ and $p \to q \in \mathcal{R}$.

[30] More generally, for any partition of X into sets $X_1, ..., X_k$, a simplicity relation $<$ is defined by $Z < Y :\Leftrightarrow [(|Z \cap X_1|, ..., |Z \cap X_n|) \text{ is lexicographically smaller than } (|Y \cap X_1|, ..., |Y \cap X_n|)]$.

1. First let $\mathrm{C}(q)\backslash\mathrm{C}(p) \subseteq \bigcup_{s\in R_p} \mathrm{C}(s)$. Suppose all $r \in R$ hold in world w of interpretation $(W, (f_r), (v_w))$. We have to show that $p \to q$ holds in w, i.e. that all $a \in \mathrm{C}(q)$ hold in all $w^* \in f_p(w)$. Let $a \in \mathrm{C}(q)$ and $w^* \in f_p(w)$. By assumption, $a \in \mathrm{C}(p)$ or $a \in \mathrm{C}(s)$ for some $s \in R_p$. In the first case, a holds in w^* as p does (by $w^* \in f_p(w)$). In the second case, a holds in w^* as s does (by $v_w(p \to s) = T$ and $w^* \in f_p(w)$).

2. Conversely, suppose that $a \in \mathrm{C}(q)\backslash\mathrm{C}(p)$ but $a \notin \bigcup_{s\in R_p} \mathrm{C}(s)$. To show $R \not\models p \to q$, consider an interpretation $(W, (f_p), (v_w))$ such that: (i) W contains at least two distinct worlds w, w^*, (ii) all atomic propositions hold in w, (iii) all atomic propositions except a hold in w^*, (iv) $f_p(w) = \{w, w^*\}$ (which is allowed as p holds in w and w^*), and (v) for all $t \in \mathbf{L}\backslash\{p\}$ $f_t(w) \subseteq \{w\}$. To complete the proof, I show that all $r \in R$ hold in w but $p \to q$ doesn't. First, $v_w(p \to q) = F$ by (iv) and as $v_{w^*}(q) = F$ by (iii). To show the truth in w of all $r \in R$, I show that of every implication $t \to s$ with $t \to s \in R$ or $t \leftrightarrow s \in R$ or $s \leftrightarrow t \in R$. For such $t \to s$, if $t \neq p$ then $v_w(t \to s) = T$ by (v) and (ii); and if $t = p$ then $v_w(t \to s) = T$ by (iv) and (ii)-(iii) and using that $a \notin \mathrm{C}(s)$. ■

The next technical lemma shows that there are broadly two ways in which a subset A of the implication agenda X can be inconsistent (the second way, (20), holds for instance if $\neg(a \to (b \wedge c)), a \to b, a \to c \in A$.).

Lemma 8 *If $A \subseteq \bar{\mathcal{A}} \cup \bar{\mathcal{R}}$ is inconsistent, then either already $A\backslash\mathcal{R}^{\neg}$ is inconsistent or*

$$A \text{ contains some } \neg r \in \mathcal{R}^{\neg} \text{ such that } A \cap \mathcal{R} \vDash r. \tag{20}$$

Proof. Suppose $A \subseteq \bar{\mathcal{A}} \cup \bar{\mathcal{R}}$. Assume $A_* := A\backslash\mathcal{R}^{\neg}$ is consistent and (20) does not hold. I show that A is consistent. For all $\neg(p \to q) \in A$,

(α) there is $a_{p\to q} \in \mathrm{C}(q)\backslash\mathrm{C}(p)$ with $a_{p\to q} \notin \mathrm{C}(q')$ for all $q' \in A_p$,

as otherwise $\mathrm{C}(q)\backslash\mathrm{C}(p) \subseteq \cup_{q'\in A_p}\mathrm{C}(q')$, whence by Lemma 7 $A \cap \mathcal{R} \vDash p \to q$ (take $R := A \cap \mathcal{R}$ and note that $R_p = A_p$), implying (20). Further, for all $\neg(p \leftrightarrow q) \in A$, *either*

(β1) there is $a^1_{p\leftrightarrow q} \in \mathrm{C}(q)\backslash\mathrm{C}(p)$ with $a^1_{p\leftrightarrow q} \notin \mathrm{C}(q')$ for all $q' \in A_p$

or

(β2) there is $a^2_{p\leftrightarrow q} \in \mathrm{C}(p)\backslash\mathrm{C}(q)$ with $a^2_{p\leftrightarrow q} \notin \mathrm{C}(p')$ for all $p' \in A_q$,

as otherwise $\mathrm{C}(q)\backslash\mathrm{C}(p) \subseteq \cup_{q'\in A_p}\mathrm{C}(q')$ and $\mathrm{C}(p)|\mathrm{C}(q) \subseteq \cup_{p'\in A_q}$, whence again by Lemma 7 $A \cap \mathcal{R} \vDash p \to q$ and $A \cap \mathcal{R} \vDash p \to r$, i.e. $A \cap \mathcal{R} \vDash p \leftrightarrow q$, implying (20).

To prove A's consistency, I construct an interpretation and show that in a world all $r \in A$ hold. Notationally, for any $r \in \mathcal{R}$ let r^{mat} be r's material counterpart: $(p \to q)^{\mathrm{mat}}$ is $\neg p \vee q$, and $(p \leftrightarrow q)^{\mathrm{mat}}$ is $(p \to q)^{\mathrm{mat}} \wedge (q \to p)^{\mathrm{mat}}$. Let A^{mat}_* be the set arising from A_* by replacing all $r \in A_* \cap \mathcal{R}$ by r^{mat}. Since A_* is consistent and $r \models r^{\mathrm{mat}}$ for all $r \in \mathcal{R}$, A^{mat}_* is also consistent. So there exists an interpretation $(W, (f_p), (v_w))$ and a world w such that

(w1) all members of A^{mat}_* are true in w.

As the propositions in A^{mat}_* contain no subjunctive (bi-)implications, their truth values in w depend neither on other worlds nor on the functions $f_p, p \in \mathbf{L}$. So we may assume the following w.l.o.g.

(w2) For all $\neg(p \to q) \in A$, there is a world $w_{p\to q} \in W\backslash\{w\}$ in which all atomic proposition except $a_{p\to q}$ hold; and $w_{p\to q} \in f_p(w)$ but $w_{p\leftrightarrow q} \notin f_s(w)$ $\forall s \in \mathbf{L}\backslash\{p\}$.

(w3) For all $\neg(p \leftrightarrow q) \in A$ with (β1), there is a world $w^1_{p\leftrightarrow q} \in W\backslash\{w\}$ in which all atomic propositions except $a^1_{p\leftrightarrow q}$ hold; and $w^1_{p\leftrightarrow q} \in f_p(w)$ but $w^1_{p\leftrightarrow q} \notin f_s(w)$ $\forall s \in \mathbf{L}\backslash\{p\}$.

(w4) For all $\neg(p \leftrightarrow q) \in A$ with (β2), there is a world $w^2_{p\leftrightarrow q} \in W\backslash\{w\}$ in which all atomic propositions except $a^2_{p\leftrightarrow q}$ hold; and $w^2_{p\leftrightarrow q} \in f_q(w)$ but $w^2_{p\leftrightarrow q} \notin f_s(w)$ $\forall s \in \mathbf{L}\backslash\{q\}$.

(w5) Worlds $w' \in W$ other than those defined in (w1)-(w4) are not reachable from w: $w' \notin f_r(w)$ $\forall r \in \mathbf{L}$.

To complete the proof, I consider any $r \in A$ and show that r holds in w.

Case 1: r is atomic or negated atomic. Then $r \in A^{\text{mat}}_*$. So r holds in w by (w1).

Case 2: r is an implication $s \rightarrow t$. Let $w' \in f_s(w)$. I have to show that t holds in w'. If $w' = w$, s holds in w by $w \in f_s(w)$; so, as $(s \rightarrow t)^{\text{mat}} = \neg s \vee t$ holds in w by (w1), t holds in w. Now let $w' \neq w$. Then by (w5), w' is one of the worlds defined in (w2)-(w4). Assume $w' = w_{p\rightarrow q}$, a world defined in (w2) (proofs for (w3) and (w4) are similar). By $w_{p\rightarrow q} \in f_s(w)$ and (w2), $p = s$. By (w2), all atomic propositions except $a_{p\rightarrow q}$ hold in $w_{p\rightarrow q}$, where $a_{p\rightarrow q}$ isn't a conjunct of t by (α). So t holds in $w_{p\rightarrow q} = w'$.

Case 3: r is a bi-implication $s \leftrightarrow t$. $s \leftrightarrow t$ holds in w if $s \rightarrow t$ and $t \rightarrow s$ are true in w. The latter can be shown by a procedure analogous to that in case 2.

Case 4: r is a negated implication $\neg(p \rightarrow q)$. To show that r holds in w, I show that $p \rightarrow q$ fails in w. This is so because, by (w2), $w_{p\rightarrow q} \in f_p(w)$ where q fails in $w_{p\rightarrow q}$ as its conjunct $a_{p\rightarrow q}$ fails.

Case 5: r is a negated bi-implication $\neg(p \leftrightarrow q)$. To show that r holds in w, I show that $p \leftrightarrow q$ is false in w, i.e. that $p \rightarrow q$ or $q \rightarrow p$ is false in w. Under (β1) $p \rightarrow q$ is false in w (consider the world $w^1_{p\leftrightarrow q}$ and use (w3)), and under (β2) $q \rightarrow p$ is false in w (consider the world $w^2_{p\leftrightarrow q}$ and use (w4)). ■

To allow us to apply Corollary 5, I now define a class $\mathcal{Y}$ of inconsistent sets $Y \subseteq X$, and I show that $\mathcal{IR}^*_< \subseteq \mathcal{Y} \subseteq \mathcal{MI}$. Let $\mathcal{Y} := \mathcal{Y}_\rightarrow \cup \mathcal{Y}_\leftrightarrow \cup \mathcal{Y}_{\neg\rightarrow} \cup \mathcal{Y}_{\neg\leftrightarrow}$, where $\mathcal{Y}_\rightarrow$, $\mathcal{Y}_\leftrightarrow$, $\mathcal{Y}_{\neg\rightarrow}$ and $\mathcal{Y}_{\neg\leftrightarrow}$ are the sets that consist, respectively, of

- all $Y \subseteq X$ of type $\{p \rightarrow q, \neg a\} \cup \text{C}(p)$ where $a \in \text{C}(q)\backslash\text{C}(p)$;
- all $Y \subseteq X$ of type $\{p \leftrightarrow q, \neg a\} \cup \text{C}(p)$ or $\{q \leftrightarrow p, \neg a\} \cup \text{C}(p)$ where $a \in \text{C}(q)\backslash\text{C}(p)$;
- all $Y \subseteq X$ of type $\{\neg(p \rightarrow q)\} \cup \{p_s : s \in S\}$ where $S \in X_{p\rightarrow q}$ and $\forall s \in S$ $p_s \in \{p \rightarrow s, p \leftrightarrow s, s \leftrightarrow p\}$;
- all $Y \subseteq X$ of type $\{\neg(p \leftrightarrow q)\} \cup \{p_s : s \in S\} \cup \{q_s : s \in S'\}$ where $S \in X_{p\rightarrow q}$, $\forall s \in S$ $p_s \in \{p \rightarrow s, p \leftrightarrow s, s \leftrightarrow p\}$, $S' \in X_{q\rightarrow p}$, $\forall s \in S'$ $q_s \in \{q \rightarrow s, q \leftrightarrow s, s \leftrightarrow q\}$, and the sets $\{p_s : s \in S\}$, $\{q_s : s \in S'\}$ are either each disjoint with $\{p \leftrightarrow q, q \leftrightarrow p\}$ or each equal to $\{q \leftrightarrow p\}$ (the latter is only possible if $S = \{q\} \& S' = \{p\}$; the former holds automatically if $S \neq \{q\} \& S' \neq \{p\}$ as then $q \notin S \& p \notin S'$).

(The set $X_{p\rightarrow q}$ in the last two bullet points was defined in Section 6.)

Lemma 9 *For $\mathcal{Y}$ as defined above, $\mathcal{Y} \subseteq \mathcal{MI}$.*

Proof. Let $\mathcal{Y}$ be as specified. Consider any $Y \in \mathcal{Y}$. I show that $Y \in \mathcal{MI}$ by going through the four possible cases.

1. Let $Y \in \mathcal{Y}_\rightarrow$, i.e. $Y = \{p \rightarrow q, \neg a\} \cup \text{C}(p)$ where $a \in \text{C}(q)\backslash\text{C}(p)$. Y is inconsistent because, by $\text{C}(p) \vDash p$ and $\{p \rightarrow q, p\} \vDash q$, we have $\{p \rightarrow q\} \cup \text{C}(p) \vDash q$.

Moreover, for any $y \in Y$, the consistency of $Y\backslash\{y\}$ can be checked by finding an interpretation with a world w in which all $z \in Y\backslash\{y\}$ hold. Specifically, $\{p \to q\} \cup \mathrm{C}(p)$ is consistent: let all atomic propositions hold in w and in all other worlds; $\{\neg a\} \cup \mathrm{C}(p)$ is consistent: let all atomic propositions except a hold in w; and, for any $y \in \mathrm{C}(p)$, $\{p \to q, \neg a\} \cup \mathrm{C}(p)\backslash\{y\}$ is consistent: let the only atomic propositions true in w be those in $\mathrm{C}(p)\backslash\{y\}$, and put $f_p(w) = \emptyset$ (which is allowed as p fails in w).

2. If $Y \in \mathcal{Y}_{\leftrightarrow}$ then $Y \in \mathcal{MI}$ by a proof similar to that under 1.

3. Now let $Y \in \mathcal{Y}_{\neg\to}$, say (in the earlier notation) $Y = \{\neg(p \to q)\} \cup \{p_s : s \in S\}$. We have $\{p_s : s \in S\}_p = S \in X_{p\to q}$, whence by Lemma 7 $\{p_s : s \in S\} \models p \to q$. So Y is inconsistent. To check *minimal* inconsistency, consider any $Z \subsetneq Y$. If $\neg(p \to q) \notin Z$, Z is consistent, as seen from an interpretation such that all atomic propositions hold in all worlds. If $\neg(p \to q) \in Z$, then $Z = \{\neg p \to q)\} \cup R^*$ with $R^* = \{p_s : s \in S^*\}$ and $S^* \subsetneq S$. Note that $R^*_p = S^*$. So $R^*_p \subsetneq S$. This and $S \in X_{p\to q}$ imply that $\mathrm{C}(q)\backslash\mathrm{C}(p) \nsubseteq \cup_{s\in R^*_p}\mathrm{C}(s)\}$, whence by Lemma 7 $R^* \not\models p \to q$. So Z $(= \{\neg p \to q)\} \cup R^*)$ is consistent.

4. Finally, let $Y \in \mathcal{Y}_{\neg\leftrightarrow}$, say (in the earlier notation) $Y = \{\neg(p \leftrightarrow q)\} \cup \{p_s : s \in S\} \cup \{q_s : s \in S'\}$. It can be shown like under 3 that $\{p_s : s \in S\} \models p \to q$ and $\{q_s : s \in S'\} \models q \to p$. So $\{p_s : s \in S\} \cup \{q_s : s \in S'\} \models p \leftrightarrow q$. Hence Y is inconsistent. Now consider any $Z \subsetneq Y$. If $\neg(p \leftrightarrow q) \notin Z$, Y is consistent by an argument like in case 3. If $\neg(p \leftrightarrow q) \in Z$, then $Z = \{\neg(p \leftrightarrow q)\} \cup R^*$ with $R^* = \{p_s : s \in S^*\} \cup \{q_s : s \in S'^*\}$ and $S^* \subseteq S$, $S'^* \subseteq S'$, where $S^* \subsetneq S$ or $S'^* \subsetneq S'$. Note that $R^*_p = S^*$ and $R^*_q = S'^*$. So $R^*_p \subsetneq S$ or $R^*_q \subsetneq S'$. Hence $R^* \not\models p \to q$ or $R^* \not\models q \to p$, by an argument like that under 3. So $R^* \not\models p \leftrightarrow q$. Hence Z $(= \{\neg(p \leftrightarrow q)\} \cup R^*)$ is consistent. ■

Lemma 10 *For $<$ and $\mathcal{Y}$ as defined above, $\mathcal{IR}^*_< \subseteq \mathcal{Y}$.*

Proof. Let $<$ and $\mathcal{Y}$ $(= \mathcal{Y}_{\to} \cup \mathcal{Y}_{\leftrightarrow} \cup \mathcal{Y}_{\neg\to} \cup \mathcal{Y}_{\neg\leftrightarrow})$ be as specified. Consider a $Y \in \mathcal{IR}^*_<$. I show that $Y \in \mathcal{Y}$. I will use that $Y \in \mathcal{MI}$ by Lemma 4, and that (*) Y contains no pair $t, \neg t$ by non-triviality.

Case 1: $Y \cap \mathcal{R}^{\neg} = \emptyset$. Then (i) Y has a subset of type $\{p \to q\} \cup \mathrm{C}(p)$, *or* (ii) Y has a subset of type $\{p \leftrightarrow q\} \cup \mathrm{C}(p)$ or $\{q \leftrightarrow p\} \cup \mathrm{C}(p)$. Otherwise Y would be consistent, as seen from an interpretation with a world w in which the only true atomic propositions are those in Y and such that $f_t(w) = \emptyset$ if $t \in \mathbf{L}$ is false in w: in w, all $y \in Y \cap \bar{\mathcal{A}}$ hold by construction (and by (*)), all $p \to q \in Y$ hold by $f_p(w) = \emptyset$ (as p is false by not-(i)), all $p \leftrightarrow q \in Y$ hold by $f_p(w) = f_q(w) = \emptyset$ (as p and q are false by not-(ii)), and there are no $y \in Y \cap \mathcal{R}^{\neg}$.

Subcase 1a: (i) holds, say $\{p \to q\} \cup \mathrm{C}(p) \subseteq Y$. I show that $Y \in \mathcal{Y}_{\to}$. If there is an $a \in \mathrm{C}(q)\backslash\mathrm{C}(p)$ with $\neg a \in Y$, then $\{p \to q, \neg a\} \cup \mathrm{C}(p) \subseteq Y$, hence $\{p \to q, \neg a\} \cup \mathrm{C}(p) = Y$ (as $Y \in \mathcal{MI}$), and so $Y \in \mathcal{Y}_{\to}$. Hence it suffices to prove that such an a exists. For a contradiction, suppose (**) $\neg a \notin Y$ for all $a \in \mathrm{C}(q)\backslash\mathrm{C}(p)$. I show that Y is reducible to $Z := Y \cup \mathrm{C}(q)\backslash\{p \to q\}$, a contradiction. First, Z is indeed inconsistent: otherwise there would exist an interpretation with a world w in which all $z \in Z$ hold, where by $Z \cap \mathcal{R}^{\neg} = \emptyset$ we may assume w.l.o.g. that $f_p(w)$ contains no world other than w; thus $p \to q$ also holds in w, so that $Z \cup \{p \to q\} = Y \cup \mathrm{C}(q)$ is consistent, a contradiction. Second, we have $Z < Y$ by $|Z \cap \bar{\mathcal{R}}| = |Y \cap \bar{\mathcal{R}}| - 1$ (and by our lexicographic definition of $<$). Finally, any $y \in Z\backslash Y$ belongs to $\mathrm{C}(q)$, hence is entailed by $Z := \mathrm{C}(p) \cup \{p \to q\}$ $(\subseteq Y)$; it remains to show $Z \cup \{\neg y\} < Y$, which I do by proving that $|(Z \cup \{\neg y\}| \cap \bar{\mathcal{R}}| < |Y \cap \bar{\mathcal{R}}|$, i.e. that $|Y \cap \bar{\mathcal{R}}| > 1$. Suppose the

contrary. Then $Y = \{p \to q\} \cup \mathrm{C}(p) \cup Y'$ for some $Y' \subseteq \bar{\mathcal{A}}$. By $Y \in \mathcal{MI}$, $\mathrm{C}(p) \cup Y'$ is consistent. So there is an interpretation with a world w in which all $a \in \mathrm{C}(p) \cup Y'$ hold, where by $\mathrm{C}(p) \cup Y' \subseteq \bar{\mathcal{A}}$ we may assume w.l.o.g. that $f_p(p)$ contains no world other than w, and that all $a \in \mathcal{A}$ with $\neg a \notin Y$ hold in w. All $a \in \mathrm{C}(q)$ satisfy $\neg a \notin Y$: if $a \in \mathrm{C}(q) \backslash \mathrm{C}(p)$ by (**), and if $a \in \mathrm{C}(q) \cap \mathrm{C}(p)$ by (*). So, in w, all $a \in \mathrm{C}(q)$ and hence q hold; so $p \to q$ holds. But then all $y \in Y$ hold in w, contradicting Y's inconsistency.

Subcase 1b: (ii) holds, say $\{p \leftrightarrow q\} \cup \mathrm{C}(p) \subseteq Y$ (the proof is analogous if $p \leftrightarrow q$ is replaced by $q \leftrightarrow p$). To show that $Y \in \mathcal{Y}_{\leftrightarrow}$, it suffices to slightly adapt the proof in Subcase 1a: replace "$\to$" by "$\leftrightarrow$", and in both interpretations assume w.l.o.g. that $f_q(w)$ (in addition to $f_p(w)$) contains no world other than w.

Case 2: $Y \cap \mathcal{R}^{\neg} \neq \emptyset$. Then $Y \backslash \mathcal{R}^{\neg} \subsetneq Y$, whence $Y \backslash \mathcal{R}^{\neg}$ is consistent by $Y \in \mathcal{MI}$. So by Lemma 8 Y contains a $\neg r \in \mathcal{R}^{\neg}$ such that $Y \cap \mathcal{R} \vDash r$. Let $R := Y \cap \mathcal{R}$. As Y is *minimal* inconsistent, $Y = \{\neg r\} \cup R$. I consider two subcases.

Subcase 2a: r is an implication $p \to q$. I show that $Y \in \mathcal{Y}_{\neg\to}$. As $Y = \{\neg(p \to q)\} \cup R \in \mathcal{MI}$, R is minimal subject to entailing $p \to q$. So, by Lemma 7, R is minimal subject to $\mathrm{C}(q) \backslash \mathrm{C}(p) \subseteq \bigcup_{s \in R_p} \mathrm{C}(s)$. This implies that $R_p \in X_{p \to q}$ and that $R = \{p_s : s \in R_p\}$ for some $p_s \in \{p \to s, p \leftrightarrow s, s \leftrightarrow p\}$, $s \in R_p$. So Y $(= \{\neg(p \to q)\} \cup R)$ is in $\mathcal{Y}_{\neg\to}$.

Subcase 2b: r is an bi-implication $p \leftrightarrow q$. I show $Y \in \mathcal{Y}_{\neg\leftrightarrow}$. Write $R = R^1 \cup R^2 \cup T$ with $R^1 := R \cap \{p \to s, p \leftrightarrow s, s \leftrightarrow p : s \in \mathbf{L}\}$, $R^2 := R \cap \{q \to s, q \leftrightarrow s, s \leftrightarrow q : s \in \mathbf{L}\}$ and $T := R \backslash (X_p \cup X_q)$. As $Y = \{\neg(p \leftrightarrow q)\} \cup R$ is minimal inconsistent, R is minimal subject to entailing $p \leftrightarrow q$, i.e. minimal subject to entailing each of $p \to q$ and $q \to p$. So, by Lemma 7 and using that $R_p = R^1_p$ and $R_q = R^2_q$, the set R is minimal subject to satisfying both (a) $\mathrm{C}(q) \backslash \mathrm{C}(p) \subseteq \bigcup_{s \in R^1_p} \mathrm{C}(s)$ and (b) $\mathrm{C}(p) \backslash \mathrm{C}(q) \subseteq \bigcup_{s \in R^2_q} \mathrm{C}(s)$. It follows that $R = R^1 \cup R^2$ (i.e. $T = \emptyset$).

First suppose $q \leftrightarrow p \in R^1$ or $q \leftrightarrow p \in R^2$. Then $Y = \{\neg(p \leftrightarrow q)\} \cup R \supseteq \{\neg(p \leftrightarrow q), q \leftrightarrow p\}$, hence by minimal inconsistency $Y = \{\neg(p \leftrightarrow q), q \leftrightarrow p\}$. So $Y \in \mathcal{Y}_{\neg\leftrightarrow}$, as desired.

Now suppose $q \leftrightarrow p \notin R^1$ and $q \leftrightarrow p \notin R^2$. As also $p \leftrightarrow q \notin R^1$ and $p \leftrightarrow q \notin R^2$ by (*), we have $R^1 \cap R^2 = \emptyset$. This and the fact that the set $Y = R^1 \cup R^2$ is minimal subject to (a)&(b) imply that R^1 is minimal subject to (a) and that R^2 is minimal subject to (b). So (like in Subcase 2a) $R^1_p \in X_{p \to q}$ with $R^1 = \{p_s : s \in R_p\}$ for some $p_s \in \{p \to s, p \leftrightarrow s, s \leftrightarrow p\}$, $s \in R_p$, and $R^2_q \in X_{q \to p}$ with $R^2 = \{q_s : s \in R_q\}$ for some $q_s \in \{q \to s, q \leftrightarrow s, s \leftrightarrow q\}$, $s \in R_q$. So Y $(= \{\neg(p \to q)\} \cup R^1 \cup R^2)$ is in $\mathcal{Y}_{\neg\leftrightarrow}$, as desired. ■

By Lemmas 9 and 10, we can apply Corollary 5 to characterise consistent quota rules. I finally prove that this characterisation can be simplified into that in Theorem 2.

Proof of Theorem 2. Let $F_{(m_p)_{p \in X^+}}$ be a quota rule, and $\mathcal{Y}, \mathcal{Y}_{\to}, \mathcal{Y}_{\leftrightarrow}, \mathcal{Y}_{\neg\to}, \mathcal{Y}_{\neg\leftrightarrow}$ the sets defined above. By Corollary 5 (using Lemmas 6, 9 and 10) I have to show that (a)&(b) hold iff for all $Y \in \mathcal{Y}$ $(= \mathcal{Y}_{\to} \cup \mathcal{Y}_{\leftrightarrow} \cup \mathcal{Y}_{\neg\to} \cup \mathcal{Y}_{\neg\leftrightarrow})$

$$\sum_{y \in Y} (n - m_y) < n. \tag{21}$$

I will build up this equivalence in the following four steps.

Claim 1. The LHS inequalities in (a) hold iff (21) holds for all $Y \in \mathcal{Y}_{\rightarrow}$.

Claim 2. Given (b), the RHS inequalities in (a) hold iff (21) holds for all $Y \in \mathcal{Y}_{\neg\rightarrow}$.

Claim 3. (b) holds iff (21) holds for all $Y \in \mathcal{Y}_{\leftrightarrow}$.

By Claims 1-3, (a)&(b) hold iff (21) holds for all $Y \in \mathcal{Y}_{\rightarrow} \cup \mathcal{Y}_{\leftrightarrow} \cup \mathcal{Y}_{\neg\rightarrow}$; which is the case iff (21) holds for all $Y \in \mathcal{Y}_{\rightarrow} \cup \mathcal{Y}_{\leftrightarrow} \cup \mathcal{Y}_{\neg\rightarrow} \cup \mathcal{Y}_{\neg\leftrightarrow}$, because of our last claim which completes the proof.

Claim 4. If (21) holds for all $Y \in \mathcal{Y}_{\rightarrow} \cup \mathcal{Y}_{\leftrightarrow}$ then it holds for all $Y \in \mathcal{Y}_{\neg\leftrightarrow}$ (hence the inequalities for $Y \in \mathcal{Y}_{\neg\leftrightarrow}$ are redundant in the system).

Proof of Claim 1. The inequalities (21) for all $Y \in \mathcal{Y}_{\rightarrow}$ are given by

$$(n - m_{\neg a}) + (n - m_{p\rightarrow q}) + \sum_{a' \in \mathrm{C}(p)} (n - m_{a'}) < n \ \forall p \rightarrow q \in X \ \forall a \in \mathrm{C}(q) \backslash \mathrm{C}(p).$$

Using that $n - m_{\neg a} = m_a - 1$, these inequalities can be rewritten as

$$m_a + \sum_{a' \in \mathrm{C}(p)} (n - m_{a'}) \leq m_{p\rightarrow q} \ \forall p \rightarrow q \in X \ \forall a \in \mathrm{C}(q) \backslash \mathrm{C}(p),$$

which by taking the maximum over a is equivalent to the LHS inequalities in (a).

Proof of Claim 2. Suppose (b). The inequalities (21) for all $Y \in \mathcal{Y}_{\neg\rightarrow}$ are given by

$$n - m_{\neg(p\rightarrow q)} + \sum_{s \in S} (n - m_{p_s}) < n \quad \begin{array}{l} \forall p \rightarrow q \in X \ \forall S \in X_{p\rightarrow q} \\ \forall (p_s)_{s\in S} \in (\{p \rightarrow s, p \leftrightarrow s, s \leftrightarrow p\} \cap X)^S. \end{array}$$

These inequalities can (by $n - m_{\neg(p\rightarrow q)} = m_{p\rightarrow q} - 1$) be rewritten as

$$m_{p\rightarrow q} + \sum_{s \in S} (n - m_{p_s}) \leq n \quad \begin{array}{l} \forall p \rightarrow q \in X \ \forall S \in X_{p\rightarrow q} \\ \forall (p_s)_{s\in S} \in (\{p \rightarrow s, p \leftrightarrow s, s \leftrightarrow p\} \cap X)^S, \end{array}$$

or equivalently as

$$m_{p\rightarrow q} + \max_{(p_s)_{s\in S} \in (\{p\rightarrow s, p\leftrightarrow s, s\leftrightarrow p\} \cap X)^S} \sum_{s \in S} (n - m_{p_s}) \leq n \ \forall p \rightarrow q \in X \ \forall S \in X_{p\rightarrow q}. \quad (22)$$

Note that

$$\max_{(p_s)_{s\in S} \in (\{p\rightarrow s, p\leftrightarrow s, s\leftrightarrow p\} \cap X)^S} \sum_{s \in S} (n - m_{p_s}) = \sum_{s \in S} (n - \min_{p_s \in \{p\rightarrow s, p\leftrightarrow s, s\leftrightarrow p\} \cap X} m_{p_s}). \quad (23)$$

For all $s \in S$ and all $p_s \in \{p \leftrightarrow q, q \leftrightarrow p\}$ we have $m_{p_s} = n$ by (b). So, for all $s \in S$, $\min_{p_s \in \{p\rightarrow s, p\leftrightarrow s, s\leftrightarrow p\} \cap X} m_{p_s}$ is n if $p \rightarrow s \notin X$ and $m_{p\rightarrow s}$ if $p \rightarrow s \in X$. Hence in (23) the term $(n - \min_{p_s \in \{p\rightarrow s, p\leftrightarrow s, s\leftrightarrow p\} \cap X} m_{p_s})$ drops out if $p \rightarrow s \notin X$ and equals $(n - m_{p\rightarrow s})$ if $p \rightarrow s \in X$. Therefore (23) implies

$$\max_{(p_s)_{s\in S} \in (\{p\rightarrow s, p\leftrightarrow s, s\leftrightarrow p\} \cap X)^S} \sum_{s \in S} (n - m_{p_s}) = \sum_{s \in S : p\rightarrow s \in X} (n - m_{p\rightarrow s}).$$

Using this, the inequalities (22) are equivalent to

$$m_{p\rightarrow q} + \sum_{s \in S : p\rightarrow s \in X} (n - m_{p\rightarrow s}) \leq n \ \forall p \rightarrow q \in X \ \forall S \in X_{p\rightarrow q},$$

and hence, as desired, to

$$m_{p\to q} + \max_{S\in X_{p\to q}} \sum_{s\in S: p\to s\in X} (n - m_{p\to s}) \leq n \ \forall p \to q \in X.$$

Proof of Claim 3. 1. First assume (21) holds for all $Y \in \mathcal{Y}_{\leftrightarrow}$, and let $p \leftrightarrow q \in X$.

1.1. Here I show $m_{p\leftrightarrow q} = n$. As $p \leftrightarrow q$ is non-degenerate, there exist $a \in \mathrm{C}(p)\backslash\mathrm{C}(q)$ and $b \in \mathrm{C}(q)\backslash\mathrm{C}(p)$. By assumption,

$$\begin{aligned}(n - m_{p\leftrightarrow q}) + (n - m_{\neg b}) + \sum_{a'\in \mathrm{C}(p)} (n - m_{a'}) &< n,\\ (n - m_{p\leftrightarrow q}) + (n - m_{\neg a}) + \sum_{b'\in \mathrm{C}(q)} (n - m_{b'}) &< n.\end{aligned}$$

Rewriting this (by using that $n - m_{\neg s} = m_s - 1$ for all $s \in X$), we obtain

$$\begin{aligned}m_{p\leftrightarrow q} - m_b + 1 &> \sum_{a'\in \mathrm{C}(p)} (n - m_{a'}) \geq n - m_a, \qquad (24)\\ m_{p\leftrightarrow q} - m_a + 1 &> \sum_{b'\in \mathrm{C}(q)} (n - m_{b'}) \geq n - m_b.\end{aligned}$$

So

$$m_{p\leftrightarrow q} \geq n - m_a + m_b \text{ and } m_{p\leftrightarrow q} \geq n - m_b + m_a. \qquad (25)$$

Adding both inequalities, we get $2m_{p\leftrightarrow q} \geq 2n$, whence $m_{p\leftrightarrow q} = n$.

1.2. Next I show that all $a \in \mathrm{C}(p)\Delta\mathrm{C}(q)$ have the same threshold. As $\mathrm{C}(p)\Delta\mathrm{C}(q)$ is the union of the non-empty sets $\mathrm{C}(p)\backslash\mathrm{C}(q)$ and $\mathrm{C}(q)\backslash\ \mathrm{C}(p)$, it is sufficient to show that $m_a = m_b$ for all $a \in \mathrm{C}(p)\backslash\mathrm{C}(q)$ and $b \in \mathrm{C}(q)\backslash\mathrm{C}(p)$. Consider such a, b. The argument in 1.1 yields (25), which by $m_{p\leftrightarrow q} = n$ implies $m_a \geq m_b$ and $m_b \geq m_a$, whence $m_a = m_b$.

1.3. Let m be the common threshold of all $a \in \mathrm{C}(p)\Delta\mathrm{C}(q)$. I suppose $m < n$ and show that $|\mathrm{C}(p)\Delta\mathrm{C}(q)| \leq 2$. The first inequality in (24) (where $b \in \mathrm{C}(q)\Delta\mathrm{C}(p)$) implies

$$m_{p\leftrightarrow q} - m_b + 1 > \sum_{a'\in \mathrm{C}(p)\backslash\mathrm{C}(q)} (n - m_{a'}),$$

which after substituting $m_{p\leftrightarrow q} = n$ and $m_b = m_{a'} = m$ gives

$$n - m \geq |\mathrm{C}(p)\backslash\mathrm{C}(q)|(n - m), \text{ i.e. } |\mathrm{C}(p)\backslash\mathrm{C}(q)| \leq 1.$$

It can be shown similarly that $|\mathrm{C}(q)\backslash\mathrm{C}(p)| \leq 1$. So $|\mathrm{C}(p)\Delta\mathrm{C}(q)| \leq 2$.

1.4. Finally, let $a'' \in \mathrm{C}(p) \cap \mathrm{C}(q)$. I show that $m_{a''} = n$. Let a, b be as in 1.1. The first inequality in (24) implies

$$m_{p\leftrightarrow q} - m_b + 1 > (n - m_{a''}) + (n - m_a),$$

which by $m_{p\leftrightarrow q} = n$ and $m_a = m_b$ implies $1 > (n - m_{a''})$, i.e. $m_{a''} = n$.

2. Conversely, assume (b). Consider any $Y \in \mathcal{Y}_{\leftrightarrow}$, say $Y = \{r, \neg a\} \cup \mathrm{C}(p)$ where $r \in \{p \leftrightarrow q, q \leftrightarrow p\}$ and $a \in \mathrm{C}(q)\backslash\mathrm{C}(p)$, and let me show (21). Using (b) and $n - m_{\neg a} = m_a - 1$,

$$\sum_{y\in Y}(n - m_y) = m - 1 + |\mathrm{C}(p)\backslash\mathrm{C}(q)|(n - m),$$

where m denotes the common threshold of all $a' \in \mathrm{C}(p)\Delta\mathrm{C}(q)$. Note that if $|\mathrm{C}(p)\backslash\mathrm{C}(q)| \geq 2$ then $|\mathrm{C}(p)\Delta\mathrm{C}(q)| \geq 3$, hence $m = n$. So, as desired,

$$\sum_{y\in Y}(n - m_y) = \begin{cases} m - 1 + n - m < n & \text{if } |\mathrm{C}(p)\backslash\mathrm{C}(q)| = 1 \\ n - 1 + 0 < n & \text{if } |\mathrm{C}(p)\backslash\mathrm{C}(q)| \geq 2. \end{cases}$$

Proof of Claim 4. Suppose (21) holds for all $Y \in \mathcal{Y}_{\rightarrow} \cup \mathcal{Y}_{\leftrightarrow}$. Consider any $Y \in \mathcal{Y}_{\neg\leftrightarrow}$, say (in the earlier notation) $Y = \{\neg(p \leftrightarrow q)\} \cup \{p_s : s \in S\} \cup \{q_s : s \in S'\}$. To prove the corresponding inequality,

$$(n - m_{\neg(p\leftrightarrow q)}) + \sum_{s\in S}(n - m_{p_s}) + \sum_{s\in S'}(n - m_{q_s}) < n,$$

I show that $m_{p_s} = n \ \forall s \in S$ and that $m_{q_s} = n \ \forall s \in S'$; in fact, I only show the former as the latter holds analogously. Let $s \in S$. Recall that $S \in X_{p\rightarrow q}$ and $p_s \in \{p \rightarrow s, p \leftrightarrow s, s \leftrightarrow p\}$.

If $p_s \in \{p \leftrightarrow s, s \leftrightarrow p\}$ then already by Claim 3 $m_{p_s} = n$, as desired.

Now assume $p_s = p \rightarrow s$. Since $s \in S \in X_{p\rightarrow q}$, $\mathrm{C}(q)\backslash\mathrm{C}(p)$ is a subset of $\cup_{s^*\in S}\mathrm{C}(s^*)$ but not of $\cup_{s^*\in S\backslash\{s\}}\mathrm{C}(s^*)$. So there is a $b \in \mathrm{C}(s) \cap \mathrm{C}(q)\backslash\mathrm{C}(p)$. Moreover, as $p \leftrightarrow q$ is non-degenerate, there is an $a \in \mathrm{C}(p)\backslash\ \mathrm{C}(q)$. As $a, b \in \mathrm{C}(p)\Delta\mathrm{C}(q)$, we have $m_a = m_b$ by Claim 3. Using Claim 1,

$$m_{p\rightarrow q} \geq \sum_{a'\in\mathrm{C}(p)}(n - m_{a'}) + \max_{b'\in\mathrm{C}(q)\backslash\mathrm{C}(p)} m_{b'} \geq n - m_a + m_b = n,$$

whence $m_{p\rightarrow q} = n$, as desired. ■

Chapter 7

On the informational basis of collective judgments

Paper: Aggregation theory and the relevance of some issues to others, *unpublished manuscript*, London School of Economics, 2006

Aggregation theory and the relevance of some issues to others

Franz Dietrich

Abstract. I propose a general collective decision problem consisting in many issues that are interconnected in two ways: by mutual constraints and by connections of relevance. Aggregate decisions should respect the mutual constraints, and be based on relevant information only. This general informational constraint has many special cases, including premise-basedness and Arrow's independence condition; they result from special notions of relevance. The existence and nature of (non-degenerate) aggregation rules depends on both types of connections. One result, if applied to the preference aggregation problem and adopting Arrow's notion of (ir)relevance, becomes Arrow's Theorem, without excluding indifferences unlike in earlier generalisations.

Keywords: aggregation, informational constraints, (ir)relevance, premise-based procedure

JEL Classification Numbers: D70, D71

1 Introduction

Most complex decision problems can be formalised as consisting of many binary decisions: decisions of accepting or rejecting certain propositions. For instance, establishing a preference relation R over a given set of alternatives Q consists in deciding, for each pair of alternatives $x, y \in Q$, whether or not xRy. Judging the values of different variables consists of judging, for each variable V and each of its potential values v, whether or not $V = v$. Producing a report that contains qualitative economic forecasts might involve deciding for or against many propositions: atomic ones like "inflation will increase" and compound ones like "*if* consumption will increase *and* foreign demand does *not* decrease, *then* inflation will increase" (where logical operators are italicised).

Although this division into binary issues is usually possible, there are arguably two distinct types of interconnections – to be called *logical* connections and *relevance* connections – that can prevent us from treating the issues independently. First, the decisions on the issues may logically constrain each other; in the above examples, the preference judgments must respect conditions like transitivity, the variables might constrain each other, and the propositions stated in the economic report must be logically consistent with each other, respectively. Second – and this is the topic of the paper – some issues may be *relevant* to (the decision on) other issues. The nature and interpretation of relevance connections is context-specific. A proposition r may be relevant to another one p on the grounds that r is an (argumentative) premise of p, or that r is a causal factor bringing about p, or that r and p share some other (semantic) relation. Relevance connections are not reducible to logical connections. Two issues – say, whether traffic lights are necessary and whether the diplomatic relations to a

country should be interrupted – may be considered irrelevant to each other and yet be indirectly logically related via other issues under consideration. Conversely, an issue – say that of whether country X has weapons of mass destruction – may be considered relevant to another issue – say that of whether measure Y against country X is appropriate – without a (direct or indirect) logical connection in the complex decision problem considered.

Now suppose that the complex decision problem is faced by a group of individuals and should be settled by aggregating the individual judgments on each proposition (issue). Many concrete aggregation models and procedures in the literature in effect account, in different ways, both for logical connections and relevance connections. Logical connections are represented by delimiting the set of admissible decisions, for instance in the form of rationality conditions like transitivity in preference aggregation, or in the form of an overall budget constraint if different budget items are decided simultaneously. By contrast, relevance connections are accounted for through "informational" constraints on the way in which the decision (output of the aggregation rule) may depend on the individuals' input: only *relevant* information may be used. For instance, Arrow's condition of *independence of irrelevant alternatives* ("IIA") excludes the use of (arguably) irrelevant information. The *premise-based procedure* in judgment aggregation makes the decision on certain conclusion-type propositions dependent on people's judgments on other premise-type propositions considered relevant. In general, the question of "what is relevant to what?" may be controversial: some researchers reject Arrow's IIA condition, and in judgment aggregation it may be unclear which propositions to consider as premises and which as conclusions, and moreover the same conclusion-type proposition could be explained in more than one way in terms of premises.

While accounted for in concrete aggregation problems and procedures, the notions of relevance and of (ir)relevant information have not been treated in general terms. As relevance connections are not reducible to logical connections, both connections should be separate ingredients of a general aggregation model. More precisely, I propose to consider, in addition to logical connections, a (binary) *relevance* relation $\mathcal{R}$ between propositions (issues), and to aggregate in accordance with *independence of irrelevant information* ("III"). To allow broad applications, I leave general the type of complex decision problem and the interpretation and relation-theoretic properties of the relevance relation $\mathcal{R}$: it might be highly partial (few inter-relevances) or close to complete (many inter-relevances), and it need not be symmetric, or transitive, or reflexive (i.e. self-irrelevance is allowed).

In the special case that every proposition is considered relevant just to itself (i.e. $p\mathcal{R}q \Leftrightarrow p = q$ for any propositions p, q), III reduces to the restrictive condition of *proposition-wise independence* (often simply called *independence*): here, each proposition is decided via an isolated vote, using an arbitrary voting rule but ignoring people's judgments on other propositions. A number of general results have been obtained on proposition-wise independent aggregation, in abstract aggregation models (starting with Wilson 1975) or models of logic-based judgment aggregation (starting with List and Pettit 2002). Essentially, these results establish limits to the possibility of (non-degenerate) proposition-wise independent aggregation in the presence of logical connections between propositions. Impossibility results with necessary conditions on logical connections are derived, for instance, by Wilson (1975),

List and Pettit (2002), Pauly and van Hees (2006), Dietrich (2006), Gärdenfors (2006), Mongin (2005-a) and van Hees (forthcoming). Nehring and Puppe (2002, 2005, 2006) derive the first results with minimal conditions on logical connections, and Dokow and Holzman (2005) introduce minimal conditions of an algebraic kind. Other (im)possibility results are given, for instance, in Dietrich (forthcoming), Dietrich and List (forthcoming-a, forthcoming-b) and Nehring (2005). Possibilities of proposition-wise independent aggregation arise if the individual judgments fall into particular domains (List 2003, Dietrich and List 2006) or if logical connections are modelled using subjunctive implications (Dietrich 2005).

The proposition-wise independence condition is often criticised (e.g., Chapman 2002, Mongin 2005-a), but has rarely been weakened in the general aggregation literature. The normative appeal of the condition is easily challenged by concrete examples: why, for instance, should the collective judgment on whether to introduce taxes on kerosene be independent of people's judgments on whether global warming should be prevented? All weaker independence conditions proposed in the literature are special cases of III: each implicitly uses some notion of relevance $\mathcal{R}$. Let me mention the literature's two most notable independence weakenings.[1]

One departure from proposition-wise independence aims to represent non-binary variables.[2] Suppose again the decision problem consists in estimating the values of different typically non-binary (interconnected) variables V like GDP growth. Then propositions take the form $V = v$, where V is a variable and v belongs to a set $Rge(V)$ of possible values of V. Suppose the collective estimate of each variable V must be a function of people's estimates of V (e.g. a weighted average). Then the collective judgment on whether $V = v$ depends on people's attitudes towards the propositions $V = v', v' \in Rge(V)$ (each individual accepts exactly one of them).[3] So aggregation is *variable-wise* independent – not *proposition-wise* as the decision on whether $V = v$ depends not just on people's views on whether $V = v$. Variable-wise independence is an example of III, where any $V = v$ and $V = v'$ are now inter-relevant. Variable-wise independence is often imposed: for instance in probability aggregation theory, where a variable is an event's probability and variable-wise independence leads (under other constraints) to *linear* aggregation rules (e.g. Genest and Zidek 1986); or in abstract aggregation theory, where Rubinstein and Fishburn (1986) derive more general linearity results on variable-wise independent aggregation; or in judgment aggregation, where Claussen and Roisland (2005) introduce a variable-wise version of the discursive paradox and show results on when it occurs. Also Pauly and van Hees' (2006) multi-valued logic approach can be viewed as using variable-wise independence.

A second weakening of proposition-wise independence aims to represent the different status of different propositions. Here the independence condition is applied

[1] A more radical move consists in imposing *no* independence condition (informational constraint) on aggregation. This route is taken in the literature on *belief merging* in artificial intelligence (e.g. Konieczni and Pino-Perez 2002, Eckert and Pigozzi 2005). Apart from the absence of informational constraints, belief merging is closely related to judgment aggregation: it also aims to merge sets of logical propositions.

[2] It was originally not intended as a weakening of proposition-wise independence, but is one under our division of the decision problem into (binary) decisions on propositions of the form $V = v, v \in Rge(V)$.

[3] E.g. the collective accepts $V = v$ if and only if v is a certain weighted average of the individual estimates v' of V, i.e. of the values $v' \in Rge(V)$ for which $V = v'$ is accepted.

only to *some* propositions, e.g. to "premises" (Dietrich 2006) or to *atomic* propositions (Mongin 2005-a). Mongin (2005-a) argues that the collective judgment on a compound proposition like $p \wedge q$ should not ignore how the individuals judge p and judge q; our relevance relation $\mathcal{R}$ would then have to satisfy $p\mathcal{R}(p \wedge q)$ and $q\mathcal{R}(p \wedge q)$.

This paper has an expository and a technical focus. On the expository dimension, I introduce the relevance-based aggregation model; I discuss different types of relevance relations, including transitive relevance, asymmetric relevance, and relevance as premisehood; I introduce relevance-based conditions of III, agreement preservation and dictatorship (generalising for instance Arrow's conditions of IIA, weak Pareto and weak dictatorship); and I introduce significantly generalised forms of premise-based and prioritarian aggregation rules. On the technical dimension, I prove two possibility and four impossibility results on III aggregation. One result, if applied to the preference aggregation problem, becomes Arrow's Theorem. While Arrow's Theorem has been generalised earlier under the simplifying assumption that not only individuals but also the collective are never indifferent between distinct options,[4] one might view as an embarrassment of the growing literature that, despite its intended generality, its theorems do not generalise Arrow's (unrestricted) theorem; and its aggregation conditions do not have as special cases Arrow's conditions of IIA, weak Pereto and weak dictatorship.

2 Basic definitions

We consider a set $N = \{1, ..., n\}$ of individuals, where $n \geq 2$, faced with a collective decision problem of a general kind.

Agenda, judgment sets. The *agenda* is an arbitrary non-empty (possibly infinite) set X of *propositions* on which a decision (acceptance or rejection) is needed. The agenda includes negated propositions: $X = \{p, \neg p : p \in X^+\}$, where X^+ is some set of non-negated propositions and "$\neg p$" stands for "not p". Notationally, double-negations cancel each other out.[5] A *judgment set* is a set $A \subseteq X$ of (accepted) propositions; it is *complete* if it contains a member of each pair $p, \neg p \in X$ ("no abstentions").

Logical interconnections. Not all judgment sets are consistent. For the agenda $X = \{a, \neg a, b, \neg b, a \wedge b, \neg(a \wedge b)\}$, the (complete) judgment set $\{a, b, \neg(a \wedge b)\}$ is inconsistent. Let $\mathcal{J}$ be a non-empty set of judgment sets, each containing exactly one member of each pair $p, \neg p \in X$, and suppose the *consistent* judgment sets are precisely the sets in $\mathcal{J}$ and their subsets; all other judgment sets are *inconsistent*.[6] A judgment set $A \subseteq X$ *entails* a proposition $p \in X$ (written $A \vdash p$) if $A \cup \{\neg p\}$ is inconsistent. I write $q \vdash p$ for $\{q\} \vdash p$.

It is natural (though for the present results not necessary) to take the propositions in X to be *statements* of a formal language, and to take consistency/entailment to

[4] See Wilson (1975), Dietrich and List (forthcoming-b), and Dokow and Holzman (2005). Nehring (2003) shows an Arrow-like result.

[5] That is: whenever I write "$\neg q$" (where $q \in X$), I mean the other member of the pair $p, \neg p \in X$ to which q belongs; hence "$\neg\neg q$" stands for q.

[6] So $\mathcal{J}$ contains the consistent *and complete* judgment sets. Any set $\{p, \neg p\} \subseteq X$ is inconsistent. Any subset of a consistent set is consistent. Finally, $\emptyset$ is consistent, and any consistent set has a superset that is consistent and complete (hence in $\mathcal{J}$).

be standard logical consistency/entailment, as is usually assumed in the judgment aggregation literature. The formal language, if sufficiently expressive, can mimic the natural language in which the real decision problem arises.[7]

A proposition $p \in X$ is a *contradiction* if $\{p\}$ is inconsistent, and a *tautology* if $\{\neg p\}$ is inconsistent. I call $A \subseteq X$ *consistent with* $B \subseteq X$ if $A \cup B$ is consistent; and I call $A \subseteq X$ *consistent with* $p \in X$ (and p *consistent with* A) if $A \cup \{p\}$ is consistent.

Aggregation. The *(judgment) aggregation rule* is a function F that assigns to every profile $(A_1, \ldots, A_n)$ of (individual) judgment sets in some domain of "admissible" profiles a (collective) judgment set $F(A_1, \ldots, A_n) = A \subseteq X$. It is often required that F has *universal* domain, i.e. allows as an input precisely all profiles $(A_1, ..., A_n)$ of consistent and complete (individual) judgment sets. An important question is how rational the (collective) judgment sets generated by F are: are they consistent? Complete? If F has universal domain and consistent and complete outputs, it is a function $F : \mathcal{J}^n \to \mathcal{J}$. Majority rule on $\mathcal{J}^n$, given by

$$F(A_1, ..., A_n) = \{p \in X : |\{i : p \in A_i\}| > n/2\} \text{ for all } (A_1, ..., A_n) \in \mathcal{J}^n,$$

may for most agendas generate inconsistent outputs: it is not a function $F : \mathcal{J}^n \to \mathcal{J}$.

Abstract aggregation. One may (re)interpret the elements of X as arbitrary *attributes*, which may but need not be *propositions/judgments*, and may but need not be expressed in formal logic. Then judgment sets become *attribute sets*, and the aggregation rule maps profiles of individual attribute sets to a collective attribute set. Of course, the attribute holders $i \in N$ need not be humans.

I give two examples here; more examples follow in the next section.

Example 1: preference aggregation. For a given set of (exclusive) alternatives Q ($|Q| \geq 3$), consider the agenda

$$X := \{xRy, \neg xRy : x, y \in Q\} \text{ (the } \textit{preference agenda}\text{)},$$

where xRy is the proposition "x is at least as good as y". Throughout the paper, I often write xPy for $\neg yRx$. Let $\mathcal{J}$ be the set of all judgment sets $A \subseteq X$ that represent fully rational preferences, i.e. for which there exists a weak ordering[8] $\succeq$ on Q such that

$$A = \{xRy \in X : x \succeq y\} \cup \{\neg xRy \in X : x \not\succeq y\}.$$

Note that there is a bijective correspondence between weak orderings on Q and judgment sets in $\mathcal{J}$; and between judgment aggregation rules $F : \mathcal{J}^n \to \mathcal{J}$ and Arrowian social welfare functions (with universal domain). The agenda X and its consistency

[7] The formal language can be one of classical (propositional or predicate) logic or one of a non-classical logic such as a modal logic, as long as the logic satisfies certain regularity conditions. This follows Dietrich's (forthcoming) model of judgment aggregation in general logics, which generalises List and Pettit's (2002) original model in classical propositional logic.

[8] A weak ordering on Q is a binary relation $\succeq$ on Q that is reflexive, transitive, and connected (but not necessarily anti-symmetric, so that non-trivial indifferences are allowed).

notion belong to a predicate logic, as defined in Dietrich (forthcoming), drawing on List and Pettit (2004).[9]

Example 2: judging values of and constraints between variables. Suppose a group (e.g. a central bank's board or research panel) debates the values of different variables (e.g. macroeconomic variables measuring GDP, prices or consumption). Let $\mathbf{V}$ be a non-empty set of "variables". For each $V \in \mathbf{V}$ let $Rge(V)$ be a non-empty set of possible "values" of V (numbers or other objects), called the *range* of V. For any variable $V \in \mathbf{V}$ and any value $v \in Rge(V)$, the group has to judge the proposition $V = v$ stating that V takes the value v.[10] These judgments should respects the (causal) constraints between variables; but, not surprisingly, the nature of these constraints is itself disputed, for instance because the group members believe in different (e.g. econometric) estimation techniques. If the variables are real-valued, some *linear* constraints like $V + 3W - U = 5$, or non-linear ones like $V^2 = W$, might be debated. Let $\mathbf{C}$ be any non-empty set of "constraints" under consideration.[11] The agenda is given by

$$X = \{V = v, \neg(V = v) : V \in \mathbf{V}, v \in Rge(V)\} \cup \{c, \neg c : c \in \mathbf{C}\}.$$

A judgment set $A \subseteq X$ thus states that certain variables do (not) take certain values, and that the variables do (not) constrain each other in certain ways. To define logical connections, note first that some constraints may conflict with others (e.g., $V > W$ conflicts with $W > V$), and that some constraints may conflict with other *negated* constraints (e.g., $V \log(W) > 2$ conflicts with $\neg(V \log(W) > 0)$). Let $\mathcal{J}^*$ be some non-empty set containing for each constraint $c \in \mathbf{C}$ either c or $\neg c$ (not both); the sets in $\mathcal{J}^*$ represent consistent judgments on the constraints. Now let $\mathcal{J}$ be the set of all judgment sets $A \subseteq X$ containing exactly one member of each pair $p, \neg p \in X$ such that:[12]

(i) each variable $V \in \mathbf{V}$ has a single value $v \in Rge(V)$ with $V = v \in A$;
(ii) the family of values in (i) obeys all accepted constraints $c \in A \cap \mathbf{C}$;
(iii) the judgments on constraints are consistent: $A \cap \{c, \neg c : c \in \mathbf{C}\} \in \mathcal{J}^*$.

Note that it may be consistent to hold a negated constraint $\neg c$ and yet to assign values to variables in accordance with c. Indeed, variables can stand in certain relations by pure coincidence, i.e. without a constraint to this effect.[13]

[9] See Dietrich and List (forthcoming-b) for a logic representing *strict* preference aggregation.

[10] More generally, the group might consider propositions stating that V's value belongs to certain sets $S \subseteq val(V)$.

[11] A constraint might be formalised by a subset of the "joint range" $\Pi_{V \in \mathbf{V}} Rge(V)$ of the family of variables $(V)_{V \in \mathbf{V}}$ (e.g. a subset of $\mathbf{R}^3$ if $\mathbf{V}$ consists of three real-valued variables), or by an expression in a logical language (see below).

[12] It is easily possible to add *exogenous* constraints (which cannot be rejected, unlike those in $\mathbf{C}$), by further restricting in (iii) the allowed value assignments.

[13] More precisely, a constraint states not just an *actual* relation r between variables but a *necessary* relation "necessarily r", that is (in modal logical terms) "in all possible worlds r" ($\Box r$). The negation of this constraint ($\neg\Box r$) is equivalent to "possibly $\neg r$" ($\Diamond\neg r$), which is indeed consistent with "r", i.e. with the relation holding.

3 Independence of irrelevant information

The conditions I will impose on the aggregation rule are based on a *relevance relation*, whose nature and interpretation is context-specific, as indicated earlier. Such a relevance relation is not simply reducible to logical interconnections (of inconsistency or entailment). Suppose the proposition a : "country X has weapons of mass destruction" (and $\neg a$) is considered relevant to the proposition b: "country X should be attacked" (and to $\neg b$), but not vice versa. This asymmetry of relevance between the two issues need not be reflected in logical connections: $\mathcal{J}$ can be perfectly symmetric in the two issues. This is clear if X contains no issues other than these two (logically independent) ones, i.e. if $X = \{a, \neg a, b, \neg b\}$. But even additional propositions in X that create (indirect) logical links between the two issues need not reveal a direction of relevance, as is seen from examples.[14] Hence any relevance relation derived from logical interconnections would have to declare, against our intuition, the two issues as either mutually relevant or mutually irrelevant.

So relevance must be taken on board as an additional structure. I do this in the form of a *relevance relation*. Not any binary relation on X can reasonably count as a relevance relation. I call a binary relation $\mathcal{R}$ on the agenda X a *relevance relation* (where "$r\mathcal{R}p$" means "r is relevant to p") if the following condition holds.

No underdetermination. Each $p \in X$ is settled by the judgments on the relevant propositions: for every consistent set $E \subseteq \{r, \neg r : r\mathcal{R}p\}$ containing a member of each pair $r, \neg r$ in $\{r, \neg r : r\mathcal{R}p\}$, either $E \vdash p$ or $E \vdash \neg p$. (I call such an E an *($\mathcal{R}$-)explanation* of p or *($\mathcal{R}$-)refutation* of p, respectively.)
No underdetermination. Each $p \in X$ is settled by the judgments on the relevant propositions: for every consistent set $E = \{r^* : r\mathcal{R}p\}$ where each r^* is r or $\neg r$, either $E \vdash p$ or $E \vdash \neg p$. (I call such an E an *($\mathcal{R}$-)explanation* of p or *($\mathcal{R}$-)refutation* of p, respectively.)

This definition of a relevance relation has two main characteristics.

First, it requires no relation-theoretic properties like reflexivity or symmetry. This generality is essential to represent different notions of relevance (see below); and it is appropriate since no relation-theoretic property is uncontroversially adequate for all decision problems. Below I suggest relation-theoretic conditions on relevance for special decision problems, but different ones across decision problems.

Second, it requires "no underdetermination": a proposition's truth value must be fully determined by the relevant propositions' truth values. To illustrate this condition (which I justify in the next section), note first that it holds trivially for *self-relevant* propositions $p \in X$, as p's truth value settles p's truth value; here all explanations of p contain p, and all refutations of p contain $\neg p$.[15] In particular, all reflexive

[14] Suppose $X = \{a, \neg a, a \to b, \neg(a \to b), b, \neg b\}$. The symmetry argument is simple. A truth-value assignment $(t_1, t_2, t_3) \in \{T, F\}^3$ to the propositions $a, a \to b, \neg b$ is consistent if and only if (t_3, t_2, t_1) (in which the truth-values of a and $\neg b$ are interchanged) is consistent. This is so whether $a \to b$ represents a *subjunctive* or a *material* implication. In the first case, the only inconsistent truth-value assignment is (T, T, T). In the second case, there are other inconsistent truth-value assignments (as $a \to b$ is equivalent to $\neg a \vee b$), yet without breaking the symmetry between a and $\neg b$.

[15] If p is the *only* proposition relevant to p, p's only explanation is $\{p\}$ (except if p is a contradiction: then p has no explanation), and p's only refutation is $\{\neg p\}$ (except if $\neg p$ is a contradiction: then p has no refutation).

relations $\mathcal{R}$ satisfy "no underdetermination", i.e. are relevance relations. This said, "no underdetermination" is a weak condition. It only has a bite for propositions that are non-self-relevant, hence "externally" explained. For instance, suppose to a conjunction $a \wedge b$ only the conjuncts a and b, not $a \wedge b$ itself, are deemed relevant. (Such an idea underlies the premise-based procedure for the agenda given by $X^+ = \{a, b, a \wedge b\}$; see Example 4 below.) Here, $a \wedge b$'s truth value is indeed determined by a's and b's truth values; $a \wedge b$ has a single explanation ($\{a, b\}$) and three possible refutations ($\{\neg a, b\}$, $\{a, \neg b\}$, $\{\neg a, \neg b\}$). Dropping a's or b's relevance to $a \wedge b$ would lead to underdetermination.

Hereafter, let $\mathcal{R}$ be a given relevance relation. I denote the set of propositions relevant to $p \in X$ by $\mathcal{R}(p) := \{r \in X : r\mathcal{R}p\}$. The following condition requires the collective judgment on any proposition $p \in X$ to be formed on the basis of how the individuals judge the propositions relevant to p.

Independence of Irrelevant Information (III). For all propositions $p \in X$ and all profiles $(A_1, ..., A_n)$ and $(A'_1, ..., A'_n)$ in the domain, if $A_i \cap \mathcal{R}(p) = A'_i \cap \mathcal{R}(p)$ for every individual i then $p \in F(A_1, ..., A_n) \Leftrightarrow p \in F(A'_1, ..., A'_n)$.

Many informational constraints on aggregation used in social choice theory can be viewed as being the III condition relative to some notion of relevance. Roughly, the more propositions are relevant to each other, the weaker the informational constraint III is. III is empty if all propositions are relevant to all propositions, i.e. if $\mathcal{R} = X \times X$. III is the standard proposition-wise independence condition if each proposition is just self-relevant, i.e. $\mathcal{R}(p) = \{p\}$ for all $p \in X$. III is Gärdenfors' "weak" (yet still quite strong) independence if $\mathcal{R}(p) = \{p, \neg p\}$ for all $p \in X$. III is Dietrich's (2006) independence restricted to a subset $Y \subseteq X$ if $\mathcal{R}(p) = \{p\}$ for $p \in Y$ and $\mathcal{R}(p) = X$ for $p \in X \backslash Y$. III is Mongin's (2005-a) independence restricted to the atomic propositions (of an agenda X in a propositional language) if $\mathcal{R}(p) = \{p\}$ for atomic p and $\mathcal{R}(p) = X$ for compound p (e.g. $p = a \wedge \neg b$).

I now discuss further examples of relevance relations. These examples make the convenient assumption that relevance is *negation-invariant*:[16]

$p\mathcal{R}q \Leftrightarrow \tilde{p}\mathcal{R}\tilde{q}$ for all $p, q \in X$ and all $\tilde{p} \in \{p, \neg p\}$, $\tilde{q} \in \{q, \neg q\}$ (negation invariance).

So $\mathcal{R}$ is determined by its restriction to the set $X^+ \subseteq X$ of non-negated propositions. Let $\mathcal{R}^+$ be this restriction, and for all $p \in X^+$ let $\mathcal{R}^+(p) := \mathcal{R}(p) \cap X^+$ ($= \{r \in X^+ : r\mathcal{R}p\} = \{r \in X^+ : r\mathcal{R}^+p\}$).

Example 1 (continued). For the preference agenda, III is equivalent to Arrow's *independence of irrelevant alternatives* ("IIA") in virtue of defining relevance by

$$\mathcal{R}^+(xRy) := \{xRy, yRx\} \text{ for all } xRy \in X. \tag{1}$$

I call this the *Arrowian* relevance relation. Indeed, to socially decide on xRy, Arrow considers as relevant whether people weakly prefer x to y *and also* whether they

[16] The relevance relation underlying proposition-wise independence (given by $\mathcal{R}(p) = \{p\}$ for all $p \in X$) is *not* negation-invariant.

weakly prefer y to x. By contrast, the standard proposition-wise independence condition is stronger than IIA, as it denies the relevance of yRx to xRy.

Example 2 (continued). For the agenda of Example 2, one might put

$$\begin{array}{ll} \mathcal{R}^+(V=v) = \{V=v' : v' \in Rge(V)\} & \text{for all } V=v \in X \\ \mathcal{R}^+(c) = \{c\} & \text{for all constraints } c \in \mathbf{C}. \end{array} \tag{2}$$

On a modified assumption, some distinct constraints $c, c' \in \mathbf{C}$ might be declared inter-relevant, for instance if they involve the same variables.

Example 3: relevance as an equivalence relation, and topic-wise independence. Examples 1 and 2 are instances of the general case where relevance is an equivalence relation: $\mathcal{R}$ is reflexive (which requires self-relevance), symmetric, and transitive. Each of these three conditions is a substantial assumption on the notion of relevance. The agenda X is then partitioned into equivalence classes (of inter-relevant propositions), each one interpretable as a *topic*; so III is a *topic-wise* (rather than proposition-wise) independence condition. A topic can be binary (of the form $\{p, \neg p\}$) or non-binary. For the preference agenda (Example 1), the Arrowian relevance relation creates topics of the form $\{xRy, \neg xRy, yRx, \neg yRx\}$ (for options $x, y \in Q$): the topic of x's and y's relative ranking.

An example of topic-wise independence is the *variable-wise* independence condition mentioned in the introduction. Consider a variant of Example 2, in which the inter-variable constraints are exogenously imposed rather than under decision. So the agenda is given by $X^+ = \{V = v : V \in \mathbf{V} \text{ and } v \in Rge(V)\}$, and relevance by $\mathcal{R}^+(V=v) = \{V=v' : v' \in Rge(V)\}$. To each variable $V \in \mathbf{V}$ corresponds an equivalence class: $\{V=v, \neg(V=v) : v \in Rge(V)\}$, the topic of V's value. Judging this topic boils down to specifying a value $v \in Rge(V)$ of V (i.e. $V=v$ is accepted and all $V=v', v' \in Rge(V)\backslash\{v\}$ are rejected). So a judgment set $A \in \mathcal{J}$ can be identified with a function b assigning to each variable $V \in \mathbf{V}$ a value $v \in Rge(V)$. Then $\mathcal{J}$ becomes a set B of such functions, and an aggregation rule $F : \mathcal{J}^n \to \mathcal{J}$ becomes a function $f : B^n \to B$.[17]

Example 4: relevance as premisehood, and generalised premise-based rules. If we interpret "$r\mathcal{R}p$" as "r is a premise/reason/argument for (or against) p", III is the condition that the aggregation rule be *premise-based*: that the collective judgment on any proposition $p \in X$ be determined by people's reasons for their judgments on p.

In principle, $\mathcal{R}$ could define an arbitrarily complex premisehood structure over a possibly complex agenda, generalising the classical *premise-based procedure* (PBP) usually defined for simple agendas like agendas 1 and 2 in Figure 1. For agenda 1,

[17] Rubinstein and Fishburn (1986) analyse variable-wise independent aggregators $f : B^n \to B$, assuming that there are only finitely many variables V, all with the same range, an algebraic field $\mathcal{F}$ (e.g. $\mathcal{F} = \mathbf{R}$). So $B \subseteq \mathcal{F}^{\mathbf{V}}$. Their two main results establish correspondences between algebraic properties of B, like being a *hyperplane* of the $\mathcal{F}$-vector space $\mathcal{F}^{\mathbf{V}}$, and algebraic properties of "admissible" aggregators $f : B^n \to B$, like *linearity* or *additivity*. In practice, the hyperplane condition on B seems restrictive: variables are interconnected by exactly one equation, and this equation is linear. But note that linearity can sometimes be achieved by transforming the variables.

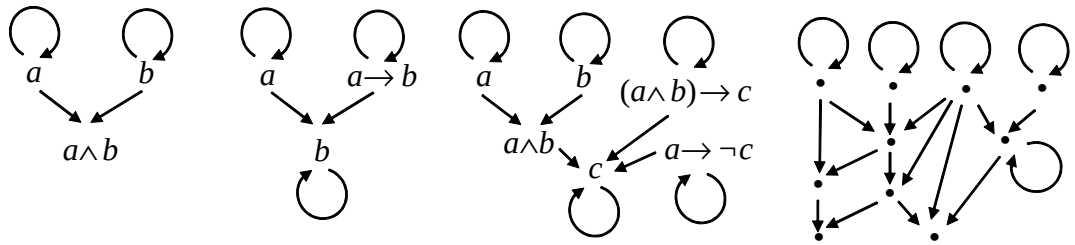

Figure 1: Relevance ("premisehood") relations over four agendas. Arrows indicate relevance. Agenda 1: $X^+ = \{a, b, a \wedge b\}$. Agenda 2: $X^+ = \{a, a \to b, b\}$. Agenda 3: $X^+ = \{a, b, c, a \wedge b, (a \wedge b) \to c, a \to \neg c\}$. Agenda 4: X^+ contains ten propositions indicated by "$\cdot$".

the classical PBP decides each "premise" a and b by a majority vote, and decides the "conclusion" $a \wedge b$ by logical entailment from the decisions on a and b. This PBP is III for the relevance ("premisehood") relation indicated in Figure 1:

$$\mathcal{R}^+(a) = \{a\},\ \mathcal{R}^+(b) = \{b\},\ \mathcal{R}^+(a \wedge b) = \{a, b\}. \tag{3}$$

For agenda 2 in Figure 1, the classical PBP takes majority votes on each "premise" a and $a \to b$; if the resulting decisions logically constrain the "conclusion" b,[18] b is decided accordingly; otherwise b is (for instance) decided by a majority vote on b. This PBP is III for relevance as given in Figure 1. Unlike in (3), the conclusion is self-relevant: individual judgments on b may matter for deciding b.

In general, call $p \in X$ a *root* proposition if p has no premise other than p (and $\neg p$). In (3), a and b are root propositions. Any root proposition $p \in X$ must be a premise to itself: otherwise p would have no premises at all, violating "no underdetermination".[19] So the collective judgment on any root proposition p is (by III) formed solely on the basis of people's judgments on p via some voting method – majority voting if we stick closely to the standard premise-based procedure – while decisions on non-root propositions may depend on external premises.

When interpreting $\mathcal{R}$ as a premisehood relation, additional requirements on $\mathcal{R}$ may be appropriate. Surely, symmetry should not be required (unlike in Examples 1-3). Indeed, one might require that $\mathcal{R}$ is *anti-symmetric* on X^+ (so that no distinct propositions in X^+ are premises to each other) or, more strongly, *acyclic* on X^+ (so that in X^+ there is no cycle $p_1 \mathcal{R} p_2 \mathcal{R} p_3 ... \mathcal{R} p_m \mathcal{R} p_1$ where the p_i's are pairwise distinct and $m \geq 2$).

For some agendas X, specifying $\mathcal{R}$ is non-trivial: it is not obvious which propositions should count as reasons for/against which others. One might for instance draw on the *syntax* of the propositions in X: if $X^+ = \{a, a \to b, b\}$, one might argue that $a \mathcal{R} b$ because $a \to b \in X$, and not $b \mathcal{R} a$ because $b \to a \notin X$. Finding objective criteria for relevance would be an interesting research goal on its own.[20]

[18] Whether this is so may depend on whether "$\to$" is a material or subjunctive implication.

[19] Unless p is a tautology or contradiction: then even the empty set settles p.

[20] For different purposes, *relevance logicians* (e.g. Parikh 1999) propose syntactic and other criteria for when a proposition is relevant to another. Although this enterprise is controversial and its notion of relevance may differ from ours, one might use such criteria in defining $\mathcal{R}$.

4 Justifying the "no underdetermination" condition

I now give a technical and a conceptual motivation for the "no underdetermination" requirement on a relevance relation.

The technical reason is that "no underdetermination" is crucial for the existence of non-degenerate III aggregation rules. The condition of *judgment-set unanimity preservation*, whereby $F(A, ..., A) = A$ for all unanimous profiles $(A, ..., A)$ in the domain, is very mild (unlike the unrestricted *proposition*wise unanimity condition mentioned in Section 7).

Theorem 1 *Let $\mathcal{R}$ be an arbitrary binary relation on X. There exists a judgment-set unanimity preserving III aggregation rule with universal domain if and only if $\mathcal{R}$ satisfies "no underdetermination".*

This defence of the "no underdetermination" condition needs no collective completeness or consistency condition, not even a non-dictatorship condition. So, given underdetermination, not even rules with incomplete, or inconsistent, or dictatorial outputs can satisfy the conditions.

Proof. First, suppose "no underdetermination" is violated for $p \in X$. Then there are sets $A, A' \in \mathcal{J}$ such that $p \in A$ and $p \notin A'$ but $A \cap \mathcal{R}(p) = A' \cap \mathcal{R}(p)$ (i.e. A and A' disagree on p but agree on the relevant propositions). If we apply an III aggregation rule F with universal domain to the two unanimous profiles $(A, ..., A)$ and $(A', ..., A')$, the resulting judgment sets $F(A, ..., A)$ and $F(A', ..., A')$ agree on p by III. So, as A and A' disagree on p, $F(A, ..., A) \neq A$ or $F(A', ..., A') \neq A'$, violating judgment-set unanimity preservation.

Second, suppose $\mathcal{R}$ satisfies "no underdetermination". I show that (for instance) the unanimity rule with universal domain, given by $F(A_1, ..., A_n) = A_1 \cap ... \cap A_n$, satisfies III (it obviously also preserves judgment-set unanimity). Consider any $p \in X$ and $(A_1, ..., A_n), (A'_1, ..., A'_n) \in \mathcal{J}^n$ such that $A_i \cap \mathcal{R}(p) = A'_i \cap \mathcal{R}(p)$ for all i. For all i, we have $A_i \cap \{r, \neg r : r \in \mathcal{R}(p)\} = A'_i \cap \{r, \neg r : r \in \mathcal{R}(p)\}$. By "no underdetermination", this set entails p or entails $\neg p$; in the first case, $p \in A_i$ and $p \in A'_i$, and in the second case $p \notin A_i$ and $p \notin A'_i$. So in any case $p \in A_i \Leftrightarrow p \in A'_i$. Hence $p \in \cap_i A_i \Leftrightarrow p \in \cap_i A'_i$, i.e. $p \in F(A_1, ..., A_n) \Leftrightarrow F(A'_1, ..., A'_n)$, as desired. ■

I now turn to a conceptual defence of "no underdetermination". This condition can be violated for $p \in X$ only if p is self-irrelevant, a rather special assumption. I can see only one (albeit prominent) case in which self-irrelevance has a clear motivation: the case of premise-based collective decision making. Here the decision on $p \in X$ should depend on people's *reasons* (*grounds*) for accepting or rejecting p (as in Examples 4). A person's reasons for accepting (rejecting) p can be viewed as a set E of sentences that, if specified exhaustively, logically entails p ($\neg p$).[21] But not

[21] If the set of reasons E did *not* entail p, it wouldn't be exhaustive, i.e. some "reasons" have been forgotten. For instance, if in inferring p from E the persons implicitly uses that $(\wedge_{e \in E} e) \rightarrow p$, i.e. that p follows from the members of E, then $(\wedge_{e \in E} e) \rightarrow p$ should be added as a reason to E. The so-enlarged set of reasons now logically entails p. (In other situations, $(\wedge_{e \in E} e) \rightarrow p$ is not the missing reason, i.e. E has to be enlarged differently.)

all sets of sentences E that entail p ($\neg p$) have to count as "sets of reasons" for accepting (rejecting) p; $\{p\}$ might not count as a set of reasons for accepting p. For instance, $\{a\}$ and $\{b\}$ might count as the sets of reasons for accepting a disjunction $a \vee b$; and $\{\neg a, \neg b\}$ might count as the only set of reasons for rejecting $a \vee b$. Let $\mathcal{E}(p)$ be the set of all sets of reasons for accepting or rejecting p. In the example, $\mathcal{E}(a \vee b) = \{\{a\}, \{b\}, \{\neg a, \neg b\}\}$. Plausibly, $\mathcal{E}(p)$ should contain sufficiently many sets of reasons so that p cannot be accepted or rejected without any set of reasons $E \in \mathcal{E}$. For instance, we cannot remove $\{a\}$ from $\mathcal{E}(a \vee b)$: otherwise someone who accepts a but not b would accept $a \vee b$ for *no* set of reasons $E \in \mathcal{E}(a \vee b)$. Given our assumption of reason-based aggregation, every reason for or against p (i.e. every member of any set of reasons $E \in \mathcal{E}(p)$) should be considered relevant to p: that is, $\cup_{E \in \mathcal{E}(p)} E \subseteq \mathcal{R}(p)$ (one might even argue that $\cup_{E \in \mathcal{E}(p)} E = \mathcal{R}(p)$). In this case, "no underdetermination" holds.[22]

5 Possibility or impossibility?

Are there appealing III aggregation rules, and how do they look? General answers to this question are harder to give than for proposition-wise independence. The reason is that criteria for the (in)existence of (non-degenerate) III aggregation rules typically concern not just logical interconnections (as for proposition-wise independence) but also relevance interconnections. More precisely, we need criteria on the *interplay* between logic and relevance. One such criterion is "no underdetermination", which is (by Theorem 1) necessary and sufficient for a limited possibility: "limited" because collective incompleteness is allowed (but collective consistency, agreement preservation, and non-dictatorship could have been required in Theorem 1, as the proof shows).

Below I derive one more possibility theorem – with creteria for the possibility of priory rules – and four impossibility theorems. I deliberately sacrifice some generality (of the criteria) for simplicity and elegance.[23]

6 Priority rules

In this section, I adopt Example 4's interpretation of relevance as premisehood; and I assume again that $\mathcal{R}$ is negation-invariant. Do there exist appealing premise-based

[22] What if some reasons $e \in \cup_{E \in \mathcal{E}(p)} E$ are outside the agenda X (i.e. not part of the decision problem), so that we cannot have $\cup_{E \in \mathcal{E}(p)} E \subseteq \mathcal{R}(p)$? One might *either* argue that such agendas are simply misspecified (in the context of reason-based aggregation): if $a \vee b \in X$ then X should include $a \vee b$'s reasons. *Or* one might defend "no underdetermination" for such agendas: since "$r \in X$ is relevant to $p \in X$" more precisely means "the individuals' judgments on r are relevant information for deciding p", if X excludes some of p's reasons then other propositions in X (perhaps p itself) become relevant to p as people's judgments on them are information on people's non-available reasons. If $X = \{a \vee b, \neg(a \vee b)\}$, the individuals' judgments on $a \vee b$ are relevant information for deciding $a \vee b$ as they reflect (partially) people's non-available reasons; hence we have $(a \vee b)\mathcal{R}(a \vee b)$ (but not so if X contains all reasons of $a \vee b$).

[23] I make no conjecture on the nature of *minimal* criteria for the (im)possibilities considered below, except that such conditions would not have a unified or structured form but the form of disjunctions of several cases. The reason is that the conditions must capture the joint and non-separable behaviour of relevance and logical connections, which is left general and uncontrolled in the framework.

(i.e. III) aggregation rules? I now introduce *priority rules* (generalising List 2004) and give simple criteria for when they can be used.

An impossibility threat comes not only from logical interconnections between root propositions (or other propositions), but also from transitivity violations of relevance $\mathcal{R}$. To see why, let $p \in X$ and suppose the premises of p's premises – call them the "pre-premises" – are *not* premises of p. The decision on p is settled by the decisions on p's premises (by "no underdetermination"), which in turn depend on how people judge the pre-premises (by III). This forces the decision on p to be some function f of how people judge the pre-premises. But by III the decision on p must be a function of how people judge p's premises (not pre-premises). So f depends on people's pre-premise judgments only indirectly: only through people's premise judgments as entailed by their pre-premise judgments – a strong restriction on f that suggests that impossibility is looming.

It is debatable whether premisehood (more generally, relevance) is inherently a transitive concept. If $\mathcal{R}$ is assumed transitive – whether for conceptual reasons or just to remove one impossibility source – interesting candidates for III aggregation arise, as explained now. List (2004) introduces *sequential priority rules* in judgment aggregation (generalising sequential rules in standard social choice theory). Here the propositions of a (finite) agenda are put in a priority order $p_1, p_2, ...$ and decided sequentially, where earlier decisions logically constrain later ones. As is easily seen, such a rule is III if relevance is given by $p_j \mathcal{R} p_{j'} \Leftrightarrow j \leq j'$, a linear order on X^+. I now introduce similar rules relative to an arbitrary (possibly quite partial) relevance relation. Informally, these rules decide the propositions in the order of relevance: each $p \in X$ is decided by logical entailment from previously accepted relevant propositions *except* if the latter propositions do not settle p, in which case p is decided via some local decision method (e.g. via majority voting on p). Formally, a *priority rule* is an aggregation rule F with universal domain such that there is for every proposition $p \in X^+$ a ("local") aggregation rule D_p for the binary agenda $\{p, \neg p\}$ (where D_p has for this agenda universal domain and consistent and complete outcomes) with

$$F(A_1, ..., A_n) \cap \{p, \neg p\} = \begin{cases} \{\tilde{p} \in \{p, \neg p\} : F(A_1, ..., A_n) \cap & \text{if this set is} \\ \qquad\qquad \mathcal{R}(p) \backslash \{p, \neg p\} \vdash \tilde{p}\} & \text{non-empty} \\ D_p(A_1 \cap \{p, \neg p\}, ..., A_n \cap \{p, \neg p\}) & \text{otherwise} \end{cases} \quad (4)$$

for all profiles $(A_1, \ldots, A_n) \in \mathcal{J}^n$. So the pair $p, \neg p$ is decided locally via D_p unless the previous decisions $F(A_1, ..., A_n) \cap \mathcal{R}(p) \backslash \{p, \neg p\}$ are logically constraining (hence "priority" to the previous decisions). In practice, first every root proposition $p \in X^+$ and $\neg p$ are decided by a local vote using D_p. Then every non-root proposition $p \in X^+$ to which only root propositions (and possibly p and $\neg p$) are relevant is decided: *either* by entailment from the previous decisions on relevant root propositions *or* (if neither p nor $\neg p$ is entailed) by a local vote using D_p. And so on.

The local rule D_p may be chosen as the same rule for all $p \in X^+$ (e.g. majority rule). Or D_p may vary: D_p might assign more weight to individuals with expertise on p (e.g. to physicists if p is "Nuclear energy is safe"), or to individuals personally affected by the decision on p (e.g. to the citizens of towns X and Y if p is "A road between X and Y should be built"). Such "expert rights" or "liberal rights" are (unlike those in Dietrich and List 2004) *conditional rights*: they can be overruled by previous decisions on relevant propositions. If the group can be partitioned into

experts on different fields, and a proposition q's premises fall each into exactly one of the fields, the decision on each premise $p \in X^+$ of q could be delegated entirely to the experts on p, i.e. D_p uses only these experts' judgments. This generalises List's (2005) *distributed premise-based procedure*. The premises from different fields might form different subtrees preceding q.

Once we specify the family $(D_p)_{p \in X^+}$ of local rules, the recursive formula (4) defines a unique priority rule $F_{(D_p)_{p \in X^+}} := F$, provided that relevance $\mathcal{R}$ is *well-founded* on X^+.[24]

The following theorem shows that, for transitive relevance, $F_{(D_p)_{p \in X^+}}$ (a) satisfies III, and (b) generates consistent outcomes if certain logical independencies hold within X. Result (a) is surprising in one respect: one might have expected that III can be violated for non-self-relevant $p \in X$ due to the second case in (4). Let me motivate result (b). $F_{(D_p)_{p \in X^+}}$ could generate inconsistent outcomes if there are logical dependencies between root propositions, or more generally between any propositions $p_i \in X, i \in I$, that are *mutually irrelevant* (i.e. for no distinct $i, i' \in I$ $p_i \mathcal{R} p_{i'}$). To see why, notice that no p_i's precede other p_i's in the priority order (by irrelevance), whence the decisions on the p_i's ignore each other. But even if the (mutually irrelevant) p_i's are logical independent, inconsistent outcomes may still arise if there are logical interconnections between the sets $\mathcal{R}(p_i), i \in I$, as is easily imagined. This is why result (b) requires certain logical independencies between the sets $\mathcal{R}(p_i), i \in I$. To define these logical independencies, some terminology is needed. As usual, negation-closed sets $A_i, i \in I$, are called *logically independent* if $\cup_{i \in I} B_i$ is consistent for all consistent sets $B_i \subseteq A_i, i \in I$. Logical independence fails whenever $A_i \cap A_{i'} \neq \emptyset$ for some $i \neq i'$, because the sets B_i and $B_{i'}$ can pick different members of a pair $p, \neg p \in A_i \cap A_{i'}$.[25] This "easy" way to render $\cup_{i \in I} B_i$ inconsistent is excluded in the following weaker definition. I call negation-closed sets $A_i, i \in I$, *logically quasi-independent* if $\cup_{i \in I} B_i$ is consistent for all consistent sets $B_i \subseteq A_i, i \in I$, such that any pair $p, \neg p$ in an intersection $A_i \cap A_{i'}$ $(i \neq i')$ has a member that is both in B_i and in $B_{i'}$. (So $A_i \cap A_{i'}$ has the same intersection with B_i as with $B_{i'}$).

The theorem moreover requires relevance $\mathcal{R}$ to be *vertically finite*: there is no infinite sequence $(p_k)_{k=1,2,...}$ in X^+ that is ascending (i.e. each p_k is relevant to and distinct from p_{k+1}) or descending (i.e. each p_{k+1} is relevant to and distinct from p_k). In short, the network of inter-relevances is nowhere "infinitely deep", but possibly "infinitely broad". This exclusion of "infinite relevance chains" is a debatable condition on the concept of relevance;[26] without it the theorem would not hold.

[24] $\mathcal{R}$ is *well-founded* on X^+ if every non-empty set $S \subseteq X^+$ has an $\mathcal{R}$-minimal element s (i.e. for no $r \in S \backslash \{s\}$ $r \mathcal{R} s$); or, more intuitively, if there is no infinite sequence $(p_k)_{k=1,2,...}$ in X^+ such that each p_{k+1} is relevant to and distinct from p_k. The priority rule $F \equiv F_{(D_p)_{p \in X^+}}$ uniquely exists because, for every $(A_1, ..., A_n) \in \mathcal{J}^n$, $F(A_1, ..., A_n)$ is the union of the sets $f(p) := F(A_1, ..., A_n) \cap \{p, \neg p\}$, $p \in X^+$, where the function f is uniquely defined on X^+ by recursion on $\mathcal{R}$ using the well-founded recursion theorem (e.g. Fenstad 1980). If $\mathcal{R}$ is *not* well-founded on X^+, there could exist no or many priority rules with local rules $(D_p)_{p \in X^+}$.

[25] Unless p is a tautology or contradiction.

[26] For instance, one might argue (like Gärdenfors 2006) that every proposition can, in principle, be explained in terms of more fundamental premises; this creates infinite descending relevance chains. On the other hand, realistic agendas might still be vertically finite: they might not *include* all arbitrarily fundamental premises.

Theorem 2 *Let relevance $\mathcal{R}$ be transitive, vertically finite and negation-invariant.*

(a) *Every priority rule satisfies III.*[27]

(b) *Every priority rule generates consistent judgment sets if, for all mutually irrelevant propositions $p_i \in X, i \in I$, the sets $\mathcal{R}(p_i), i \in I$, are logically quasi-independent.*

The logical quasi-independence condition reduces to a logical independence condition if no mutually irrelevant propositions share any relevant proposition – but often the relevance (premisehood) relation is not of this special kind. Consider for instance case 4 in Figure 1.[28] Or consider a scientific board using a priority rule to derive collective judgments on several scientific propositions: then mutually irrelevant propositions, e.g. "Species X survives in Hawaii" and "Species Y survives in Australia", might well share premises, e.g. "The ozone hole exceeds size Z" or general biological or chemical hypotheses.

Proof. Let $\mathcal{R}$ and X be as specified. I leave it to the author to verify that $\mathcal{R}$'s vertical finiteness implies (in fact, is equivalent to) the following: every non-empty set $S \subseteq X^+$ has an $\mathcal{R}$-maximal element s (i.e. for no $r \in S \backslash \{s\}$ $s\mathcal{R}r$) and an $\mathcal{R}$-minimal element s (i.e. for no $r \in S \backslash \{s\}$ $r\mathcal{R}s$). In short:

$$\max_{\mathcal{R}} S \neq \emptyset \text{ and } \min_{\mathcal{R}} S \neq \emptyset, \text{ for all } \emptyset \neq S \subseteq X^+. \tag{5}$$

In particular, $\mathcal{R}$ is well-founded on X^+. Let $F \equiv F_{(D_p)_{p \in X^+}}$ be a priority rule.

(a) To show III, I prove that all $p \in X^+$ have the following property: for all $(A_1, ..., A_n), (A'_1, ..., A'_n) \in \mathcal{J}^n$, if $A_i \cap \mathcal{R}(p) = A'_i \cap \mathcal{R}(p)$ for all i then

$$F(A_1, ..., A_n) \cap \{p, \neg p\} = F(A'_1, ..., A'_n) \cap \{p, \neg p\}. \tag{6}$$

Suppose for a contradiction that the property fails for some $p \in X^+$. By (5) there is a $p \in X^+$ that is $\mathcal{R}$-minimal such that the property fails. So there are $(A_1, ..., A_n), (A'_1, ..., A'_n) \in \mathcal{J}^n$ with $A_i \cap \mathcal{R}(p) = A'_i \cap \mathcal{R}(p)$ for all i such that (6) is false. By p's minimality property and $\mathcal{R}$'s transitivity,

$$F(A_1, ..., A_n) \cap \mathcal{R}(p) \backslash \{p, \neg p\} = F(A'_1, ..., A'_n) \cap \mathcal{R}(p) \backslash \{p, \neg p\}. \tag{7}$$

Let $Y := \{\tilde{p} \in \{p, \neg p\} : \text{the set (7) entails } \tilde{p}\}$.

Case 1: $Y \neq \emptyset$. Then, by the first case in (4), $F(A_1, ..., A_n) \cap \{p, \neg p\} = Y$, and for the same reason $F(A'_1, ..., A'_n) \cap \{p, \neg p\} = Y$. This implies (6), contradicting the choice of p.

Case 2: $Y = \emptyset$. Then, by the second case in (4), $F(A_1, ..., A_n) \cap \{p, \neg p\} = D_p(A_1 \cap \{p, \neg p\}, ..., A_n \cap \{p, \neg p\})$ and $F(A'_1, ..., A'_n) \cap \{p, \neg p\} = D_p(A'_1 \cap \{p, \neg p\}, ..., A'_n \cap \{p, \neg p\})$. These two sets are distinct (as (6) is violated), and so for some i $A_i \cap \{p, \neg p\} \neq A'_i \cap \{p, \neg p\}$. So, as $A_i \cap \mathcal{R}(p) = A'_i \cap \mathcal{R}(p)$, $\mathcal{R}(p)$ does not contain both of $p, \neg p$, hence contains none of $p, \neg p$ by negation-invariance. So the set (7) equals $F(A_1, ..., A_n) \cap \mathcal{R}(p)$, which contains a member of each pair $r, \neg r \in \mathcal{R}(p)$, and hence entails p or $\neg p$ by "no underdetermination". This contradicts that $Y = \emptyset$.

[27] This still holds if the vertical finiteness condition is weakened to well-foundedness on X^+.

[28] If $\mathcal{R}$ is wished to be transitive, close the plotted inter-relevances under transitivity.

(b) Assume the condition. For all $p \in X$, put $\mathcal{R}^p := \mathcal{R}(p) \cup \{p, \neg p\}$ and $\mathcal{R}_p := \mathcal{R}(p) \backslash \{p, \neg p\}$. Let $(A_1, ..., A_n) \in \mathcal{J}^n$. The (desired) consistency of $A := F(A_1, ..., A_n)$ follows from the following claims.

Claim 1: $X = \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p$; in particular, $A = \cup_{p \in \max_{\mathcal{R}} X^+} (A \cap \mathcal{R}^p)$.

Claim 2: for any mutually irrelevant propositions $p_i \in X, i \in I$, the sets $\mathcal{R}^{p_i}, i \in I$, are logically quasi-independent; in particular, the sets $\mathcal{R}^p, p \in \max_{\mathcal{R}} X^+$, are logically quasi-independent.

Claim 3: $A \cap \mathcal{R}^p$ is consistent for all $p \in X^+$ (hence for all $p \in \max_{\mathcal{R}} X^+$).

Proof of Claim 1. For a contradiction, suppose $X \backslash \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p \neq \emptyset$. Then, by negation-invariance, $X^+ \backslash \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p \neq \emptyset$. Hence by (5) there is a $q \in \max_{\mathcal{R}} \left(X^+ \backslash \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p\right)$. By $q \notin \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p$, $q \notin \max_{\mathcal{R}} X^+$. So q is relevant to some $q' \in X^+ \backslash \{q\}$. We have $q' \notin X^+ \backslash \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p$, as q is maximal in $X^+ \backslash \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p$. So $q' \in \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p$. Hence, as $\mathcal{R}$ is transitive, q is relevant to some $p \in \max_{\mathcal{R}} X^+$, a contradiction as $q \notin \cup_{p \in \max_{\mathcal{R}} X^+} \mathcal{R}^p$.

Proof of Claim 2. Consider mutually irrelevant $p_i \in X, i \in I$, and consistent sets $B_i \subseteq \mathcal{R}^{p_i}, i \in I$, such that any pair $p, \neg p$ in an intersection $\mathcal{R}^{p_i} \cap \mathcal{R}^{p_{i'}}$ $(i \neq i')$ has a member that is in B_i and in $B_{i'}$. I show that $\cup_{i \in I} B_i$ is consistent. W.l.o.g. let each B_i contain a member of each pair $p, \neg p \in \mathcal{R}^{p_i}$ (otherwise extend the B_i's to consistent sets $\bar{B}_i \subseteq \mathcal{R}^{p_i}$ with the property; the present proof shows the consistency of $\cup_{i \in I} \bar{B}_i$, hence of $\cup_{i \in I} B_i$). As the sets $\mathcal{R}(p_i), i \in I$, are logically quasi-independent, (*) $\cup_{i \in I} (B_i \cap \mathcal{R}(p_i))$ is consistent. By "no underdetermination", (**) each $B_i \cap \mathcal{R}(p_i)$ entails a $\tilde{p}_i \in \{p_i, \neg p_i\}$. Each B_i is either $(B_i \cap \mathcal{R}(p_i)) \cup \{\tilde{p}_i\}$ or $(B_i \cap \mathcal{R}(\tilde{p}_i)) \cup \{\neg \tilde{p}_i\}$; so, as the latter set is inconsistent by (**) whereas B_i is consistent, $B_i = (B_i \cap \mathcal{R}(p_i)) \cup \{\tilde{p}_i\}$. Hence $\cup_{i \in I} B_i = \cup_{i \in I} ((B_i \cap \mathcal{R}(p_i)) \cup \{\tilde{p}_i\})$, which is consistent by (*) and (**).

Proof of Claim 3. Suppose the contrary: there is a $p \in X^+$ for which $A \cap \mathcal{R}^p$ is inconsistent. By (5), there is a $p \in X^+$ that is $\mathcal{R}$-maximal subject to $A \cap \mathcal{R}^p$ being inconsistent. By an argument similar to that for Claim 1,

$$\mathcal{R}_p = \cup_{q \in \max_{\mathcal{R}}(X^+ \cap \mathcal{R}_p)} \mathcal{R}^q; \text{ hence } A \cap \mathcal{R}_p = \cup_{q \in \max_{\mathcal{R}}(X^+ \cap \mathcal{R}_p)} (A \cap \mathcal{R}^q). \quad (8)$$

By Claim 2, the sets $\mathcal{R}^q, q \in \max_{\mathcal{R}}(X^+ \cap \mathcal{R}_p)$, are logically quasi-independent. Hence, as each $A \cap \mathcal{R}^q$ in (8) is consistent (by the maximality of p), the set $A \cap \mathcal{R}_p$ is consistent. So, as $A \cap \{p, \neg p\}$ is related via (4) to $A \cap \mathcal{R}_p$ $(= A \cap \mathcal{R}(p) \backslash \{p, \neg p\})$, the union $(A \cap \mathcal{R}_p) \cup (A \cap \{p, \neg p\}) = A \cap \mathcal{R}^p$ is also consistent. This contradicts the choice of p. ■

7 A defensible (restricted) unanimity condition

It is natural to require (as in later theorems) a unanimity preservation condition. But it would be against the relevance-based approach to require *global* unanimity preservation, i.e. to require for *all* $p \in X$ that a unanimity for p implies social acceptance of p. Indeed, a unanimity for p can be *spurious*: different individuals i can hold p for different reasons, that is (in the relevance terminology) they may hold different explanations $E_i \subseteq \{r, \neg r : r \mathcal{R} p\}$ of p.[29] I will not require spurious unanimities to be respected. This follows the frequent view that spurious unanimities have less normative force. It also follows our relevance-based approach, since propositions relevant

[29] On spurious unanimities, see for instance Mongin (2005-b) and Bradley (forthcoming).

to p should not suddenly be treated as irrelevant if a unanimity accepts p. Instead, I will impose a unanimity condition restricted to a fixed set $\mathcal{P} \subseteq X$ of "privileged" propositions:

Agreement Preservation. For every profile $(A_1, ..., A_n)$ in the domain and every privileged proposition $p \in \mathcal{P}$, if $p \in A_i$ for all individuals i then $p \in F(A_1, ..., A_n)$.

I assume that $\mathcal{P}$ is chosen such that a unanimity for a $p \in \mathcal{P}$ cannot be spurious, i.e. such that each $p \in \mathcal{P}$ can be explained in just one way:[30]

$$\mathcal{P} \subseteq \{p \in X : p \text{ has a single } \mathcal{R}\text{-explanation}\}. \tag{9}$$

By default (i.e. if $\mathcal{P}$ is not explicitly defined otherwise), I assume that (9) holds with a "=". This maximal choice of $\mathcal{P}$ is often natural, though not necessary for the theorems below. Another potentially natural choice is to include in $\mathcal{P}$ only propositions $p \in X$ to which just p itself is relevant; so $\{p\}$ is p's only explanation, i.e. p has no "external" explanation.[31]

In the "classical" case that each $p \in X$ is just self-relevant, $\mathcal{P}$ by default equals X,[32] i.e. agreement preservation applies globally. So the "classical" relevance notion renders not only III equivalent to standard independence but also agreement preservation equivalent to standard (proposition-wise) unanimity preservation. If $X^+ = \{a, b, a \wedge b\}$, with (negation-invariant) relevance given by (3), $\mathcal{P}$ may contain $a \wedge b$ (which has a single explanation: $\{a, b\}$) but not $\neg(a \wedge b)$ (which has three explanations: $\{\neg a, b\}, \{a, \neg b\}, \{\neg a, \neg b\}$). So a unanimity for $\neg(a \wedge b)$ can be spurious and need not be respected.

Example 1 (continued) For the preference agenda, agreement preservation is equivalent to the weak Pareto principle, in virtue of defining $\mathcal{P}$ as

$$\mathcal{P} := \{\neg xRy : x, y \in Q, x \neq y\} = \{yPx : x, y \in Q, x \neq y\}, \tag{10}$$

the set of *strict* ranking propositions yPx. I call (10) the *Arrowian* set of privileged propositions. Note that (under the Arrowian relevance relation) each $\neg xRy = yPx \in \mathcal{P}$ has indeed a single explanation: $\{\neg xRy, yRx\}$.

8 Semi-vetodictatorship and semi-dictatorship

Hereafter, we consider an III and agreement preserving aggregation rule $F : \mathcal{J}^n \to \mathcal{J}$, relative to some fixed relevance relation $\mathcal{R}$ and some fixed set of privileged propositions $\mathcal{P}$. I give conditions (on logical and relevance links) that force F to be degenerate: a (semi-)dictatorship or (semi-)vetodictatorship.

First, how should these degenerate rules be defined? The relevance-based framework allows us to generalise the standard social-choice-theoretic definitions. Recall

[30] "Agreement" means "non-spurious unanimity". Hence the term "agreement preservation".

[31] Interesting normative questions can be raised about the choice of $\mathcal{P}$. For instance, Nehring (2005) suggests in his analysis of the Pareto/unanimity condition that unanimities are normatively binding if they reflect "self-interested" judgments, or if they carry "epistemic priority".

[32] Unless X contains contradictions: these have *no* explanation, hence are not in $\mathcal{P}$.

that an (Arrowian) "dictator" is an individual who can socially enforce his *strict* preferences between options, but not necessarily his *indifferences.* Similarly, a "vetodictator" can prevent ("veto") any *strict* preference, but not necessarily any indifference. Put in our terminology, a dictator (vetodictator) can enforce (veto) any *privileged* proposition of the preference agenda (given (10)). The following definitions generalise this to arbitrary agendas.

Definition 1 *An individual i is*

(a) *a* dictator *(respectively,* semi-dictator*) if, for every privileged proposition $p \in \mathcal{P}$, we have $p \in F(A_1, ..., A_n)$ for all $(A_1, ..., A_n) \in \mathcal{J}^n$ such that $p \in A_i$ (respectively, such that $p \in A_i$ and $p \notin A_j$, $j \neq i$);*

(b) *a* vetodictator *(respectively,* semi-vetodictator*) if, for every privileged proposition $p \in \mathcal{P}$, i has a* veto *(respectively,* semi-veto*) on p, i.e. a judgment set $A_i \in \mathcal{J}$ not containing p such that $p \notin F(A_1, ..., A_n)$ for all $A_j \in \mathcal{J}$, $j \neq i$ (respectively, for all $A_j \in \mathcal{J}$, $j \neq i$, containing p).*

In the standard models without a relevance relation, *conditional entailment* between propositions (first used by Nehring and Puppe 2002/2005) has proven useful to understand agendas. Roughly, $p \in X$ conditionally entails $q \in X$ if p together with other propositions in X entails q (with a non-triviality condition on the choice of "other" propositions). I cannot use conditional entailments here, as they reflect only *logical* links between propositions. Rather, I now define *constrained entailments*, a related notion that reflects both logical and relevance links. It will turn out that certain paths of constrained entailments lead to degenerate aggregation rules.[33]

Definition 2 *For propositions $p, q \in X$, if $\{p\} \cup Y \vdash q$ for a set $Y \subseteq \mathcal{P}$ consistent with every explanation of p and with every explanation of $\neg q$, I say that p* constrained entails q (in virtue of Y)*, and I write $p \vdash_* q$ or $p \vdash_Y q$.*[34]

As this definition is symmetric in p and $\neg q$, constrained entailment satisfies contraposition:

Lemma 1 *For all $p, q \in X$ and all $Y \subseteq \mathcal{P}$, $p \vdash_Y q$ if and only if $\neg q \vdash_Y \neg p$.*

The amount of constrained entailments in X is crucial for whether impossibilities arise. Trivially, every *unconditional* entailment is also a constrained entailment (namely in virtue of $Y = \emptyset$). Intuitively, if there are more inter-relevances between propositions, $\mathcal{P}$ becomes smaller, and propositions have more and larger explanations;

[33] Nehring and Puppe (2002/2005) use paths of *conditional* entailment to define their *totally blocked* agendas. For such agendas, they obtain strong dictatorship by imposing that F satisfies proposition-wise independence, an unrestricted unanimity condition, and a monotonicity condition. I impose relevance-based conditions on F (III and agreement preservation); these conditions, like Arrow's conditions, imply less than strong dictatorship.

[34] Many alternative notions of constrained entailment turn out to be non-suitable: they do not preserve interesting properties along paths of constrained entailments. The present definition is the weakest one to preserve semi-winning coalitions. The requirement that $Y \subseteq \mathcal{P}$ allows one to apply agreement preservation. In view of different results to those derived here, it might be fruitful to impose additional requirements on Y, e.g. that Y be consistent also with explanations of $\neg p$ and/or of q.

so the requirements on Y in constrained entailments become stronger; hence there are fewer constrained entailments, and more room for possibilities of aggregation.

The preference agenda X (Example 1, with Arrowian $\mathcal{R}$ and $\mathcal{P}$) displays many constrained entailments (hence impossibilities). For instance, $xRy \vdash_{\{yPz\}} xPz$ (if x, y, z are pairwise distinct options), as yPz is in $\mathcal{P}$ and is consistent with each explanation of xRy ($\{xRy, yRx\}$ and $\{xRy, \neg yRx\}$) and the only explanation of $\neg xPz = zRx$. By contrast, *no* non-trivial constrained entailments arise in our example $X^+ = \{a, b, a \wedge b\}$ with (negation-invariant) relevance given by (3): for instance, it is not the case that $a \vdash_{\{\neg(a \wedge b)\}} \neg b$ since $\neg(a \wedge b) \notin \mathcal{P}$; and it is not the case that $a \vdash_{\{b\}} a \wedge b$, as $\{b\}$ is inconsistent with the explanation $\{a, \neg b\}$ of $\neg(a \wedge b)$. As a result, our impossibilities will not apply to this agenda – and cannot, as the premise-based procedure for odd n (see Example 4) satisfies all conditions.

To obtain impossibility results, richness in constrained entailments is not sufficient. At least one constrained entailment $p \vdash_* q$ must hold in a "truly" constrained sense. By this I mean more than that p does not *un*conditionally entail q, i.e. more than that p is consistent with $\neg q$: I mean that every explanation of p is consistent with every explanation of $\neg q$.

Definition 3 *For propositions $p, q \in X$, p* truly constrained entails *q if $p \vdash_* q$ and moreover every explanation of p is consistent with every explanation of $\neg q$.*

For instance, if relevance is an equivalence relation (as in Examples 1-3) that partitions X into *pairwise logically independent* subagendas[35] (as for the preference agenda) then all constrained entailments across equivalence classes are truly constrained. Also, $p \vdash_* q$ is truly constrained if $p \nvdash q$ and moreover p and q are root propositions (see Example 4).

Our impossibility results rest on the following path conditions.

Definition 4 (a) *For propositions $p, q \in X$, if X contains propositions $p_1, ..., p_m$ ($m \geq 2$) with $p = p_1 \vdash_* p_2 \vdash_* ... \vdash_* p_m = q$, I write $p \vdash\vdash q$; if moreover one of these constrained entailments is truly constrained, I write $p \vdash\vdash_{\text{true}} q$.*

(b) *A set $Z \subseteq X$ is* pathlinked (in X) *if $p \vdash\vdash q$ for all $p, q \in Z$, and* truly pathlinked (in X) *if moreover $p \vdash\vdash_{\text{true}} q$ for some (hence all) $p, q \in Z$.*

While pathlinkedness forces to a limited form of *neutral* aggregation (see Lemma 4), true pathlinkedness forces to the following degenerate aggregation rules.

Theorem 3 *If the set $\mathcal{P}$ of privileged propositions is inconsistent and truly pathlinked, there is a semi-vetodictator.*

Theorem 4 *If the set $\{p, \neg p : p \in \mathcal{P}\}$ of privileged or negated privileged propositions is truly pathlinked, there is a semi-dictator.*

In the present (and all later) theorems, the qualification "truly" can be dropped if relevance is restricted to taking a form for which pathlinkedness (of the set in question)

[35] That is, if X_1, X_2 are distinct subagendas, $A \cup B$ is consistent for all consistent $A \subseteq X_1, B \subseteq X_2$.

implies true pathlinkedness, for instance if $\mathcal{R}$ is restricted to being an equivalence relation that partitions X into logically independent subagendas.[36]

Under the conditions of Theorems 3 and 4, there may be more than one semi-(veto)dictator, and moreover there need not exist any (veto)dictator.[37]

There are many applications. The preference agenda (Example 1) is discussed later. If in Example 4 we let $\mathcal{P}$ be the set of root propositions, and if these root propositions are interconnected in the sense of Theorem 4 (3), then some individual is semi-(veto)decisive on all "fundamental issues"; and hence, premise-based or prioritarian aggregation rules take a degenerate form (at least with respect to the local decision methods D_p for root propositions $p \in X$). Let me discuss Example 2 in more detail.

Example 2 (continued) For many instances of this aggregation problem (of judging values of and constraints between variables), the conditions of Theorems 3 and 4 hold, so that semi-(veto)dictatorships are the only solutions. To make this point, let relevance be again given by (2), and let the privileged propositions be given by

$$\mathcal{P} = \{V = v : V \in \mathbf{V} \& v \in Rge(V)\} \cup \{c, \neg c : c \in \mathbf{C}\}. \tag{11}$$

Also, let $|\mathbf{V}| \geq 2$ (to make it interesting), and assume[38]

$$\{\neg c : c \in \mathbf{C}\} \notin \mathcal{J}^*. \tag{12}$$

First, consider Theorem 3. Obviously, $\mathcal{P}$ is inconsistent, as $\mathbf{C} \neq \emptyset$ by (12). Often, $\mathcal{P}$ is also truly pathlinked. The latter could be shown by establishing that

(a) $\mathcal{P}_1 := \{V = v : V \in \mathbf{V} \& v \in Rge(V)\}$ is truly pathlinked, and

(b) for all $c \in \mathbf{C}$ there are $p, q, r, s \in \mathcal{P}_1$ with $c \vdash_* p$, $q \vdash_* c$, $\neg c \vdash_* r$, $s \vdash_* \neg c$.

Part (a) might even hold in the sense of, for all $V = v, V' = v' \in \mathcal{P}_1$ with $V \neq V'$, a truly constrained entailment $V = v \vdash_* V' = v'$ (rather than an indirect path $V = v \vdash\vdash V' = v'$); indeed, there might be a set of constraints $C \subseteq \mathbf{C}$ and a set of value assignments $D \subseteq \mathcal{P}_1$ such that $V = v \vdash_{C \cup D} V' = v'$ (hence, under the constraints in C, the set of value assignments $\{V = v\} \cup D$ implies that $V' = v'$).

[36] The argument for the latter is as follows. By Lemma 2, all constrained entailments *within* any of the subagendas are unconditional entailments. This implies that the pathlinked set in question contains propositions linked by a path containing a constrained entailment *across* subagendas. The latter is truly constrained by an earlier remark.

[37] Suppose $\mathcal{R}(p) = \{p\}$ for all $p \in X$ (only *self*-relevance allowed), $\mathcal{P} = X$ (all propositions privileged), and $|X| < \infty$. Then constrained entailment reduces to standard conditional entailment, and pathlinkedness of X reduces to Nehring and Puppe's (2002/2005) *total blockedness* condition whereby there is a path of *conditional* entailments between any $p, q \in X$. Dokow and Holzman (2005) show that *parity rules* F, defined on $\mathcal{J}^n$ by $F(A_1, ..., A_n) = \{p \in X : |\{i \in M : p \in A_i\}|$ is odd$\}$ for an odd-sized subgroup $M \subseteq N$, take values in $\mathcal{J}$ for certain agendas X that are totally blocked (hence pathlinked, in fact *truly* pathlinked) and satisfy an algebraic condition. Such a parity rule is also III and agreement preserving, and hence provides the required counterexample because every $i \in M$ a semi-dictator and a semi-vetodictator, but not a dictator and not a vetodictator (unless $|M| = 1$).

[38] Condition (12) requires that *at least one* constraint between variables holds, i.e. that the variables are not totally independent from each other. This assumption is natural in cases where the question is not *whether* but only *how* the variables affect each other, as it is the case for macroeconomic variables.

Part (b) might hold for the following reasons. Consider a constraint $c \in \mathbf{C}$. Plausibly, $V = v \vdash_D \neg c$ for some $V = v \in \mathcal{P}_1$ and $D \subseteq \mathcal{P}_1$; here, $\{V = v\} \cup D$ is a set of value assignments violating the constraint c. It is also plausible that $c \vdash_D V = v$ for some $V = v \in \mathcal{P}_1$ and some $D \subseteq \mathcal{P}_1$; here, the value assignments in D imply, under the constraint c, that $V = v$. Moreover, we could have $V = v \vdash_D c$ for some $V = v \in \mathcal{P}_1$ and $D \subseteq \mathcal{P}_1$: this is so if the set of value assignments $\{V = v\} \cup D$ violates all constraints in $\mathbf{C}$ except c, hence entails c by (12). Finally, we could have $\neg c \vdash_{\tilde{C} \cup D} V = v$ for some $V = v \in \mathcal{P}_1$ and $D \subseteq \mathcal{P}_1$, and some set $\tilde{C}$ of negated constraints; indeed, suppose $\{\neg c\} \cup \tilde{C}$ contains the negations of all except of one constraint in $\mathbf{C}$, hence entails the remaining constraint by (12); under this remaining constraint, the value assignments in D could imply that $V = v$.

Now consider Theorem 4. The special form (11) of $\mathcal{P}$ in fact implies that the conditions of Theorem 4 hold whenever those of Theorem 3 hold (hence in many cases, as argued above). Specifically, let $\mathcal{P}$ be truly pathlinked. To prove that also $\{p, \neg p : p \in \mathcal{P}\}$ is truly pathlinked, it suffices to show that, for all $V = v \in \mathcal{P}$, there is a $p \in \mathcal{P}$ with $\neg(V = v) \vdash\vdash p$ and $p \vdash\vdash \neg(V = v)$. Consider any $V = v \in \mathcal{P}$, and choose any $p \in \mathbf{C}$ ($\subseteq \mathcal{P}$). As $\mathcal{P}$ is pathlinked and by (11) contains $\neg p$ and $V = v$, we have $\neg p \vdash\vdash V = v$ and $V = v \vdash\vdash \neg p$; hence (using Lemma 1 below) $\neg(V = v) \vdash\vdash p$ and $p \vdash\vdash \neg(V = v)$, as desired.

I now derive lemmas that will help both prove the theorems and understand constrained entailment. I first give a sufficient condition for when a constrained entailment reduces to an unconditional entailment.

Lemma 2 *For all $p, q \in X$ with $\mathcal{R}(p) \subseteq \mathcal{R}(\neg q)$ or $\mathcal{R}(\neg q) \subseteq \mathcal{R}(p)$, $p \vdash_* q$ if and only if $p \vdash q$.*

Proof. Let p, q be as specified. Obviously, $p \vdash q$ implies $p \vdash_\emptyset q$. Suppose for a contradiction that $p \vdash_* q$, say $p \vdash_Y q$, but $p \nvdash q$. Then $\{p, \neg q\}$ is consistent. So there is an $B \in \mathcal{J}$ containing p and $\neg q$. Then

- the set $B \cap \{r, \neg r : r \mathcal{R} p\}$ is an explanation of p;
- the set $B \cap \{r, \neg r : r \mathcal{R} \neg q\}$ is an explanation of $\neg q$.

One of these two sets is a superset of the other one, as $\mathcal{R}(p) \subseteq \mathcal{R}(\neg q)$ or $\mathcal{R}(\neg q) \subseteq \mathcal{R}(p)$; call this superset A. As $p \vdash_Y q$, $A \cup Y$ is consistent. So, as $A \vdash p$ and $A \vdash \neg q$, $\{p, \neg q\} \cup Y$ is consistent. It follows that $\{p\} \cup Y \nvdash q$, in contradiction to $p \vdash_Y q$. ■

The next fact helps in choosing the set Y in a constrained entailment.

Lemma 3 *For all $p, q \in X$, if $p \vdash_* q$ then $p \vdash_Y q$ for some set Y containing no proposition relevant to p or to $\neg q$.*

Proof. Let $p, q \in X$, and assume $p \vdash_* q$, say $p \vdash_Y q$. The proof is done by showing that $p \vdash_{Y \setminus (\mathcal{R}(p) \cup \mathcal{R}(\neg q))} q$. Suppose for a contradiction that not $p \vdash_{Y \setminus (\mathcal{R}(p) \cup \mathcal{R}(\neg q))} q$. Then

(*) $\{p, \neg q\} \cup Y \setminus (\mathcal{R}(p) \cup \mathcal{R}(\neg q))$ is consistent.

I show that

(**) $p \vdash p'$ for all $p' \in Y \cap \mathcal{R}(p)$ and $\neg q \vdash q'$ for all $q' \in Y \cap \mathcal{R}(\neg q)$,

which together with (*) implies that $\{p, \neg q\} \cup Y$ is consistent, a contradiction since $p \vdash_Y q$. Suppose for a contradiction that $p' \in Y \cap \mathcal{R}(p)$ but $p \not\vdash p'$. Then there is a $B \in \mathcal{J}$ containing p and $\neg p'$. The set $A := B \cap \{r, \neg r : r \mathcal{R} p\}$ does not entail $\neg p$, hence is an explanation of p (as $\mathcal{R}$ is a relevance relation). So $A \cup Y$ is consistent (as $p \vdash_Y q$), a contradiction since $A \cup Y$ contains both p' and $\neg p'$. For analogous reasons, for all $q' \in Y \cap X^l$ it cannot be that $\neg q \not\vdash q'$. ■

Now I introduce notions of decisive and semi-decisive coalitions, and I show that semi-decisiveness is preserved along paths of constrained entailments.

Definition 5 *A possibly empty coalition $C \subseteq N$ is* decisive *(respectively,* semi-decisive*)* for $p \in X$ *if its members have judgment sets $A_i \in \mathcal{J}$, $i \in C$, containing p, such that $p \in F(A_1, ..., A_n)$ for all $A_i \in \mathcal{J}$, $i \in N \backslash C$ (respectively, for all $A_i \in \mathcal{J}$, $i \in N \backslash C$, not containing p).*

While a decisive coalition for p can (by appropriate judgment sets) always socially enforce p, a semi-decisive coalition can do so provided all other individuals reject p. Let $\mathcal{W}(p)$ and $\mathcal{C}(p)$ be the sets of decisive and semi-decisive coalitions for $p \in X$, respectively.

Lemma 4 *For all $p, q \in X$, if $p \vdash_* q$ then $\mathcal{C}(p) \subseteq \mathcal{C}(q)$. In particular, if $Z \subseteq X$ is pathlinked, all $p \in Z$ have the same semi-decisive coalitions.*[39]

Proof. Suppose $p, q \in X$, and $p \vdash_* q$, say $p \vdash_Y q$, where by Lemma 3 w.l.o.g. $Y \cap \mathcal{R}(p) = Y \cap \mathcal{R}(\neg q) = \emptyset$. Let $C \in \mathcal{C}(p)$. So there are sets $A_i \in \mathcal{J}$, $i \in C$, containing p, such that $p \in F(A_1, ..., A_n)$ for all $A_i \in \mathcal{J}$, $i \in N \backslash C$, containing $\neg p$. By Y's consistency with every explanation of p, it is possible to change each A_i, $i \in C$, into a set (still in $\mathcal{J}$) that contains every $y \in Y$ and has the same intersection with $\mathcal{R}(p)$ as A_i; this change preserves the required properties, i.e. it preserves that $p \in A_i$ for all $i \in C$ (as $\mathcal{R}$ is a relevance relation), and preserves that $p \in F(A_1, ..., A_n)$ for all $A_i \in \mathcal{J}$, $i \in N \backslash C$, containing $\neg p$ (by $Y \cap \mathcal{R}(p) = \emptyset$ and III). So we may assume w.l.o.g. that $Y \subseteq A_i$ for all $i \in C$. Hence, by $\{p\} \cup Y \vdash q$, all A_i, $i \in C$, contain q.

To establish that $C \in \mathcal{C}(q)$, I consider any sets $A_i \in \mathcal{J}$, $i \in N \backslash C$, all containing $\neg q$, and I show that $q \in F(A_1, ..., A_n)$. We may assume w.l.o.g. that $Y \subseteq A_i$ for all $i \in N \backslash C$, by an argument like the one above (using that Y is consistent with any explanation of $\neg q$, $\mathcal{R}$ is a relevance relation, $Y \cap \mathcal{R}(\neg q) = \emptyset$, and III). As $\{\neg q\} \cup Y \vdash \neg p$, all A_i, $i \in N \backslash C$, contain $\neg p$. Hence $p \in F(A_1, ..., A_n)$. Moreover, $Y \subseteq F(A_1, ..., A_n)$ by $Y \subseteq \mathcal{P}$. So, as $\{p\} \cup Y \vdash q$, $q \in F(A_1, ..., A_n)$, as desired. ■

I can now prove Theorems 3 and 4.

Proof of Theorem 3. Let $\mathcal{P}$ be inconsistent and truly pathlinked. I first prepare the proof by establishing three simple claims.

Claim 1. (i) The set $\mathcal{C}(p)$ is the same for all $p \in \mathcal{P}$; call it $\mathcal{C}_0$. (ii) The set $\mathcal{C}(\neg p)$ is the same for all $p \in \mathcal{P}$.

Part (i) follows from Lemma 4. Part (ii) follows from it too because, by Lemma 1, $\{\neg p : p \in \mathcal{P}\}$ is like $\mathcal{P}$ pathlinked, q.e.d.

[39] Constrained entailments preserve semi-decisiveness but usually not decisiveness.

Claim 2. $\emptyset \notin \mathcal{C}_0$ and $N \in \mathcal{C}_0$.

By agreement preservation, $N \in \mathcal{C}_0$. Suppose for a contradiction that $\emptyset \in \mathcal{C}_0$. Consider any judgment set $A \in \mathcal{J}$. Then $F(A, ..., A)$ contains all $p \in \mathcal{P}$, by $N \in \mathcal{C}_0$ if $p \in A$, and by $\emptyset \in \mathcal{C}_0$ if $p \notin A$. Hence $F(A, ..., A)$ is inconsistent, a contradiction, q.e.d.

By Claim 2, there is a minimal coalition C in $\mathcal{C}_0$ (with respect to inclusion), and $C \neq \emptyset$. By $C \neq \emptyset$, there is a $j \in C$. Write $C_{-j} := C \backslash \{j\}$. As $\mathcal{P}$ is truly pathlinked, there exist $p \in \mathcal{P}$ and $r, s \in X$ such that $p \vdash\vdash r$, $r \vdash_* s$ truly, and $s \vdash\vdash p$.

Claim 3. $\mathcal{C}(r) = \mathcal{C}(s) = \mathcal{C}_0$; hence $C \in \mathcal{C}(r)$ and $C_{-j} \notin \mathcal{C}(s)$.

By Lemma 4, $\mathcal{C}(p) \subseteq \mathcal{C}(r) \subseteq \mathcal{C}(s) \subseteq \mathcal{C}(p)$. So $\mathcal{C}(r) = \mathcal{C}(s) = \mathcal{C}(p) = \mathcal{C}_0$, q.e.d.

Now let Y be such that $r \vdash_Y s$, where by Lemma 3 w.l.o.g. $Y \cap \mathcal{R}(r) = Y \cap \mathcal{R}(\neg s) = \emptyset$. By $C \in \mathcal{C}(r)$, there are judgment sets $A_i \in \mathcal{J}$, $i \in C$, containing r, such that $r \in F(A_1, ..., A_n)$ for all $A_i \in \mathcal{J}$, $i \in N \backslash C$, not containing r. I assume w.l.o.g. that

$$\text{for all } i \in C_{-j},\ Y \subseteq A_i, \text{ hence (by } \{r\} \cup Y \vdash s)\ s \in A_i, \tag{13}$$

which I may do by an argument like that in the proof of Lemma 4 (using that Y is consistent with any explanation of q, $\mathcal{R}$ is a relevance relation, $Y \cap \mathcal{R}(r) = \emptyset$, and III). By (13) and as $C_{-j} \notin \mathcal{C}(s)$ (see Claim 3), there are sets $B_i \in \mathcal{J}$, $i \in N \backslash C_{-j}$, containing $\neg s$, such that, writing $B_i := A_i$ for all $i \in C_{-j}$,

$$\neg s \in F(B_1, ..., B_n). \tag{14}$$

I may w.l.o.g. modify the sets B_i, $i \in N \backslash C_{-j}$, into new sets in $\mathcal{J}$ as long as their intersections with $\mathcal{R}(\neg s)$ stays the same (because the new sets then still contain $\neg s$ as $\mathcal{R}$ is a relevance relation, and still satisfy (14) by III). First, I modify the set B_i for $i = j$: as $r \vdash_* s$ truly, $B_j \cap \{t, \neg t : t \in \mathcal{R}(\neg s)\}$ (an explanation of $\neg s$) is consistent with any explanation of r, hence with $A_j \cap \{t, \neg t : t \in \mathcal{R}(r)\}$, so that I may assume that $A_j \cap \{t, \neg t : t \in \mathcal{R}(r)\} \subseteq B_j$; which implies that

$$B_i \cap \mathcal{R}(r) = A_i \cap \mathcal{R}(r) \text{ for all } i \in C. \tag{15}$$

Second, I modify the sets B_i, $i \in N \backslash C$: I assume (using that $Y \cap \mathcal{R}(\neg s) = \emptyset$ and Y's consistency with any explanation of $\neg s$) that

$$\text{for all } i \in N \backslash C,\ Y \subseteq B_i, \text{ hence (as } \{\neg s\} \cup Y \vdash \neg r)\ \neg r \in B_i. \tag{16}$$

The definition of the sets A_i, $i \in C$, and (16) imply, via (15) and III, that

$$r \in F(B_1, ..., B_n). \tag{17}$$

By (14), (17), and the inconsistency of $\{r, \neg s\} \cup Y$, the set Y is not a subset of $F(B_1, ..., B_n)$. So there is a $y \in Y$ with $y \notin F(B_1, ..., B_n)$. We have $\{j\} \in \mathcal{C}(\neg y\}$ for the following two reasons.

- B_j contains $\neg y$; otherwise $y \in B_i$ for all $i \in N$, so that $y \in F(B_1, ..., B_n)$ by $y \in \mathcal{P}$.
- Consider any sets $C_i \in \mathcal{J}$, $i \neq j$, not containing $\neg y$, i.e. containing y. I show that $\neg y \in A := F(C_1, ..., C_{j-1}, B_j, C_{j+1}, ..., C_n)$. For all $i \neq j$, $C_i \cap \{t, \neg t : t \in \mathcal{R}(y)\}$ is consistent with y, hence is an explanation of y (as $\mathcal{R}$ satisfies

"no underdetermination"); for analogous reasons, $B_i \cap \{t, \neg t : t \in \mathcal{R}(y)\}$ is an explanation of y. These two explanations must be identical by $y \in \mathcal{P}$. So $C_i \cap \mathcal{R}(y) = B_i \cap \mathcal{R}(y)$. Hence, by $y \notin F(B_1, ..., B_n)$ and III, $y \notin A$. So $\neg y \in A$, as desired.

By $\{j\} \in \mathcal{C}(\neg y)$ and Claim 1, $\{j\} \in \mathcal{C}(\neg q)$ for all $q \in \mathcal{P}$. So j is a semi-vetodictator. ■

Proof of Theorem 4. Let $\{p, \neg p : p \in \mathcal{P}\}$ be truly pathlinked. I will reduce the proof to that of Theorem 3. I start again with two simple claims.

Claim 1. The set $\mathcal{C}(q)$ is the same for all $q \in \{p, \neg p : p \in \mathcal{P}\}$; call it $\mathcal{C}_0$.

This follows immediately from Lemma 4, q.e.d.

Claim 2. $\emptyset \notin \mathcal{C}_0$ and $N \in \mathcal{C}_0$.

By agreement preservation, $N \in \mathcal{C}(p)$ for all $p \in \mathcal{P}$; hence $N \in \mathcal{C}_0$. This implies, for all $p \in \mathcal{P}$, that $\emptyset \notin \mathcal{C}(\neg p)$; hence $\emptyset \notin \mathcal{C}_0$, q.e.d.

Now by an analogous argument to that in the proof of Theorem 3, but based this time on the present Claims 1 and 2 rather than on the two first claims in Theorem 3's proof, one can show that there exists an individual j such that $\{j\} \in \mathcal{C}(\neg q)$ for all $q \in \mathcal{P}$. So, by the present Claim 1 (which is stronger than the first claim in Theorem 3's proof),

$$\{j\} \in \mathcal{C}(q) \text{ for all } q \in \mathcal{P}. \tag{18}$$

So j is a semi-dictator, for the following reason. Let $q \in \mathcal{P}$ and let $(A_1, ..., A_n) \in \mathcal{J}^n$ be such that $q \in A_j$ and $q \notin A_i$, $i \neq j$. By (18) there is a set $B_j \in \mathcal{J}$ containing q such that $q \in F(B_1, ..., B_n)$ for all $B_i \in \mathcal{J}$, $i \neq j$, not containing q. Since q has only one explanation (by $q \in \mathcal{P}$), the two explanations $A_j \cap \{t, \neg t : t \in \mathcal{R}(q)\}$ and $B_j \cap \{t, \neg t : t \in \mathcal{R}(q)\}$ are identical. So $A_j \cap \mathcal{R}(q) = B_j \cap \mathcal{R}(q)$. Hence, using III and the definition of B_j, $q \in F(A_1, ..., A_n)$, as desired. ■

9 Dictatorship and strong dictatorship

In fact, the semi-dictator of Theorem 4 is in many cases (including the preference aggregation problem) a dictator, and in some cases even a strong dictator in the sense of the following definition that generalises the classical notion of strong dictatorship in social choice theory.

Definition 6 *An individual i is a* strong dictator *if $F(A_1, ..., A_n) = A_i$ for all $(A_1, ..., A_n) \in \mathcal{J}^n$.*

So a strong dictator imposes his judgments on all rather than just privileged propositions. I will give simple criteria for obtaining (weak or strong) dictatorship, in terms of the following irreversibility property.

Definition 7 *For $p, q \in X$, p* irreversibly *constrained entails q if $p \vdash_Y q$ for a set Y for which $\{q\} \cup Y \nvdash p$.*

So a constrained entailment $p \vdash_* q$ is irreversible if the constrained entailment is not a "constrained equivalence", i.e. if p and q do not conditionally entail each other (for at least one choice of Y). If X is the preference agenda (with Arrowian $\mathcal{R}$ and

$\mathcal{P}$), *all* constrained entailments between (distinct) propositions are irreversible. For instance, $xRy \vdash_* xRz$ is irreversible (for distinct options x, y, z), since $xRy \vdash_{\{yPz\}} xRz$, where $\{xRz, yPz\} \nvdash xRy$.

By the next result, the semi-dictatorship of Theorem 4 becomes a dictatorship if we only slightly strengthen the pathlinkedness condition: in *at least one* path, *at least one* constrained entailment should be irreversible.

Definition 8 (a) *For propositions $p, q \in X$, I write $p \vdash\vdash_{\text{irrev}} q$ if X contains propositions $p_1, ..., p_m$ ($m \geq 2$) with $p = p_1 \vdash_* p_2 \vdash_* ... \vdash_* p_m = q$, where at least one of these constrained entailments is irreversible.*

(b) *A pathlinked set $Z \subseteq X$ is* irreversibly pathlinked (in X) *if $p \vdash\vdash_{\text{irrev}} q$ for some (hence all) $p, q \in Z$.*

Theorem 5 *If the set $\{p, \neg p : p \in \mathcal{P}\}$ of privileged or negated privileged propositions is truly and irreversibly pathlinked, some individual is a dictator.*

As an application, I obtain the full Arrow theorem by proving that, if X is the preference agenda, $\{p, \neg p : p \in \mathcal{P}\}$ is truly and irreversibly pathlinked.[40]

Corollary 1 *(Arrow's Theorem) For the preference agenda (with Arrowian $\mathcal{R}$ and $\mathcal{P}$), some individual is a dictator.*

Proof. Let X be the preference agenda, with $\mathcal{R}$ and $\mathcal{P}$ Arrowian. I show that (i) $\mathcal{P}$ is pathlinked, and (ii) there are $r, s \in \mathcal{P}$ with true and irreversible constrained entailments $r \vdash_* \neg s \vdash_* r$. Then, by (i) and Lemma 1, $\{\neg p : p \in \mathcal{P}\}$ is (like $\mathcal{P}$) pathlinked, which together with (ii) implies that $\{p, \neg p : p \in \mathcal{P}\}$ is truly and irreversibly pathlinked, as desired.

(ii): For any pairwise distinct options $x, y, z \in Q$, we have $xPy \vdash_{\{yPz\}} xRz$ ($= \neg zRx$), and $xRz \vdash_{\{zPy\}} xPy$, in each case truly and irreversibly.

(i): Consider any $xPy, x'Py' \in \mathcal{P}$. I show that $xPy \vdash\vdash x'Py'$. The paths to be constructed depend on whether $x \in \{x', y'\}$ and whether $y \in \{x', y'\}$. As $x \neq y$ and $x' \neq y'$, the following list of cases is exhaustive. Case $x \neq x', y' \& y \neq x', y'$: $xPy \vdash_{\{x'Px, yPy'\}} x'Py'$. Case $y = y' \& x \neq x', y'$: $xPy \vdash_{\{x'Px\}} x'Py = x'Py'$. Case $y = x' \& x \neq x', y'$: $xPy \vdash_{\{yPy'\}} xPy' \vdash_{\{x'Px\}} x'Py'$. Case $x = x' \& y \neq y', x'$: $xPy \vdash_{\{yPy'\}} xPy'$. Case $x = y' \& y \neq x', y'$: $xPy \vdash_{\{x'Px\}} x'Py \vdash_{\{yPx\}} x'Px$. Case $x = x' \& y = y'$: $xPy \vdash_{\emptyset} xPy$. Case $x = y' \& y = x'$: taking any $z \in Q \backslash \{x, y\}$, $xPy \vdash_{\{yPz\}} xPz \vdash_{\{yPx\}} yPz \vdash_{\{zPx\}} yPx$. ■

The proof of Theorem 5 uses two further lemmas. For any set $\mathcal{S}$ of coalitions $C \subseteq N$, I define $\overline{\mathcal{S}} := \{C' \subseteq N : C \subseteq C' \text{ for some } C \in \mathcal{S}\}$.

Lemma 5 *For all $p, q \in X$,*

(a) *$p \vdash_* q$ irreversibly if and only if $\neg q \vdash_* \neg p$ irreversibly;*

(b) *if $p \vdash_* q$ irreversibly then $\overline{\mathcal{C}(p)} \subseteq \mathcal{C}(q)$.*

[40] This property of $\{p, \neg p : p \in \mathcal{P}\}$ strengthens Nehring's (2003) finding that the preference agenda is totally blocked, which gave him already a weaker version of Arrow's theorem. Part (i) of our proof is analogous to Nehring's proof and also to proofs for the *strict* preference agenda by Dietrich and List (forthcoming-b) and Dokow and Holzman (2005).

Proof. Let $p, q \in X$. Part (a) follows from Lemma 1 and the fact that, for all $Y \subseteq \mathcal{P}$, $\{q\} \cup Y \nvdash p$ if and only if $\{\neg p\} \cup Y \nvdash \neg q$.

Regarding (b), suppose $p \vdash_* q$ irreversibly, say $p \vdash_Y q$ with $\{q\} \cup Y \nvdash p$. We can assume w.l.o.g. that $Y \cap \mathcal{R}(p) = Y \cap \mathcal{R}(\neg q) = \emptyset$, since otherwise we could replace Y by $Y' := Y \backslash (\mathcal{R}(p) \cup \mathcal{R}(\neg q))$, for which still $p \vdash_{Y'} q$ (by the proof of Lemma 3) and $\{q\} \cup Y' \nvdash p$. To show $\overline{\mathcal{C}(p)} \subseteq \mathcal{C}(q)$, consider any $C' \in \overline{\mathcal{C}(p)}$. So there is a $C \in \mathcal{C}(p)$ with $C \subseteq C'$. Hence there are $A_i \in \mathcal{J}$, $i \in C$, containing p, such that $p \in F(A_1, ..., A_n)$ for all $A_i \in \mathcal{J}$, $i \in N \backslash C$, containing $\neg p$. Like in earlier proofs, I may suppose w.l.o.g. that, for all $i \in C$, $Y \subseteq A_i$ (using that Y is consistent with all explanations of p, $\mathcal{R}$ is a relevance relation, III, and $Y \cap \mathcal{R}(p) = \emptyset$); hence, by $\{p\} \cup Y \vdash q$, $q \in A_i$ for all $i \in C$. Further, as $\{\neg p, q\} \cup Y$ is consistent (by $\{q\} \cup Y \nvdash p$), there are sets $A_i \in \mathcal{J}$, $i \in C' \backslash C$, such that $\{\neg p, q\} \cup Y \subseteq A_i$ for all $i \in C' \backslash C$.

I have to show that $q \in F(A_1, ..., A_n)$ for all $A_i \in \mathcal{J}$, $i \in N \backslash C'$, containing $\neg q$. Consider such sets A_i, $i \in N \backslash C'$. Again, we may assume w.l.o.g. that for all $i \in N \backslash C'$, $Y \subseteq A_i$ (as Y is consistent with all explanations of $\neg q$, $\mathcal{R}$ is a relevance relation, III, and $Y \cap \mathcal{R}(\neg q) = \emptyset$), which by $\{\neg q\} \cup Y \vdash \neg p$ implies that $\neg p \in A_i$ for all $i \in N \backslash C'$. In summary then,

$$A_i \supseteq \begin{cases} \{p, q\} \cup Y & \text{for all } i \in C \\ \{\neg p, q\} \cup Y & \text{for all } i \in C' \backslash C \\ \{\neg p, \neg q\} \cup Y & \text{for all } i \in N \backslash C'. \end{cases}$$

So $p \in F(A_1, ..., A_n)$ (by the choice of the sets A_i, $i \in C$) and $Y \subseteq F(A_1, ..., A_n)$ (by $Y \subseteq \mathcal{P}$). Hence, as $\{p\} \cup Y \vdash q$, $q \in F(A_1, ..., A_n)$, as desired. ■

In the following characterisation of decisive coalitions it is crucial that $p \in \mathcal{P}$.

Lemma 6 *If $p \in \mathcal{P}$, $\mathcal{W}(p) = \{C \subseteq N$: all coalitions $C' \supseteq C$ are in $\mathcal{C}(p)\}$.*

Proof. Let $p \in \mathcal{P}$ and $C \subseteq N$. If $C \in \mathcal{W}(p)$ then clearly all coalitions $C' \supseteq C$ are in $\mathcal{C}(p)$. Conversely, suppose all coalitions $C' \supseteq C$ are in $\mathcal{C}(p)$. As $C \in \mathcal{C}(p)$, there are sets A_i, $i \in C$, containing p, such that $p \in F(A_1, ..., A_n)$ for all sets A_i, $i \in N \backslash C$, not containing p. To show that $C \in \mathcal{W}(p)$, consider any sets A_i, $i \in N \backslash C$ (containing or not containing p); I show that $p \in F(A_1, ..., A_n)$. Let $C' := C \cup \{i \in N \backslash C : p \in A_i\}$. By $C \subseteq C'$, $C' \in \mathcal{C}(p)$. So there are sets B_i, $i \in C'$, containing p, such that $p \in F(B_1, ..., B_n)$ for all sets B_i, $i \in N \backslash C'$, not containing p. As p has a single explanation, we have for all $i \in C'$ $A_i \cap \{r, \neg r : r \in \mathcal{R}(p)\} = B_i \cap \{r, \neg r : r \in \mathcal{R}(p)\}$, hence $A_i \cap \mathcal{R}(p) = B_i \cap \mathcal{R}(p)$. So, by III and the definition of the sets B_i, $i \in C'$, and since $p \notin A_i$ for all $i \in N \backslash C'$, $p \in F(A_1, ..., A_n)$, as desired. ■

Proof of Theorem 5. Let $\{p, \neg p : p \in \mathcal{P}\}$ be truly and irreversibly pathlinked. By Theorem 4, there is a semi-dictator i. I show that i is a dictator.

Claim. For all $q \in \{p, \neg p : p \in \mathcal{P}\}$, $\mathcal{C}(q)$ contains all coalitions containing i.

Consider any $q \in \{q, \neg q : q \in \mathcal{P}\}$ and any coalition $C \subseteq N$ containing i. By true pathlinkedness there exist $p \in \mathcal{P}$ and $r, s \in X$ such that $p \vdash\vdash r \vdash_* s \vdash\vdash q$, where $r \vdash_* s$ is a truly constrained entailment. By $\{i\} \in \mathcal{C}(p)$ and Lemma 4, $\{i\} \in \mathcal{C}(r)$. So, by Lemma 5(b), $C \in \mathcal{C}(s)$. Hence, by Lemma 4, $C \in \mathcal{C}(q)$, q.e.d.

By this claim and Lemma 6, $\{i\} \in \mathcal{W}(p)$ for all $p \in \mathcal{P}$. This implies that i is a dictator, by an argument similar to the one that completed the proof of Theorem 4. ■

Finally, for what agendas do we even obtain *strong* dictatorship? Surely not for the preference agenda, as it is well-known that Arrow's conditions only imply weak dictatorship.[41]

Trivially, if *all* propositions are privileged, every dictatorship is strong:

Corollary 2 *If $\mathcal{P} = X$, and X is truly and irreversibly pathlinked, some individual is a strong dictator.*

But the assumption $\mathcal{P} = X$ removes nearly all generality: agreement preservation becomes global unanimity preservation, and the relevance relation is forced to give each $p \in X$ a single explanation. However, strong dictatorship follows under a much less restrictive condition than $\mathcal{P} = X$. Call $p \in X$ *logically equivalent to* $A \subseteq X$ if A entails p and p entails all $q \in A$ (i.e. intuitively, if p is equivalent to the conjunction of all $q \in A$). For instance, $a \wedge b$ is equivalent to $\{a, b\}$ (where $a, b, a \wedge b \in X$).

Theorem 6 *If $\{p, \neg p : p \in \mathcal{P}\}$ is truly and irreversibly pathlinked and each proposition in X is logically equivalent to a set of negated privileged propositions $A \subseteq \{\neg p : p \in \mathcal{P}\}$, some individual is a strong dictator.*

Proof. Let the assumptions hold. By Theorem 5, there is a dictator i. To show that i is a strong dictator, I consider any $(A_1, ..., A_n) \in \mathcal{J}^n$, and I show that $A_i = F(A_1, ..., A_n)$. Obviously, it suffices to show that $F(A_1, ..., A_n) \subseteq A_i$. Suppose $q \in F(A_1, ..., A_n)$. By assumption, q is logically equivalent to some $A \subseteq \{\neg p : p \in \mathcal{P}\}$. For all $\neg p \in A$, we have $\neg p \in F(A_1, ..., A_n)$ (by $q \vdash \neg p$), hence $p \notin F(A_1, ..., A_n)$, and so $p \notin A_i$ (as $p \in \mathcal{P}$ and i is a dictator), implying that $\neg p \in A_i$. This shows that $A \subseteq A_i$. So $q \in A_i$ (since $A \vdash q$), as desired. ■

The preference agenda X, which has *not strongly* dictatorial solutions, indeed violates the extra condition in Theorem 6: some propositions in X (namely precisely the privileged propositions xPy) are not logically equivalent to any set of negated privileged propositions xRy.

Example 2 (continued) As argued earlier, $\{p, \neg p : p \in \mathcal{P}\}$ is truly pathlinked in many instances of this aggregation problem. The other conditions in Theorem 6 also often hold, so that strong dictatorship follows. The reasons are simple.

First, X is often rich in irreversible constrained entailments. For instance, if X contains value assignments $V = 3$ and $W = 3$ and the constraint $W > V$**,** then $V = 3 \vdash_* \neg(W = 3)$ irreversibly, since $V = 3 \vdash_{\{W>V\}} \neg(W = 3)$ but $\{\neg(W = 3), W > V\} \nvdash V = 3$; or, if X contains a constraint c that is strictly stronger than another constraint $c' \in \mathbf{C}$, then $c \vdash_* c$ irreversibly, since $c \vdash_\emptyset c'$ but $\{c'\} \cup \emptyset = \{c'\} \nvdash c$.

Second, if $\mathcal{P}$ is again given by (11), each proposition $q \in X$ is indeed logically equivalent to a set of negated privileged propositions $A \subseteq \{\neg p : p \in \mathcal{P}\}$: if q has the form $V = v$, one should take $A = \{\neg(V = v') : v' \in Rge(V) \backslash \{v\}\}$; otherwise q has the form $\neg p$ with $p \in \mathcal{P}$, and so one may take $A = \{q\}$.

[41] Lexicographic dictatorships satisfy all conditions but are only weak dictatorships.

10 Conclusion

The impossibility findings might be interpreted as showing how relevance $\mathcal{R}$ should *not* be specified. Indeed, in order to enable non-degenerate III aggregation rules, $\mathcal{R}$ must display sufficiently many inter-relevances. But such richness in inter-relevances may imply that collective decisions have to be made in a "holistic" manner: many semantically unrelated decisions must be bundled and decided simultaneously. Two propositions, say one on traffic regulations and one on diplomatic relations with Argentina, have to be treated simultaneously if the relevance relation (specified sufficiently richly to enable non-degenerate aggregation rules) displays some possibly indirect link between the two.[42] Large and semantically disparate decision problems are a hard challenge in practice.

As I began to discuss in Section 3, several types of relevance relations, hence informational constraints, are of interest to aggregation theory. Which III aggregation rules are there if relevance is, for instance, transitive? Or asymmetric? Or well-founded? In addition to such questions, it is worth exploring further the premise-based and prioritarian approaches. If a distance-based approach compatible with the informational constraint III could be developed, the theories of belief merging and judgment aggregation would meet. Developing different types of III aggregation procedures should go hand in hand with developing objective criteria for when to consider a proposition as relevant to another, i.e. for which informational restriction to impose. This second research goal has a normative dimension. Reaching both goals would enable us to give concrete recommendations for practical group decision-making.

11 References

Bradley, R (forthcoming) Taking advantage of difference in opinion, *Episteme*, forthcoming

Chapman, B (2002) Rational Aggregation, Polit Philos Econ 1(3): 337-354

Claussen, CA, Roisland, O (2005) Collective economic decisions and the discursive paradox, working paper, Central Bank of Norway Research Division

Dietrich, F (2006) Judgment aggregation: (im)possibility theorems, *J Econ Theory* 126(1): 286-298

Dietrich, F (forthcoming) A generalised model of judgment aggregation, *Soc Choice Welfare*

Dietrich, F (2005) The possibility of judgment aggregation on agendas with subjunctive implications, working paper, Maastricht Univ

Dietrich, F, List, C (2004) A liberal paradox for judgment aggregation, working paper, London School of Economics

Dietrich, F, List, C (forthcoming-a) Judgment aggregation by quota rules, *J Theor Polit*

[42] That is, if the two propositions are related in terms of the transitive, symmetric and reflexive closure of relevance $\mathcal{R}$. This closure partitions the totality of propositions (questions) into equivalence classes of irreducible decision problems. If there is a single equivalence class, *all* decisions (including those on traffic regulations and on diplomatic relations) have to be treated simultaneously.

Dietrich, F, List, C (forthcoming-b) Arrow's theorem in judgment aggregation, *Soc Choice Welfare*

Dietrich, F, List, C (2006) Judgment aggregation on restricted domains, working paper, Maastricht Univ

Dokow, E, Holzman, R (2005) Aggregation of binary evaluations, working paper, Technion Israel Institute of Technology

Eckert, D, Pigozzi, G (2005) Belief merging, judgment aggregation and some links with social choice theory, in *Belief Change in Rational Agents*, J. Delgrande et al. (eds.)

Fenstad, JF (1980) *General recursion theory: an axiomatic approach*, Springer

Genest, C, Zidek, JV (1986) "Combining Probability Distributions: A Critique and Annotated Bibliography", *Statistical Science* 1(1): 113-135

Gärdenfors, P (2006) An Arrow-like theorem for voting with logical consequences, *Econ Philos* 22(2): 181-190

Konieczny, S, Pino-Perez, R (2002) Merging information under constraints: a logical framework, *J Logic Comput* 12(5): 773-808

List, C (2004) A model of path-dependence in decisions over multiple propositions, *Am Polit Sci Rev* 98(3): 495-513

List, C (2005) Group knowledge and group rationality: a judgment aggregation perspective, *Episteme* 2(1): 25-38

List, C, Pettit, P (2002) Aggregating sets of judgments: an impossibility result. *Econ Philos* 18: 89-110

List, C, Pettit, P (2004) Aggregating sets of judgments: two impossibility results compared. *Synthese* 140(1-2): 207-235

Mongin, P (2005-a) Factoring out the impossibility of logical aggregation, working paper, CNRS, Paris

Mongin, P (2005-b) Spurious Unanimity and the Pareto Principle, *LSE Choice Group Working Papers* 1 (5)

Nehring, K (2003) Arrow's theorem as a corollary. *Econ Letters* 80(3): 379-382

Nehring, K (2005) The (im)possibility of a Paretian rational, working paper, Univ California at Davies

Nehring, K, Puppe, C (2002) Strategy-proof social choice on single-peaked domains: possibility, impossibility and the space between, working paper, Univ California at Davies

Nehring, K, Puppe, C (2005) The structure of strategy-proof social choice, part II: non-dictatorship, anonymity and neutrality, working paper, Karlsruhe Univ

Nehring, K., Puppe, C (2006) Consistent judgment aggregation: the truth-functional Case, working paper, Karlsruhe Univ

Parikh, R (1999) Beliefs, belief revision and splitting languages, *Logic, language and computation* 2: 266-278

Pauly, M, van Hees, M (2006) Logical constraints on judgment aggregation, *J Philos Logic* 35: 569-585

Rubinstein, A, Fishburn, P (1986) Algebraic aggregation theory, *J Econ Theory* 38: 63-77

van Hees, M (forthcoming) The limits of epistemic democracy, *Soc Choice Welfare*

Wilson, R (1975) On the Theory of Aggregation, *J Econ Theory* 10: 89-99