

Judgment aggregation: (im)possibility theorems

Franz Dietrich¹

December 2003

forthcoming in *Journal of Economic Theory*

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. *JEL 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";

$a \rightarrow b$: "if we can afford a budget deficit *then* spending on education should be raised".

	a	$a \rightarrow b$	b
Camp 1	Yes	Yes	Yes
Camp 2	Yes	No	No
Camp 3	No	Yes	No
Majority	Yes	Yes	No

Table 1: Collective inconsistency under majority voting

¹I wish to thank Luc Bovens, Christian List, Jocelyn Paine, Marc Pauly and Martin van Hees, for very inspiring discussions. I am also grateful to 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* (University of Konstanz, June 2004).

The cabinet is split into three camps of equal size. As indicated in Table 1, 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 the implication $a \rightarrow b$, and the third camp accepts the implication $a \rightarrow b$ 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 (*conclusion-based voting*) or rather accept b on the grounds of accepting both premises of b (*premise-based voting*)?

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 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, \dots (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 $A \models p$, 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

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

²Specifically, 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.

The example had $X = \{a, b, a \rightarrow b, \text{negations}\}$ ("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* is an n -tuple (A_1, \dots, A_n) of (individual) judgment sets. A (*judgment*) *aggregation rule* is a function, F , that assigns to each profile (A_1, \dots, A_n) (in some set of admissible profiles called the domain of F and written $Dom(F)$) a (collective) judgment set $F(A_1, \dots, A_n) = A \subseteq X$. For instance, propositionwise majority voting (with universal domain \mathbf{A}^n) is the aggregation rule F such that, for each profile $(A_1, \dots, A_n) \in \mathbf{A}^n$, $F(A_1, \dots, A_n)$ 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, $F(A_1, \dots, A_n)$ may then be inconsistent, hence not in \mathbf{A} .

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 $A \models p$ or $A \models \neg p$. 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 examples³, 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.

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, \dots, A_n) , $(A'_1, \dots, A'_n) \in \text{Dom}(F)$, if [for every person i , $A_i \models p$ if and only if $A'_i \models p$] then $[F(A_1, \dots, A_n) \models p$ if and only if $F(A'_1, \dots, A'_n) \models p]$.

Here, I interpret " $A_i \models p$ " and " $F(A_1, \dots, A_n) \models p$ " as "acceptance of p ", even when this acceptance is not expressed explicitly (i.e. no " $\exists p$ ") but only entailed logically. Note that $A \models p$ 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 $M : \{0, 1\}^n \rightarrow \{0, 1\}$, taking vectors (t_1, \dots, t_n) of (individual) truth values to single (collective) truth values $M(t_1, \dots, t_n)$ (where 0/1 stands for rejection/acceptance of a given proposition). For instance, the *absolute majority method* M is defined by $[M(t_1, \dots, t_n) = 1$ if and only if $t_1 + \dots + t_n > n/2]$, and the *unanimity method* by $[M(t_1, \dots, t_n) = 1$ if and only if $t_1 = \dots = t_n = 1]$. I say that F "applies decision method M for p " if, for every profile $(A_1, \dots, A_n) \in \text{Dom}(F)$, we have $t = M(t_1, \dots, t_n)$, where t_1, \dots, t_n and t are the individual and collective truth values of p (i.e., t_i is 1 if $A_i \models p$ and 0 else, and t is 1 if $F(A_1, \dots, A_n) \models p$ 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 .

³For 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" $\{\bigwedge_{p \in A} p, \neg \bigwedge_{p \in A} p \mid A \in \mathbf{A}\}$, e.g. modifying $\{a, b, \text{negations}\}$ into $\{a \wedge b, (\neg a) \wedge b, a \wedge (\neg b), (\neg a) \wedge (\neg b), \text{negations}\}$.

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 $Y = \tilde{X} := \{\bigwedge_{p \in AP} : A \in \mathbf{A}\}$.*

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 $\bigwedge_{p \in AP} (A \in \mathbf{A})$ 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 $\tilde{X} = \{a \wedge b, (\neg a) \wedge b, a \wedge (\neg b), (\neg a) \wedge (\neg b)\}$.

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.

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. ■

⁴Specifically, Pauly and van Hees require that, for every $p \in Y$ and $(A_1, \dots, A_n), (A'_1, \dots, A'_n) \in Dom(F)$, if [for every person i , $p \in A_i$ if and only if $p \in A'_i$] then [$p \in F(A_1, \dots, A_n)$ if and only if $p \in F(A'_1, \dots, A'_n)$]. 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 $A \models p$] for all $p \in X$ and $A \in \mathbf{A}$.

⁵For *infinite* X , the conjunction $\bigwedge_{p \in AP}$ 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 $\bigwedge_{p \in AP}, A \in \mathbf{A}$, 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 $(I_{\tilde{X}})$ may be considered even for infinite agendas X .

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 $(I_{\widetilde{X}})/(I_{\widetilde{X^*}})$ 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 $(I_{\overline{X}})/(I_{\overline{X^*}})$ 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 $A^* \models p$ implies $A \models p$, and $A^* \models \neg p$ implies $A \models \neg p$. I only show the former, as the proof of the latter is analogous. So, let $A^* \models p$. 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 , $[t_i \leq t_i^* \text{ for all } i \text{ implies } M_q(t_1, \dots, t_n) \leq M_q(t_1^*, \dots, t_n^*)]$, and, for fixed t_1, \dots, t_n , $[q^* \models q \text{ implies } M_{q^*}(t_1, \dots, t_n) \leq M_q(t_1, \dots, t_n)]$.

Take any $p^* \in \overline{X^*}$ with $A^* \models p^*$ and $p^* \models p$ (e.g. $p^* = \bigwedge_{q \in A^*} q$). 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 $p^* \models p$ and (1), we have $t_i^* \leq t_i$ for all i . Since also $p^* \models p$, by (4) $M_{p^*}(t_1^*, \dots, t_n^*) \leq M_p(t_1, \dots, t_n)$. By $A^* \models p^*$ we have $M_{p^*}(t_1^*, \dots, t_n^*) = 1$, so $M_p(t_1, \dots, t_n) = 1$, so $A \models p$. ■

Proof of claim D. [assuming (1),(2)]. Suppose F violates (I_X) (the proof for $(I_{\overline{X}})$ 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's⁶ 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 .

⁶F. Dietrich and C. List, Strategy-Proof Judgment Aggregation, unpublished paper, Konstanz Univ., 2004.

⁷More precisely, I call F *strategy-proof on Y* ($\subseteq \overline{X}$) if, for every person i , profile $(A_1, \dots, A_n) \in Dom(F)$ and proposition $p \in Y$, if A_i disagrees with $F(A_1, \dots, A_n)$ on p (i.e. $A_i \models p$ if and only if $F(A_1, \dots, A_n) \not\models p$), then A_i still disagrees with $F(A_1, \dots, A_i^*, \dots, A_n)$ on p for every i -variant $(A_1, \dots, A_i^*, \dots, A_n) \in Dom(F)$. A game-theoretic justification for this definition may be given along the lines of Dietrich and List's analysis.

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 $\mathbf{A}^n = \mathbf{A} \times \dots \times \mathbf{A}$ of all logically possible profiles of complete and consistent individual judgment sets.

Collective Rationality (C). For any profile $(A_1, \dots, A_n) \in Dom(F)$, $F(A_1, \dots, A_n) \in \mathbf{A}$.

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 $(A_1, \dots, A_n), (A'_1, \dots, A'_n) \in Dom(F)$ such that $F(A_1, \dots, A_n) \neq F(A'_1, \dots, A'_n)$.

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") $F(A_1, \dots, A_n) = A_j$ for all profiles $(A_1, \dots, A_n) \in \mathbf{A}^n$.

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.*

As strong independence implies independence on states of the world, 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 $(I_{\tilde{X}})$ holds if and only if, for every $A \in \mathbf{A}$ and $(A_1, \dots, A_n), (A'_1, \dots, A'_n) \in Dom(F)$, if [for every person i , $A_i = A$ if and only if $A'_i = A$] then $[F(A_1, \dots, A_n) = A$ if and only if $F(A'_1, \dots, A'_n) = A]$.*

Proof. Obvious, as a judgment set in \mathbf{A} entails $\bigwedge_{p \in AP} (p \in \tilde{X})$ just in case it equals A . ■

Judgment-Set Monotonicity (JM). For any person j and any j -variants $(A_1, \dots, A, \dots, A_n), (A_1, \dots, A', \dots, A_n) \in Dom(F)$, if $F(A_1, \dots, A, \dots, A_n) = A'$ then $F(A_1, \dots, A', \dots, A_n) = A'$.

Lemma 2 *Let X contain at least two propositions (not in trivial dependence). If F satisfies (U), (C) and $(I_{\tilde{X}})$, then F satisfies (JM).*

Proof. Let X be as specified, and suppose (U), (C) and $(I_{\tilde{X}})$. To show (JM), let j be a person and $(\dots, A, \dots), (\dots, A', \dots) \in \text{Dom}(F)$ be j -variants, where " \dots " denotes the other persons' votes. Assume for contradiction that $F(\dots, A, \dots) = A'$ but $F(\dots, A', \dots) \neq A'$. In (\dots, A, \dots) and (\dots, A', \dots) exactly the same persons endorse each $A'' \in \mathbf{A} \setminus \{A, A'\}$; hence, as $F(\dots, A, \dots) \neq A''$, we have $F(\dots, A', \dots) \neq A''$ by Lemma 1, so $F(\dots, A', \dots) \in \{A, A'\}$, hence $F(\dots, A', \dots) = A$. By $|\mathbf{A}| \geq 3$ there exists an $A'' \in \mathbf{A} \setminus \{A, A'\}$. Consider the new j -variant (\dots, A'', \dots) . I apply twice Lemma 1, with contradictory implications: as $F(\dots, A, \dots) = A'$ and as in (\dots, A, \dots) and (\dots, A'', \dots) exactly the same persons endorse A' (in neither profile person j), $F(\dots, A'', \dots) = A'$; but, as $F(\dots, A', \dots) = A$ and as in (\dots, A', \dots) and (\dots, A'', \dots) , exactly the same persons endorse A (in neither profile person j), $F(\dots, A'', \dots) = A$. ■

Judgment-Set Unanimity Principle (JUP). $F(A, \dots, A) = A$ for all $(A, \dots, A) \in \text{Dom}(F)$.

Lemma 3 *Let X contain at least two propositions (not in trivial dependence). If F satisfies (U), (C), $(I_{\tilde{X}})$ and (R), then F satisfies (JUP).*

Proof. Let X be as specified, and assume (U), (C), $(I_{\tilde{X}})$ and (R). To show (JUP), consider any $A \in \mathbf{A}$, and suppose for contradiction that $F(A, \dots, A) \neq A$. I show that $F(A'_1, \dots, A'_n) = F(A, \dots, A)$ for all $(A'_1, \dots, A'_n) \in \mathbf{A}^n$, violating (R). Take any $(A'_1, \dots, A'_n) \in \mathbf{A}^n$ and write $A' := F(A'_1, \dots, A'_n)$. By (JM) (see Lemma 2), if the votes A'_1, \dots, A'_n are replaced one by one by A' , the decision remains A' , and so $F(A', \dots, A') = A'$. In (A', \dots, A') and (A, \dots, A) exactly the same persons (namely nobody) endorse each $A'' \in \mathbf{A} \setminus \{A, A'\}$; hence, as $F(A', \dots, A') \neq A''$, we have $F(A, \dots, A) \neq A''$ (see Lemma 1). So $F(A, \dots, A) \in \{A, A'\}$. As $F(A, \dots, A) \neq A$, we have $F(A, \dots, A) = A'$, i.e. $F(A, \dots, A) = F(A'_1, \dots, A'_n)$, as claimed. ■

Proof of Theorem 1. Let X be as specified. If F is dictatorial then F obviously satisfies all of (U), (C), $(I_{\tilde{X}})$ and (R). Now I assume (U), (C), $(I_{\tilde{X}})$ 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), $F(A, \dots, A) = A$. Modify (A, \dots, 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 $F(A'', \dots, A'') = A$, 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, $(A'', \dots, A'', \underline{A}, A, \dots, A)$ and $(A'', \dots, A'', \underline{A'}, A, \dots, A)$, exactly the same persons endorse each $A^* \in \mathbf{A} \setminus \{A, A'\}$; hence, as $F(A'', \dots, A'', \underline{A}, A, \dots, A) \neq A^*$, we have $F(A'', \dots, A'', \underline{A'}, A, \dots, A) \neq A^*$ (see Lemma 1). So $F(A'', \dots, A'', \underline{A'}, A, \dots, A) \in \{A, A'\}$. As $F(A'', \dots, A'', \underline{A'}, A, \dots, A) \neq A$, we have $F(A'', \dots, A'', \underline{A'}, A, \dots, A) = A'$, although here j is the only person to vote A' . To show that j is a dictator for A' , consider any profile $(A_1, \dots, A_{j-1}, A', A_{j+1}, \dots, A_n)$ in which j votes A' . The one-by-one substitution in $(A'', \dots, A'', \underline{A'}, A, \dots, A)$ of the votes of persons $i \neq j$ by their respective votes in $(A_1, \dots, A_{j-1}, \underline{A'}, A_{j+1}, \dots, A_n)$ leaves the outcome at A' , by (JM) if $A_i = A'$ and by Lemma 1 if $A_i \neq A'$. So $F(A_1, \dots, A_{j-1}, \underline{A'}, A_{j+1}, \dots, A_n) = A'$.

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 (I_X) implies ($I_{\tilde{X}}$) 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 $q \in X$, $p \models q$ or $p \models \neg q$. 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 ($X = \bar{X}$);
- (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 ($I_{\tilde{X}}$). This holds if every state of the world $q \in \tilde{X}$ is logically equivalent to some atom r of X . Consider any $q = \bigwedge_{p \in AP} p \in \tilde{X}$ ($A \in \mathbf{A}$). Let B be the set of all atoms of X consistent with q . B is non-empty, since otherwise $q \models \neg r$ 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 $q = \bigwedge_{p \in AP} p \models r$. Also, $r \models q$. Otherwise r would be consistent with $\neg q$, hence with $\neg p$ for some $p \in A$, so that $r \models \neg p$ for this p (since q is an atom), and hence $p \models \neg r$, in contradiction with $\bigwedge_{p \in AP} p \models r$. ■

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 ($I_{\tilde{X}}$) are logically related – of which I otherwise have little intuition except that both are of course weaker than strong independence ($I_{\bar{X}}$).

Corollary 3 *If (U) and (C) hold, ($I_{\tilde{X}}$) 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 $(I_{\tilde{X}})$ and (I_X) hold). If F satisfies $(I_{\tilde{X}})$, 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 $(I_{\tilde{X}})$. ■

6 A possibility theorem on premise-based decision-making

Despite their merits in preventing manipulation, there are good reasons to reject the independence conditions $(I_X)/(I_{\tilde{X}})/(I_{\overline{X}})$. 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 $(A_1, \dots, A_n) \in \mathbf{A}^n$, $F(A_1, \dots, A_n) = \{p \in X : P^* \models p\}$, where $P^* := \{p \in P : n_p > n_{\sim p}, \text{ or } [n_p = n_{\sim p} \text{ and } p \text{ is a negated proposition}]\}$ with n_p denoting the number of persons i with $p \in A_i$.*

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 $(A_1, \dots, A_n), (A_{\pi(1)}, \dots, A_{\pi(n)}) \in \text{Dom}(F)$, where $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ is any permutation of the individuals, $F(A_1, \dots, A_n) = F(A_{\pi(1)}, \dots, A_{\pi(n)})$.

Monotonicity on Y (M_Y). For each proposition $p \in Y$, individual i and i -variants $(A_1, \dots, A_n), (A_1, \dots, A_i^*, \dots, A_n) \in \text{Dom}(F)$ with $A_i \neq p$ and $A_i^* \models p$, if $F(A_1, \dots, A_n) \models p$ then $F(A_1, \dots, A_i^*, \dots, A_n) \models p$.

Systematicity on Y (S_Y). For every two propositions $p, p' \in Y$ and every two profiles $(A_1, \dots, A_n), (A'_1, \dots, A'_n) \in \text{Dom}(F)$, if [for every person i , $A_i \models p$ if and only if $A'_i \models p'$], then $[F(A_1, \dots, A_n) \models p \text{ if and only if } F(A'_1, \dots, A'_n) \models p']$.

(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:

⁸D. Osherson, Notes on Aggregating Belief, unpublished paper, Princeton University, 2004.

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, $\{a, b, c, c \leftrightarrow (a \wedge b), \text{negations}\}$, then $P := \{a, b, c \leftrightarrow (a \wedge b), \text{negations}\}$ 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 $(A_1, \dots, A_n) \in \mathbf{A}^n$, 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 $F(A_1, \dots, A_n) = \{p \in X : P^* \models p\}$ is consistent.

(ii) Assume $X \subseteq \overline{P}$. For each $(A_1, \dots, A_n) \in \mathbf{A}^n$, 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 $F(A_1, \dots, A_n) = \{p \in X : P^* \models p\}$ 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), $M(t_1, \dots, t_n)$ depends only on the number of persons i with $t_i = 1$, i.e. there exists a function $g : \{0, \dots, n\} \rightarrow \{0, 1\}$ such that, for all $(A_1, \dots, A_n) \in \mathbf{A}^n$ and $p \in P$, $[p \in F^*(A_1, \dots, A_n) \text{ if and only if } g(|N_p|) = 1]$, where $N_p := \{i : p \in A_i\}$. By (M_P) and (U), $g(k) \leq g(k+1)$ for all $k \in \{0, \dots, n-1\}$. Hence, by induction, (a) $k < l$ implies $g(k) \leq g(l)$, for all $k, l \in \{0, \dots, n\}$. 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 $(A_1, \dots, A_n) \in \mathbf{A}^n$, and so (b) $g(k) + g(n-k) = 1$ for all $k \in \{0, \dots, n\}$. For, as (A_1, \dots, A_n) runs through \mathbf{A}^n , $|N_p|$ runs through $\{0, \dots, n\}$ and always takes the value $n - |N_{\neg p}|$. 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 $(A_1, \dots, A_n) \in \mathbf{A}^n$, $F^*(A_1, \dots, A_n) \cap P = F(A_1, \dots, A_n) \cap P =: A^*$. 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^* , $F^*(A_1, \dots, A_n) = \{p \in X : A^* \models p\}$, 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 strategy-proofness 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.