Factoring Out the Impossibility of Logical Aggregation*

Philippe Mongin†

June 2005, revised October 2006

Abstract

According to a theorem recently proved in the theory of logical aggregation, any nonconstant social judgment function that satisfies independence of irrelevant alternatives (IIA) is dictatorial. We show that the strong and not very plausible IIA condition can be replaced with a minimal independence assumption plus a Pareto-like condition. This new version of the impossibility theorem likens it to Arrow's and arguably enhances its paradoxical value.

1 Introduction

In political science and legal theory, the so-called *doctrinal paradox* refers to the observation that if a group of voters casts separate ballots on each proposition of a given agenda, and the majority rule is applied to each of these votes separately, the resulting set of propositions may be logically inconsistent. A mathematical theory of *logical judgment aggregation* has recently grown out of this straightforward point. Its method is to introduce a mapping from profiles of individual judgments to social judgments, where

*The author thanks for comments and reactions B. Hill, F. Dietrich, M. Fleurbaey, R. Holzman, C. List, K. Nehring, M. Pauly, as well as the participants in the LGS4 conference, the Cowles Foundation Workshop on Aggregation of Opinions, and seminars at HEC, MIT, and Brown University.

†Centre National de la Recherche Scientifique (CNRS) & Ecole des Hautes Etudes Commerciales (HEC), 1 rue de la Libération, F-78350 Jouy-en-Josas. E-mail: mongin@hec.fr

Printed from: *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz.* Ed. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonsson, 2007. www.fil.lu.se/hommageawlodek

these judgments are formalized as sets of formulas in some logical language, and then investigate the effect of imposing axiomatic conditions on this mapping. Among the results obtained is a striking impossibility theorem that abstractly generalizes the doctrinal paradox (Pauly and van Hees, forthcoming; Dietrich, 2006). This theorem states that a mapping defined on a universal domain is dictatorial - i.e., collapses the social judgment set into the set of a given individual whatever the profile - if and only if it is nonconstant and satisfies *independence of irrelevant alternatives* (IIA). In the present context, the latter condition says that if a formula enters the judgment sets of exactly the same individuals in two profiles, it is a member of either both or none of the two social sets obtained from the mapping. More informally, one can decide whether a formula belongs to a social set just by considering which individual sets it belongs to, regardless of the other formulas that these sets may contain. An earlier variant relied on a stronger condition that is intuitively related to the neutrality axiom of social choice theory (List and Pettit, 2002).

Given the Arrovian undertones of the impossibility theorem, there is something puzzling about it. Arrow's social welfare function maps any conceivable profile of individual weak orderings to a social weak ordering, and his IIA says in effect that one can decide what the social preference is between two alternatives just by considering the individual preferences over these alternatives. The two aggregative schemes are closely related, and so are the two independence conditions. However, Arrow's conclusion that the social welfare condition is dictatorial - in the sense of reproducing one individual's strict preference - depends not only on IIA, but on a unanimity condition. The standard proof based on Arrow (1963) uses the weak Pareto condition; Wilson's (1972) extension dispenses with it but needs a premiss to exclude antidictatorship, i.e., the rule which amounts to reversing one individual's strict preference. The impossibility of logical judgment aggregation seems to obtain without any assumption of the kind. This should come as a surprise to anybody acquainted with the social-choice theoretic tradition, not the least to theoretical economists.

The present paper offers a new theorem that will make the impossibility conclusion less mysterious. Still granting universal domain, it derives dictatorship from a IIA condition that is restricted to the atomic components of the language, hence much weaker than the existing one, plus a unanimity condition without which the conclusion does not follow. Both the earlier result and ours deliver essentially equivalent restatements of dictatorship;

so what we do in effect is to *factor out* the strong IIA condition of the current theory. The proposed set of premisses appears to be preferable for two reasons. Firstly, unanimity-preservation and independence are conceptually different properties for a social aggregate. If one is hidden under the guise of the other, this can only be due to the peculiarities of the propositional logic framework. Elsewhere, we compare the impossibility of logical aggregation with related probabilistic impossibilities, and show that unanimity and independence cannot be blended together in that framework (Mongin, 2005). Secondly, at the normative level, the IIA condition is hardly acceptable, whereas, as will be explained, restricting it to atomic components deflates a telling objection against it, hence puts it on a safer basis. In the present version, dictatorship arises from one's combining the two separate ideas of independence (weakly stated) and unanimity-preservation (unrestricted). This suggests that the latter condition may be blamed for the impossibility. In the literature, the claim is sometimes made that the aggregation procedure should be *premiss-based* rather than *conclusion-based*, i.e., that individual judgments should be aggregated on a strict subset of formulas (the premisses), while the social judgment would be reached on the other formulas (the conclusions) by an inference step performed within the social set. Our theorem warrants this claim with some reservations that we will indicate (one of them relates to the novel argument for the conclusion-based procedure made by Bovens and Rabinowicz, 2006).

The new theory defines the *agenda* to be the set of logical formulas representing the propositions, or issues, on which the individuals and society take sides. The factoring out theorem is obtained at the price of slightly extending the agenda beyond what is required in order to get dictatorship from the current IIA. The paper will relate this special restriction to the recently developed criteria for recognizing a dictatorial agenda (Dokow and Holzman, 2005 ; Nehring and Puppe, 2005).

2 The logical framework of aggregation

In the logical framework, a judgment consists in either accepting or rejecting a formula stated in some logical language. Like most previous writers, we will be concerned with the language of propositional logic. Accordingly, our set of formulas $\mathcal{L}$ is constructed from the propositional connectives $\neg$,

$\vee$, $\wedge$, $\rightarrow$, $\leftrightarrow$ ("not", "or", "and", "implies", "is equivalent to") and a set $\mathcal{P}$ of distinct propositional variables (p.v.) $p_1, ..., p_m, ...$ serving as the atomic components of the construction. We need no finiteness restriction, but lower bounds will be considered. When we say that a p.v. p *appears* in a formula φ, we mean that either p or $\neg p$ is a subformula of φ. *Literals* are those formulas which are either p.v. or negations of p.v.; a *literal value* for $p \in \mathcal{P}$ is a choice between p or $\neg p$. In shorthand notation, $\widetilde{p}$ means either p or $\neg p$, and $\neg\widetilde{p}$ means $\neg p$ in the first case and p in the second; when we wish to emphasize that the literal value is fixed, we write $\overline{p}$ instead of $\widetilde{p}$.

The axiomatic system of propositional logic defines a inference relation holding between formulas; further notions, like that of a consistent set of formulas, are derived in the ordinary way. Beyond propositional logic itself, the framework covers those more expressive languages in which it can be embedded isomorphically, for instance the modal propositional logics that have become known to game theorists (see Bacharach et alii, 1997). A translation of Theorem 1 to any of these extensions is an mechanical affair, but it must be stressed that it does not apply beyond the fragment of the richer language that is isomorphic to propositional logic.

The *agenda* is the nonempty subset $\Phi \subseteq \mathcal{L}$ of formulas representing the actual propositions on which the n individuals and society pass a positive or negative judgment. Define $\Phi_0 = \Phi \cap \mathcal{P}$. For any $p, q \in \Phi_0$, we say that $\widetilde{p}$, $\widetilde{q}$ are *connected in terms of a k-disjunction of literals* if Φ contains some disjunction of k disjuncts, among which $\widetilde{p}$ and $\widetilde{q}$, the other disjuncts - if there are any - being also literals. To illustrate, $\widetilde{p}$ and $\widetilde{q}$ are connected in terms of a 2-disjunction iff Φ contains $\widetilde{p} \vee \widetilde{q}$, and in terms of a 3-disjunction if Φ contains $\widetilde{q} \vee \widetilde{r} \vee \widetilde{p}$ for some literal $\widetilde{r}$.

CLOSURE CONDITIONS ON Φ: (i) Closure Under Propositional Variables: if $\varphi \in \Phi$, and $p \in \mathcal{P}$ appears in φ, then $p \in \Phi_0$. (ii) Limited Disjunctive Closure: in every 3-element subset of Φ_0, there is an element p such that each literal value of p is connected in terms of 2- or 3-disjunctions with each literal value of the other two elements, q and r. (iii) The previous condition holds with at least one 3-disjunction, i.e., if $\{p, q, r\} \subseteq \Phi_0$, there is at least one choice of literals $\widetilde{p}$, $\widetilde{q}$, $\widetilde{r}$ for which $\widetilde{p} \vee \widetilde{q} \vee \widetilde{r} \in \Phi$.

An agenda is *minimal* among those satisfying the Closure Conditions if it does not include any other agenda belonging to this class. We illustrate by describing the minimal agendas for $|\Phi_0| = 3$. At one extreme are the agendas containing a single 3-disjunction:

$$\mathcal{A}^{-}_{223} = \{all\ \widetilde{p}, \widetilde{q}, \widetilde{r}\} \cup \{\overline{p} \vee \overline{q} \vee \overline{r}\} \cup \{\widetilde{p} \vee \widetilde{q},\ \widetilde{p} \vee \widetilde{r}\ for\ all\ \widetilde{p}, \widetilde{q}, \widetilde{r}\} \setminus \{\overline{p} \vee \overline{q},\ \overline{p} \vee \overline{r}\}\,.$$

At the other extreme are the agendas containing only 3-disjunctions:

$$\mathcal{A}^{-}_{33} = \{all\ \widetilde{p}, \widetilde{q}, \widetilde{r}\} \cup \{\overline{p} \vee \overline{q} \vee \overline{r},\ \overline{p} \vee \neg\overline{q} \vee \neg\overline{r},\ \neg\overline{p} \vee \overline{q} \vee \overline{r},\ \neg\overline{p} \vee \neg\overline{q} \vee \neg\overline{r}\}\,.$$

In between are various agendas $\mathcal{A}^{-}_{2233}$ containing several 3-disjunctions as well as several 2-disjunctions. As the subscript is meant to convey, each minimal agenda corresponds to a trade-off between the number and complexity of disjunctions.
Here is an example of a non-minimal agenda for $|\Phi_0| \geq 3$:

$$\mathcal{A}_K = \{all\ \widetilde{p_1}, ..., \widetilde{p_m}, ...\} \cup \{\widetilde{p_{l_1}} \vee ... \vee \widetilde{p_{l_K}}\ for\ all\ K\text{-}sequences\ \widetilde{p_{l_1}}, ..., \widetilde{p_{l_K}}\}\,,$$

where K is some fixed number ($K \geq 3$).

The theory investigates aggregative rules for judgment *sets*, where definite restrictions are imposed on what counts as such a set. Technically, it is any nonempty subset $B \subseteq \mathcal{L}$ that is *consistent*, as well as *maximal* in the following relative sense: for any $\varphi \in \Phi$, either φ or $\neg\varphi$ belongs to B. Define $\Phi^* = \Phi \cup \{\neg\varphi : \varphi \in \Phi\}$. The consistency and relative maximality of judgment sets imply the weaker property that judgment sets B are *deductively closed* relative to Φ^*, i.e., for all $\varphi \in \Phi^*$ and $C \subset B$, if φ can be inferred from C, then $\varphi \in B$. From this property, it follows that equivalent formulas may be identified with each other. In particular, we will from now on assume that double negations cancel and that a disjunction remains the same, whatever the order of disjuncts, and whether some of them are repeated or not. Since the definition of $\mathcal{A}_K$ allows for repeated literals, all types of minimal agendas for $|\Phi_0| = 3$ are included in $\mathcal{A}_3$.

Given some set D of judgment sets, a *social judgment function* (s.j.f.) is a mapping

$$F : D^n \to D, \quad (A_1, ..., A_n) \mapsto A.$$

We put $N = \{1, ..., n\}$ and often write A, A', ... instead of $F(A_1, ..., A_n)$, $F(A'_1, ..., A'_n)$, ... A s.j.f. is *dictatorial* if there is $j \in N$ - the *dictator* - such that for all $(A_1, ..., A_n) \in D^n$,

$$A_j = F(A_1, ..., A_n).$$

Equivalently, a dictatorial s.j.f. is a projection of the product D^n on one of its components. (Incidentally, Arrow's dictatorship does not have the projection property, since it retains only the strict part of the dictator's ordering.)

The lemmas and proofs below will involve the further idea of a local dictator, i.e., of an individual who dictates on some part of the agenda. We define j to be a *dictator on* Φ_0 if

$$\forall (A_1, ..., A_n) \in D^n, \forall p \in \Phi_0, \quad p \in F(A_1, ..., A_n) \Leftrightarrow p \in A_j.$$

In order to define a dictator on $\Phi_0^- \subset \Phi_0$, we fix literal values $\overline{s}$ for each $s \in \Phi_0 \setminus \Phi_0^-$. The condition that

$$B \in D^- \Longleftrightarrow (\forall s \in \Phi_0 \setminus \Phi_0^-) \quad \overline{s} \in B$$

defines a subset of judgment sets D^-, hence a subdomain $(D^-)^n$ of F, in which the individuals can disagree only in terms of some $p \in \Phi_0^-$. We say that D^- and $(D^-)^n$ are *determined* by Φ_0^-; there are as many choices of these sets as there are sets of $\overline{s}$ values. We declare j to be a *dictator on* Φ_0^- if for *all* subdomains $(D^-)^n$ determined by Φ_0^-,

$$\forall (A_1, ..., A_n) \in (D^-)^n, \forall p \in \Phi_0^-, \quad p \in F(A_1, ..., A_n) \Leftrightarrow p \in A_j.$$

We now introduce axiomatic conditions on F. It will be a maintained assumption that F satisfies *Universal Domain*, i.e., that D is the set of all logically possible judgment sets.

Axiom 1 *(Systematicity)*

$$\begin{array}{rcl} \forall \varphi, \psi & \in & \Phi, \forall (A_1, ..., A_n), (A'_1, ..., A'_n) \in D^n \\ [\varphi \in A_i \Leftrightarrow \psi \in A'_i, i = 1, ..., n] & \Rightarrow & [\varphi \in A \Leftrightarrow \psi \in A'] . \end{array}$$

Axiom 2 *(Independence of Irrelevant Alternatives)*

$$\begin{array}{rcl} \forall \varphi & \in & \Phi, \forall (A_1, ..., A_n), (A'_1, ..., A'_n) \in D^n \\ [\varphi \in A_i \Leftrightarrow \varphi \in A'_i, i = 1, ..., n] & \Rightarrow & [\varphi \in A \Leftrightarrow \varphi \in A'] . \end{array}$$

Axiom 3 *(Independence of Irrelevant Propositional Alternatives)*

$$\begin{array}{rcl} \forall p & \in & \Phi_0, \forall (A_1, ..., A_n), (A'_1, ..., A'_n) \in D^n \\ & & [p \in A_i \Leftrightarrow p \in A'_i, i = 1, ..., n] \Rightarrow [p \in A \Leftrightarrow p \in A'] . \end{array}$$

The three conditions are listed from the logically strongest to the weakest. *Systematicity* requires that two formulas be treated alike by society if they draw the support of exactly the same people, even if these formulas refer to semantically unrelated items. Although this was the condition assumed in the first place, it is quite obviously unattractive. Take a two-individual society in which 1 judges that the budget should be balanced, 2 disagrees, and the social judgment endorses 1. Then, if 1 also judges that marijuana should be legalized, and 2 disagrees again, the social judgment should endorse 1 again. Economists will recognize that this is a neutrality condition in the style of those of social choice theory and that it is no more appealing here than it is there. Samuelson (1977) once described neutrality as transparently close to dictatorship and gratuitous; it is instructive to recall the witty example he devised to reject this condition.

Instead of permitting variations in both the profile $(A_1, ..., A_n)$ and formula φ, *Independence of Irrelevant Alternatives* fixes the formula and allows only the profile to vary; in this way, it avoids the confounding of semantic contents that spoils the earlier condition. It singles out the requirement contained in Systematicity that the social judgment on φ should depend only on the individual judgments on φ. Exactly as for Arrow's condition, the best normative defence for this restriction is that it prevents some possible manipulations (see Dietrich, 2006, and Dietrich and List, 2004). However, the condition remains open to a charge of irrationality. One would expect society to pay attention not only to the individuals' judgments on φ, but also to their *reasons* for accepting or rejecting this formula, and these reasons may be represented by other formulas than φ in the individual sets. Before deciding that two profiles call for the same acceptance or rejection, society should in general take into account more information than is supposed in the condition.

The new condition of *Independence of Irrelevant Propositional Alternatives* (IIPA) amounts to reserving IIA to propositional variables alone. In the doctrinal paradox, trouble arises from the assumption that the majority rule applies to molecular formulas and propositional variables alike - i.e., that this independent and even neutral rule dictates on the whole of Φ. In contrast, when restricted to propositional variables, independence becomes more acceptable *because these formulas represent primary data.* One can object to IIA being applied to $p \vee q$, where p represents "The budget should be balanced" and q "Marijuana should be legalized" on the ground that there are two propositions involved, and that society should

know how each individual feels about either of them, and not simply about their disjunction. No similar objection arises when IIA is applied to either p or q in isolation because the reasons for accepting or rejecting them are beyond the expressive possibilities of the language. Of course, as pointed out earlier, the language of propositional logic can be embedded in more powerful ones that will analyze its primary data. In such a refined framewok, the irrationality charge would carry through to the formulas replacing p or q, but would be similarly deflated when one reaches the stage of the building blocks.

The last condition to be introduced is a straightforward analogue of the Pareto principle:

Axiom 4 *(Unanimity Preservation, UP) For all* $\varphi \in \Phi$, *and all* $(A_1, ..., A_n) \in D^n$,

$$\varphi \in A_i, i = 1, ..., n \Rightarrow \varphi \in A.$$

3 The impossibility of logical judgment aggregation

Whatever the formulation of the impossibility theorem, if one simply replaces IIA by IIPA in the assumptions, the dictatorial conclusion vanishes. This is confirmed by considering the following non-dictatorial rule, which is well-defined for any odd value of n:

(R) *Apply majority voting to the p.v. of the agenda, and close the resulting set of literals by logical inference.*

Any A thus obtained from $(A_1, ..., A_n)$ is consistent, and if Closure Condition (i) is met, it is also maximal relative to the agenda, so **(R)** defines a s.j.f., which satisfies IIPA by construction.

However, when IIPA is combined with UP, dictatorship reappears.

Theorem 1 *Assume that* $|\Phi_0| \geq 3$ *and that the Closure Conditions (i), (ii), and (iii) hold. If F satisfies Independence of Irrelevant Propositional Alternatives and Unanimity Preservation, F is dictatorial.*

The theorem trivially holds if $n = 1$, and trivially does not hold if $n \geq 2$, $|\Phi_0| = 1$. We show that if $n \geq 3$, it does not hold of $|\Phi_0| = 2$. The counterexample is **(R)**, which happens to satisfy UP when Φ_0 consists of

two distinct p, q. To see that, observe that from Closure Condition (ii), there will be only two forms of $\varphi \in \Phi$ up to logical equivalence, i.e., either $\varphi = \widetilde{p} \vee \widetilde{q}$ or $\varphi = \neg(\widetilde{p} \vee \widetilde{q})$, where p and q are not necessarily distinct. Assume that $\varphi = \widetilde{p} \vee \widetilde{q} \in A_i$, $i = 1, ..., n$. Given Closure Condition (i) and the properties of judgment sets, there must be either a majority for $\widetilde{p}$ or a majority for $\widetilde{q}$; so the conclusion that $\widetilde{p} \vee \widetilde{q} \in A$ follows from **(R)** and deductive closure of A. The case $\varphi = \neg(\widetilde{p} \vee \widetilde{q})$ is handled similarly.

Now, suppose that $n \geq 3$, $|\Phi_0| \geq 3$. The counterexample would still apply if Φ were restricted to 2 -disjunctions, which shows that the size of disjunctions, and not only the number of p.v., matters to the truth of the theorem. At this juncture, Closure Condition (iii) comes into play. Take three distinct $p, q, r \in \Phi_0$, suppose that (iii) is met with $p \vee q \vee r \in \Phi$, and consider the profiles in which $p \vee q \vee r \in A_i$, $i = 1, ..., n$. There need no longer be a majority for one of the disjuncts. Indeed, for $n = 3$, it may well be that:

$$p, \neg q, \neg r \in A_1;\ \neg p, q, \neg r \in A_2;\ \neg p, \neg q, r \in A_3,$$

so that all of $\neg p, \neg q$, and $\neg r$ obtain a majority. From **(R)** and the consistency of A, $p \vee q \vee r \notin A$, which negates the conclusion of UP. A profile like this one illustrates Theorem 1 at work; indeed, since the s.j.f. defined by **(R)** is non-dictatorial, it cannot satisfy UP on top of IIPA.

There remains a limiting case to consider: $n = 2 = |\Phi_0|$. Perhaps surprisingly, the theorem holds in this case, which can be checked both from the proof below and the simpler argument given at the end of the section.

The following lemma is key to the proof.

Lemma 2 *Suppose that for all subsets $\Phi_0^- \subset \Phi_0$ of cardinality k, where $k \geq 2$, there is a dictator on Φ_0^-. Then, there is a dictator on Φ_0.*

Proof. We first show that if Φ_0^-, $\Psi_0^- \subseteq \Phi_0$ are non-disjoint, and there is a dictator i^* on Φ_0^- and a dictator j^* on Ψ_0^-, then $i^* = j^*$.

Supposing the contrary, we can find a subdomain of profiles $(A_1, ..., A_n)$ which is determined by $\Phi_0^- \cup \Psi_0^-$, and s.t. for some $\overline{q} \in \Phi_0^- \cap \Psi_0^-$,

$$\overline{q} \in A_{i^*} \text{ and } \overline{q} \notin A_{j*}.$$

Within this subdomain, take $(B_1, ..., B_n)$ in which the B_i contain the same literal values $\overline{r}$ for $r \in \Phi_0 \setminus \Phi_0^-$. By definition, i^* dictates on $(B_1, ..., B_n)$, so

$\overline{q} \in B$. Still within the same subdomain, take $(C_1, ..., C_n)$ in which the C_i contain the same literal values $\overline{p}$ for $p \in \Phi_0 \setminus \Psi_0^-$ and

$$\overline{q} \in C_i \Leftrightarrow \overline{q} \in B_i, \quad i = 1, ..., n.$$

By definition, j^* dictates on $(C_1, ..., C_n)$, so $\overline{q} \notin C$, which contradicts IIPA.

Using the fact just proved and IIPA again, it is easy to check that the common dictator of Φ_0^- and Ψ_0^- is also a dictator on $\Phi_0^- \cup \Psi_0^-$. Now, take any sequence $\Phi_{01}^-, ..., \Phi_{0\nu}^-, ...$ of successive overlapping subsets of the same cardinality $k \geq 2$. We conclude that there is j who is a dictator on all finite Ξ_0^- s.t.

$$\Phi_{01}^- \subseteq \Xi_0^- \subseteq \Phi_0.$$

If Φ_0 is finite, this establishes the lemma. Otherwise, take $(A_1, ..., A_n) \in D^n$ and $p \in \Phi_0$. It must be that $p \in \Xi_0^-$ for some finite Ξ_0^- on which j is a dictator; so it is enough to apply IIPA to $(A_1, ..., A_n)$ and some relevant $(B_1, ..., B_n)$ from a subdomain $(D^-)^n$ determined by Ξ_0^-.

The lemma is key to the theorem because a dictator on Φ_0 is also a dictator on Φ. This broader conclusion follows from Closure Condition (i) and the maximal consistency of judgment sets. So with the lemma in hand, the proof reduces to one's showing that for all Φ_0^- of the suitable cardinality, there is a dictator on Φ_0^-.

Proof. For any $\Phi_0^- \subseteq \Phi_0$with $\left|\Phi_0^-\right| = 3$, fix some D^- and $(D^-)^n$ determined by Φ_0^-. If we prove that there is i^* s.t.

$$\forall (A_1, ..., A_n) \in (D^-)^n, \forall \pi \in \Phi_0^-, \quad \pi \in F(A_1, ..., A_n) \Leftrightarrow \pi \in A_{i^*},$$

we will have proved more generally that there is a dictator on Φ_0^-. This will follow from IIPA, for this condition makes it possible to replace "for *all* subdomains $(D^-)^n$" by "for *one* subdomain $(D^-)^n$" in the definition of a dictator on Φ_0^-.

Closure condition (ii) ensures that there is $p \in \Phi_0^-$, the literal values of which are connected in terms of either 2- or 3-disjunctions with the literal values of the remaining p.v. $q, r \in \Phi_0^-$. Using this fact, we first derive the following *Limited Systematicity* property: for all $\pi \in \Phi_0^-$, and all $(A_1, ..., A_n), (A_1^{'}, ..., A_n^{'}) \in (D^-)^n$,

$$\begin{aligned} (*)\ [p \in A_i \Leftrightarrow \pi \in A_i', i \in N] \ &\Rightarrow \ [p \in A \Leftrightarrow \pi \in A'] \text{ and} \\ (**)\ [p \in A_i \Leftrightarrow \neg\pi \in A_i', i \in N] \ &\Rightarrow \ [p \in A \Leftrightarrow \neg\pi \in A']. \end{aligned}$$

Take $\pi = q$ or $\pi = r$, and $(A_1, ..., A_n)$, $(A_1^{'}, ..., A_n^{'})$ as in the antecedent of (*). There exists $(B_1, ..., B_n) \in (D^-)^n$ s.t.

$$p \in A_i \Leftrightarrow p \in B_i, \text{ and } \pi \in A_i' \Leftrightarrow \pi \in B_i, i \in N.$$

A profile so contructed satisfies the further equivalences:

$$p \in B_i \Leftrightarrow \pi \in B_i, i \in N,$$

so that from deductive closure,

$$\neg p \vee \pi \in B_i \text{ and } p \vee \neg \pi \in B_i, i \in N,$$

and UP implies that

$$\neg p \vee \pi \in B \text{ and } p \vee \neg \pi \in B.$$

Also, $p \in A \Leftrightarrow p \in B$ and $\pi \in A' \Leftrightarrow \pi \in B$ from IIPA. Suppose now that $p \in A$. Then, $p \in B$, and by deductive closure, $\pi \in B$, so $\pi \in A'$. Conversely, suppose that $\pi \in A'$. Then, $\pi \in B$, and by deductive closure again, $p \in B$, so $p \in A$, which completes the proof of (*).

The proof of (**) proceeds similarly. Indeed, with a suitable choice of $(B_1, ..., B_n)$, the application of UP leads to:

$$p \vee \pi \in B \text{ and } \neg p \vee \neg \pi \in B,$$

and the desired conclusion follows again from IIPA and the deductive closure of B.

We have assumed that all 2-disjunctions $\widetilde{p} \vee \widetilde{\pi}$ were available in Φ, but according to Closure Condition (ii), they may be replaced by 3-disjunctions. To take this possibility into account, suppose that $\widetilde{p} \vee \widetilde{\pi} \notin \Phi$ but $\widetilde{p} \vee \widetilde{\pi} \vee \widetilde{\pi'} \in \Phi$, with $\pi, \pi' \in \{q, r\}, \pi \neq \pi'$. If we impose on the auxiliary $(B_1, ..., B_n)$ the further constraint that

$$\neg \widetilde{\pi'} \in B_i, i \in N,$$

UP entails that $\neg \widetilde{\pi'} \in B$, and the previous reasoning, with $\widetilde{p} \vee \widetilde{\pi} \vee \widetilde{\pi'}$ instead of $\widetilde{p} \vee \widetilde{\pi}$, leads again to the conclusion.

If $\pi = p$, statement (*) is simply IIPA, but (**) must be proved. Assuming that

$$p \in A_i \Leftrightarrow \neg p \in A_i', \ i = 1, ..., n,$$

we apply what has been shown for distinct variables. There exist $\pi' \in \Phi_0^-$, $\pi' \neq p$, and $(B_1, ..., B_n) \in (D^-)^n$ s.t.

$$p \in A_i \Leftrightarrow \neg\pi' \in B_i,\ i = 1, ..., n.$$

If $p \in A$, then $\neg\pi' \in B$, hence $\neg p \in A'$, as desired.

Now, we derive the following *Limited Positive Responsiveness* property. Consider two profiles $(A_1, ..., A_n)$, $(A'_1, ..., A'_n) \in (D^-)^n$, s.t. (i) $\widetilde{p} \in A$, (ii) for at least one j, $\widetilde{p} \notin A_j$ and $\widetilde{p} \in A'_j$, (iii) for no i, $\widetilde{p} \in A_i$ and $\widetilde{p} \notin A'_i$. (The last two conditions say that $\widetilde{p}$ does not disappear from any individual judgment set and appears in at least one.) Then, $\widetilde{p} \in A'$.

From (ii) and (iii), there are I, J s.t.

$$I \subset J \subseteq N,$$

and

$$\widetilde{p} \in A_i,\ i \in I,\ \neg\widetilde{p} \in A_i,\ i \in N \diagdown I;\ \widetilde{p} \in A'_i,\ i \in J,\ \neg\widetilde{p} \in A'_i,\ i \in N \diagdown J.$$

(Note that I, but not J, can be empty.) Now, take $\pi \neq p$ and $(B_1, ..., B_n) \in (D^-)^n$ s.t.

$$\widetilde{p} \in A_i \Leftrightarrow \widetilde{p} \in B_i, \text{ and } \widetilde{p} \in A'_i \Leftrightarrow \pi \in B_i, i \in N.$$

Thus, $\widetilde{p} \in A_i \Rightarrow \pi \in B_i, i \in N$, and from deductive closure,

$$\neg\widetilde{p} \vee \pi \in B_i, i \in N,$$

so $\neg\widetilde{p} \vee \pi \in B$ by UP. Now, suppose (i). IIPA entails that $\widetilde{p} \in B$, and the deductive closure of B, that $\pi \in B$. The conclusion that $\widetilde{p} \in A'$ follows from Limited Systematicity. As was explained earlier, there is no loss of generality in assuming that $\neg\widetilde{p} \vee \pi$, rather than some 3-disjunction, belongs to Φ.

The next step proceeds from the following sequence of profiles $\left(A_j^i\right)_{i,j=1,...,n}$ in $(D^-)^n$:

$$\begin{array}{llll} p \in A_1^1 & \neg p \in A_2^1 & ... & \neg p \in A_n^1 \\ p \in A_1^2 & p \in A_2^2 & \neg p \in A_3^2 ... & \neg p \in A_n^2 \\ ... & ... & ... & ... \\ p \in A_1^n & p \in A_2^n & p \in A_3^n ... & p \in A_n^n \end{array}.$$

Denote by $(A^i)_{i=1,...,n}$ the associated sequence of social judgment sets. Define i^* to be the first i such that $p \in A^i$; hence, both $p \in A^{i^*}_{i^*}$ and $p \in A^{i^*}$ hold. UP ensures that i^* exists. We will prove that i^* is a dictator on Φ_0^-, and for this, we need another preparatory step.

We aim at showing that there exists a profile $(B_1, ..., B_n) \in (D^-)^n$ s.t. $p \in B$ and

$$(*)\ p \in B_{i^*},\ \neg p \in B_i,\ i \neq i^*.$$

If $i^* = 1$, it is enough to take the $\left(A^1_j\right)_{j=1,...,n}$ line. If $i^* \geq 2$, we define three sets of individuals:

$$I = \{1, ..., i^* - 1\},\ J = \{i^*\},\ K = \{i^* + 1, ..., n\},$$

the last of which may be empty. Assume by way of contradiction that $\neg p \in B$ for all profiles satisfying $(*)$. In particular, $\neg p \in B'$ for $(B'_1, ..., B'_n)$ satisfying $(*)$ and

$$\begin{aligned} \neg q &\in B'_j,\ j \in I \cup J;\ q \in B'_j,\ j \in K, \\ r &\in B'_j,\ j \in I;\ \neg r \in B'_j,\ j \in J \cup K. \end{aligned}$$

Limited Systematicity entails that $\neg q \in B'$ in view of the $\left(A^{i^*}_j\right)_{j=1,...,n}$line, and that $\neg r \in B'$ in view of the $\left(A^{i^*-1}_j\right)_{j=1,...,n}$ line. Inspection of $(B'_1, ..., B'_n)$ shows that by deductive closure,

$$p \vee q \vee r \in B'_i, i \in N,$$

whence $p \vee q \vee r \in B$ from UP. But we have just shown that $\neg p, \neg q, \neg r \in B'$,which contradicts the consistency condition on B'. For notational simplicity, we have assumed that Closure Condition (iii) made $p \vee q \vee r$ available in Φ; a parallel reasoning would take care of any other 3-disjunction.

To sum up, we have produced a profile in which i^* alone accepts p and the social set endorses i^* on this p.v. When applied to both p and $\neg p$, Limited Positive Responsiveness entails the equivalence:

$$\forall (A_1, ..., A_n) \in (D^-)^n, \quad p \in F(A_1, ..., A_n) \Leftrightarrow p \in A_{i^*}.$$

In view of Limited Systematicity, p can be replaced by any $\pi \in \Phi_0^-$, which establishes the target statement of the first paragraph. ■

The reader will have noticed the social-theoretic undertones of this proof. The last but one paragraph established a "semi-decisiveness property" for a particular item that the last paragraph extended to a "deciseness

property" for any item. Also, the proof of Arrow's theorem for "economic" domains often depends on first proving dictatorship on three-element sets.

Notice the rôle of the assumption that individual and social sets are not only consistent and deductively closed, but also *maximal*. Gärdenfors (2005) has stressed that this is a strong assumption to make, and it turns out to be crucial at several stages of the proof.

A much simpler argument works for the case $n = 2$.

Proof. Suppose that F satisfies IIPA and UP, but is non-dictatorial. Accordingly, there are two profiles (A_1, A_2), $(A'_1, A'_2) \in D^2$, s.t. $A \neq A_1$, $A' \neq A'_2$. From Closure Condition (i) and relative maximality, two judgment sets differ from each other iff they differ in terms of some p.v. Thus, there are two literals $\widetilde{p}$, $\widetilde{q}$ s.t.

$$\neg\widetilde{p} \in A_1,\ \widetilde{p} \in A, \text{ and } \neg\widetilde{q} \in A'_2,\ \widetilde{q} \in A'.$$

If $p \neq q$, we can find $(B_1, B_2) \in D^2$ s.t.

$$(*)\ \widetilde{p} \in A_i \Leftrightarrow \widetilde{p} \in B_i,\ \widetilde{q} \in A'_i \Leftrightarrow \widetilde{q} \in B_i,\ i = 1, 2.$$

Deductive closure entails that $\neg\widetilde{p} \vee \neg\widetilde{q} \in B_i$, $i = 1, 2$, so that $\neg\widetilde{p} \vee \neg\widetilde{q} \in B$ follows from UP. However, IIPA requires that both $\widetilde{p} \in B$ and $\widetilde{q} \in B$, which violates consistency. To deal with the case $p = q$, it is enough to replace q by some $r \neq p$ in the last equivalences $(*)$.

This proof uses only two p.v., which shows that Theorem 1 extends to the limiting case $|\Phi_0| = n = 2$. Indeed, the main proof needed three p.v. (and a genuine 3-disjunction) only in the construction of a profile in which one i alone accepts p and is endorsed by the social set; but such a profile is trivially available when there are just two individuals.

We end up with the promised decomposition of IIA.

Corollary 3 *Under the conditions stated on* Φ, *Independence of Irrelevant Alternatives is equivalent to Independence of Irrelevant Propositional Alternatives conjoined with Unanimity Preservation.*

4 Comparisons and further comments

Among the existing impossibility theorems, Pauly and van Hees's (forthcoming) is the easiest of all to compare with Theorem 1. They conclude

that F is dictatorial from the two conditions that F satisfies IIA and is non-constant, assuming an agenda Φ with the following features: $|\Phi_0| \geq 2$, Closure Condition (i) holds, as well as

(ii') $\widetilde{p} \wedge \widetilde{q} \in \Phi$ for all literals $\widetilde{p}, \widetilde{q}$.

Pauly and van Hees's choice of conjunctions rather than disjunctions is immaterial, since what can be proved with the one can be proved with the other, given the logical assumptions made on judgment sets. The difference between their agenda and ours can be visualized by comparing the smallest possible agendas for which both theorems have been proved. For $n \geq 3$, the respective agendas are

$$\mathcal{A}' = \{all\ \widetilde{p}, \widetilde{q}\} \cup \{\widetilde{p} \wedge \widetilde{q}\ for\ all\ \widetilde{p}, \widetilde{q}\}$$

and the already defined $\mathcal{A}_{223}$, $\mathcal{A}_{2233}$, $\mathcal{A}_{33}$. As we have explained, Theorem 1, not just its proof, cannot possibly hold true of $\mathcal{A}'$. Thus, at least part of the difference between Pauly and van Hees's and our agendas must reflect the substantial difference between the axiomatic conditions.

Starting from a finite $\mathcal{P}$, Dietrich (2006) has proved an enlightening variant of Pauly and van Hees's theorem that weakens IIA by reserving it to the *atoms* of $\mathcal{L}$, i.e., to the complex formulas $\widetilde{p_1} \wedge ... \wedge \widetilde{p_m}$ obtained for all possible $\widetilde{p_1}, ..., \widetilde{p_m}$, where $|\mathcal{P}| = m \geq 2$. Like other agendas in the literature that do not satisfy Closure Condition (i), this one is incomparable with Pauly and van Hees's and ours except for the limiting case $m = 2$.

Some writers have recently provided necessary and sufficient conditions for an agenda Φ to be dictatorial, granted that F satisfies some standard conditions. Dokow and Holzman's (2005) criterion requires F to satisfy IIA and UP, while - in a somewhat different framework - Nehring and Puppe's (2005) s.j.f. must satisfy IIA and Positive Responsiveness. These instructive results make it possible to recover several theorems at once, including Pauly and van Hees's, but not ours, since the latter starts from the non-standard IIPA. However, once IIA is recovered through the Corollary, Dokow and Holzman's criterion for a dictatorial agenda must apply as a necessary condition, and it is easy to check that the Closure Conditions imply it. Because Nehring and Puppe's Positive Responsiveness follows from IIA and UP (this can be proved by adapting part of proof of Theorem 1), their criterion is also applicable, and it leads to the same reassuring conclusion. As a matter of fact, inspired by this very criterion, Nehring (2005) has independently recovered a version of Theorem 1 within his abstract

theory of Pareto aggregation. Nehring's result and ours are not equivalent because the frameworks they belong to are dissimilar (the Nehring-Puppe framework is based on an agenda that distinguishes premisses from conclusions and singles out a particular conclusion, as in the initial discursive dilemma).

Despite being larger than necessary for the proof, the class of agendas $\mathcal{A}_K$ is of some interest, both technical and conceptual. Developing the idea of local dictatorship in a new direction, we define $j \in N$ to be a *dictator for* $(A_1, ..., A_n) \in D^n$ if $F(A_1, ..., A_n) = A_j$, and F to *dictatorial profile by profile* if for each $(A_1, ..., A_n)$ under consideration, there exists a dictator for $(A_1, ..., A_n)$.

Proposition 4 *If $\Phi = \mathcal{A}_K$, with $K = n$, and F satisfies UP, then F is dictatorial profile by profile.*

Proof. Suppose to the contrary that there exists $(A_1, ..., A_n) \in D^n$ s.t. $A \neq A_i$, for all $i \in N$. Because of maximality relative to $\mathcal{A}_K$, for all i, there is $\widetilde{p}_i$ with the property that $\widetilde{p}_i \in A_i$ and $\neg\widetilde{p}_i \in A$. Deductive closure relative to $\mathcal{A}_K$ entails that $\vee_{i=1,...,n}\widetilde{p}_i \in A_i$ for all $i \in N$, whence $\vee_{i=1,...,n}\widetilde{p}_i \in A$ from UP. This implies that A implies is inconsistent. (The argument parallels that already used to simplify the proof of Theorem 1 when $n = 2$.)

Using this proposition as a lemma, we can devise a simpler proof of Theorem 1 that dispenses with Lemma 2 (such a proof was employed in the earlier version of the paper). More importantly, it comes as a surprise that profile by profile dictatorship follows from UP without any application of IIPA. A condition that goes such a long way towards dictatorship cannot be very appealing. Further, UP is open to an objection paralleling that raised against IIA. Unanimity over a disjunctive formula may be *spurious* in the sense of relying on conflicting endorsements of the reasons - here identified with the disjuncts - for endorsing that very formula. The issue of spurious unanimity has been discussed by Mongin (1997) in a probabilistic context, and thoroughly explored by Nehring (2005) in a general framework. If UP turns out to be dubious for the same reasons as IIA was, it seems natural to relax it in the same way, i.e., by restricting it to p.v. When this is also done, attractive - or at least well-regarded - aggregative rules emerge, such as **(R)** and further variants of majority voting. In order to obtain a social judgment on the complex formulas, it is enough to take the deductive closure of the social set relative to the agenda.

A s.j.f. that fits with this description is a special case of the so-called *premiss-based procedure* in which the premisses are the literals $\widetilde{p}$. By contrast, the initial F, which satisfies either IIA or IIPA and UP, illustrates the *conclusion-based procedure*, where the social set treats every formula in the same way, whether it appears as a premiss or as a conclusion in the individual judgment sets. In stopping the argument at this point, we would reach a diagnosis that is sometimes encountered in the literature, i.e., that the impossibility of logical judgment aggregation can be resolved by moving from the conclusion-based to the premiss-based procedure. The diagnosis would be refined by the new decomposition of this paper, which eventually localizes the source of the trouble within the unanimity side of IIA. There is, however, an argument in the way of this analysis. Bovens and Rabinowicz (2006) have recently defended the conclusion-based procedure by showing that under relevant probabilistic restrictions, majority voting over *both* premisses and conclusions is more likely to lead to the truth than majority voting over just the premisses, followed by the deductive closure operation. Bovens and Rabinowicz's framework differs from the present one in many significant respects, and their technical assumptions, like probabilistic independence, cannot even be reexpressed here, but their point seems to have a potential of wide generality. As far as we understand, voting on the conclusion is favorable when the individuals have independent information on the premisses; this enhances the probability that true conclusions deriving from these premisses will be included in the social set. Of course, if, as in a previous example, a 3-individual society unanimously supports the conclusion $p \vee q \vee r$, with each individual accepting only one of the premisses p, q, r, and if, say, only p is true, then $p \vee q \vee r$ is accepted for wrong reasons, since two individuals out three are mistaken. But this objection should be weighed against the advantage of having the true formula $p \vee q \vee r$ included in the social set; if the individuals had voted only on the literals, $p \vee q \vee r$ could not have been deduced. We take Bovens and Rabinowicz to make a general point like this; in effect, they counter our spurious unanimity argument by observing that spurious unanimity may be useful, as far as the objective of truth is concerned.

5 Conclusion

This paper has derived the impossibility of logical judgment aggregation from assumptions that, for one, highlight connections with social choice theory, and for another, appear to deliver a more interesting paradox than the essentially unique assumption of the initial theorem. IIA has been reduced to such a weak requirement as to exonerate it from its responsibility in the unpleasant conclusion. Although UP now comes first on the black list, many economists may be reluctant to give up a principle that is so closely related to the Pareto principle. It remains for them to blame (a) the logical framework of judgment, (b) universal domain, or (c) the demanding assumptions put on judgment sets. Logical judgment theorists have begun to explore weakenings of (b) and (c); however, some of the possibilities they have then disclosed - e.g., oligarchies - remain unattractive or exaggerately specific. Another avenue that recommend itself to the theoretical economist is to depart from (a) and its set of embodied constraints - typically, by moving to the probabilistic framework.

6 References

Arrow, K.J. (1963), *Social Choice and Individual Values*, New York, Wiley (1st ed.1951).

Bacharach, M., L. A. Gérard-Varet, P. Mongin, H. Shin (1997) (eds), *Epistemic Logic and the Theory of Games and Decisions*, Dordrecht, Kluwer.

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

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

Dietrich, F. and C. List (2004), "Strategy-Proof Judgment Aggregation", mimeo, The London School of Economics.

Dokow, E. and R. Holzman (2005), "Aggregation of Binary Evaluations", mimeo, Technion-Israel Institute of Technology.

Gärdenfors, P. (2006), "A Representation Theorem for Voting With Logical Consequences", *Economics and Philosophy*, 22, p. 181-190.

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

Mongin, P. (1997), "Spurious Unanimity and the Pareto Principle", THEMA Working Papers, Université de Cergy-Pontoise.
Mongin, P. (2005), "Logical Aggregation, Probabilistic Aggregation, and Social Choice", mimeo, invited lecture at the LGS4 Conference, Caen, June 2005.
Nehring, K. (2005), "The (Im)possibility of a Paretian Rational", mimeo, University of California at Davis.
Nehring, K. and C. Puppe (2005), "Consistent Judgment Aggregation: A Characterization", mimeo, University of California at Davis and Universität Karlsruhe.
Pauly, M. and M. van Hees (2003), "Logical Constraints on Judgment Aggregation", forthcoming in *Journal of Philosophical Logic.*
Samuelson, P.A. (1977), "Reaffirming the Existence of "Reasonable" Bergson-Samuelson Social Welfare Functions", *Economica*, 44, p. 81-88.
Wilson, R. (1972), "Social Choice Theory Without the Pareto Principle", *Journal of Economic Theory*, 5, p. 478-486.