Logical Aggregation, Probabilistic Aggregation and Social Choice

Philippe Mongin
CNRS and Ecole Polytechnique
philippe.mongin@shs.polytechnique.fr

LGS4, June 22, 2005

In political science and legal theory, the *doctrinal paradox* (or *discursive dilemma*) is 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, the resulting set of propositions may be logically inconsistent.

Example discussed in List and Pettit (2002).

A court decides whether a defendant is liable under a charge of breach of contract. The judges will find against the defendant iff they conclude that a valid contract was made and the defendant broke that existing valid contract.

So the three questions to be answered are: A. Valid contract in existence? B. Breach? C. Defendant liable?

There are 3 judges and they reach their conclusion by voting on the three questions separately. The votes are:

Judge 1: A. Yes. B. No. C. No

Judge 2: A. No. B. Yes. C. No

Judge 3: A. Yes. B. Yes. C. Yes

The court concludes: A. Yes. B. Yes. C. No, violating the legal doctrine, which judges have abided by individually.

A mathematical theory of *logical judgment aggregation* has grown out of this example. Its method consists in introducing a mapping F that associates a social set of judgments A with profiles of sets of individual judgments $(A_1, .., A_n)$.

A mathematical theory of *logical judgment aggregation* has grown out of this example. Its method consists in introducing a mapping F that associates a social set of judgments A with profiles of sets of individual judgments $(A_1, .., A_n)$.

Judgments are represented by formulas in some logical language, typically propositional logic or simple extensions of it (e.g., modal propositional logic). Sets of judgments satisfy standard logical properties such as consistency, etc.

The theory investigates the effect of imposing axiomatic conditions on this mapping. This is reminiscent of the method of social choice theory, and the major theorem obtained thus far has an Arrovian flavour.

This impossibility theorem accounts for the doctrinal paradox in the same sense as Arrow's theorem accounts for the related Condorcet paradox.

THE LOGICAL FRAMEWORK

The language (set of formulas) $\mathcal{L}$ is constructed from a set $\mathcal{P}$ of propositional variables $p_1, ..., p_k, ...,\ k \geq 2$, and the propositional connectives $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$ ("not", "or", "and", "implies", "equivalent to").

The axiomatic system of propositional logic fixes the inference relation, $B \vdash \varphi$, for any $B \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$, and the derivative logical notions (logical equivalence $\vdash\dashv$, logical truth $\top$, contradiction $\bot$, consistency, logical independence, etc).

A formula φ is in *normal disjunctive form* (NDF) if it is a disjunction of formulas, each of which is a conjunction of propositional variables or negations of propositional variables (denoted at once by $\widetilde{p}$). Each φ except for $\bot$ is equivalent to some φ' in normal disjunctive form:

$$\varphi' = \vee_{l=1,\dots,L}(\wedge_{m=1,\dots,M}\widetilde{p}_{lm}).$$

If $\mathcal{P}$ is finite, $\mathcal{L}$ has a finite number of formulas up to logical equivalence, and it is possible to construct the *atoms* of $\mathcal{L}$, $\widetilde{p}_1 \wedge \dots \wedge \widetilde{p}_K$. Every $\varphi \neq \bot$ is equivalent to a disjunction of atoms, and we redefine the n.d.f. to be this particular disjunction.

An *agenda* is a nonempty subset Φ of formulas representing the propositions on which the n individuals and society pass judgment. It is easier to take Φ to be some relevant sublanguage such as $\mathcal{L}(p_l,\, p_m)$ or $\mathcal{L}(p_l,\, p_m, p_o)$.

A *judgment set* is any *maximally consistent* set $B \subseteq \mathcal{L}$, where consistency is already defined, and maximality means:

for any $\varphi \in \Phi$, either φ or $\neg\varphi$ belongs to B.

Such sets are *deductively closed* relative to Φ, i.e.:

for any $\varphi \in \Phi$, if $B \vdash \varphi$, then $\varphi \in B$.

An important consequence of the definition of judgments sets: if B and B' differ from each other, they must differ vis-à-vis some $p \in \mathcal{P}$ (i.e., $\widetilde{p} \in B$ and $\widetilde{p} \notin B'$).

The maximality condition is strong and questionable (Gärdenfors, 2005).

A *social judgment function* is any mapping:

$$F : (A_1, ..., A_n) \mapsto A$$

where the A_i, $i = 1, ..., n$, and A are judgment sets.

An individual j is a *dictator for* $(A_1, ..., A_n)$ if:

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

and a *dictator* if:

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

Notation: $A = F\ (A_1, ..., A_n)$,
$A' = F(A'_1, ..., A'_n), ...$

Axiom 1 *(Universal Domain)* $F : D^n \to D$, *where* D *is the set of all possible judgment sets.*

Axiom 2 *(Independence)*
$\forall \varphi \in \Phi, \forall (A_1, ..., A_n), (A'_1, ..., A'_n),$

$$\left[\forall i, \varphi \in A_i \Leftrightarrow \varphi \in A'_i\right] \Rightarrow \left[\varphi \in A \Leftrightarrow \varphi \in A'\right].$$

Axiom 3 *(Nonconstancy)* F *is not a constant mapping.*

Theorem 1 *If* F *satisfies Universal Domain, Independence, and Nonconstancy, it is dictatorial.*

Pauly and van Hees (2003) and Dietrich (2004) prove this theorem for slightly different classes of agendas.

Pauly and van Hees prove another theorem using a stronger condition than independence (first introduced by List and Pettit, 2002).

Axiom 4 *(Systematicity)*
$\forall \varphi, \psi \in \Phi, \forall (A_1, ..., A_n), (A'_1, ..., A'_n),$

$$\left[\varphi \in A_i \Leftrightarrow \psi \in A'_i, i = 1, ..., n\right] \Rightarrow [\varphi \in A \Leftrightarrow \psi \in A']$$

Theorem 2 *If F satisfies Universal Domain and Systematicity, F is dictatorial.*

SKETCH OF A PROOF for $n = 2$ and $\Phi = \mathcal{L}(p_1, p_2)$.

There are four judgment sets to be considered:

$$B_1 = \{p_1, p_2, p_1 \wedge p_2, p_1 \longleftrightarrow p_2, ...\},$$

$$B_2 = \{\neg p_1, p_2, \neg p_1 \wedge p_2, \neg p_1 \longleftrightarrow p_2, ...\},$$

$$B_3 = \{p_1, \neg p_2, p_1 \wedge \neg p_2, \neg p_1 \longleftrightarrow p_2, ...\},$$

$$B_4 = \{\neg p_1, \neg p_2, \neg p_1 \wedge \neg p_2, p_1 \longleftrightarrow p_2, ...\}.$$

Suppose that $A_1 = B_1$ and $A_2 = B_2$. Then, $p_1 \wedge p_2$ has the same membership pattern as p_1; hence, by Systematicity, $A \neq B_3$ (since $p_1 \in B_3$ and $p_1 \wedge p_2 \notin B_3$, contradicting $A = B_3$).

Similarly, $\neg p_1 \wedge p_2$ has the same membership pattern as $\neg p_1$; hence $A \neq B_4$ (since $\neg p_1 \in B_4$ and $\neg p_1 \wedge p_2 \notin B_4$).

Surveying all possibilities for A_1 and A_2, we conclude each time that A cannot be different from both A_1 and A_2, i.e., that F *is dictatorial for each profile*.

If there were no (overall) dictator, there would exist (A_1, A_2) and (A'_1, A'_2) with:

$$A_1 \neq A_2,\ A'_1 \neq A'_2$$

and

$$A = A_1,\ A' = A'_2.$$

So there would exist $p, q \in \mathcal{P}$ such that $\tilde{p} \in A_1, \tilde{p} \notin A_2$, and $\tilde{q} \in A'_1, \tilde{q} \notin A'_2$. From systematicity $\tilde{p} \in A$ iff $\tilde{q} \in A'$, a contradiction.

The case of $n > 2$ is not as easy, and anyway Theorem 1 is far from being as direct as Theorem 2.

SYSTEMATICITY is normatively unattractive. It requires that two formulas be treated alike if they draw the support of exactly the same people. Take a two-individual group in which 1 thinks that the European constitution is worthless, 2 disagrees, and society endorses 1's judgment. Then, if 1 also thinks that the American constitution is worthless, and 2 disagrees again, the society should endorse 1's judgment once again; and similarly with other unrelated topics.

Instead of permitting variations in *both* $(A_1, ..., A_n)$ and φ, INDEPENDENCE fixes the formula and only the profile varies; thus, it avoids the objection against the other condition.

It singles out the requirement contained in Systematicity that the social judgment on φ should depend only on the individual judgments on φ. The best normative defence for INDEPENDENCE is that it prevents manipulation (Dietrich, 2004, and Dietrich and List, 2004). But this argument works only if the agenda is allowed to vary, which is not the case here.

And the manipulation argument does not respond to the charge of irrationality. One would expect the social judgment to care about individual reasons for accepting or rejecting φ and these reasons are implicit in the individual judgment sets. The social judgment should depend on more information drawn from these sets.

SUGGESTED WEAKENINGS:

Axiom 5 *(Independence Justified by Disjuncts φ_l, cf. NDF of φ).*

$$\forall \varphi \in \Phi, \forall (A_1, ..., A_n), (A'_1, ..., A'_n),$$

$$(\forall i, \varphi \in A_i \Leftrightarrow \varphi \in A'_i)$$

$$\& \left(\exists l \ \ s.t. \ \forall i, \varphi_l \in A_i \Leftrightarrow \varphi_l \in A'_i \right)$$

$$\Rightarrow [\varphi \in A \Leftrightarrow \varphi \in A'] \, .$$

The reasons for accepting or rejecting φ are located in the disjuncts φ_l. If you accept $p_1 \vee p_2$, this is because you accept p_1 or/and p_2, and your judgment set says what is the case. This condition restricts Independence to a situation where *individuals focus on the same argument for acceptance or rejection*.

For finite languages, the condition implies:

Axiom 6 *(Independence Limited to Atoms)*. $\forall \varphi = \widetilde{p}_1 \wedge ... \wedge \widetilde{p}_K \in \Phi$,

$$\forall (A_1, ..., A_n), (A'_1, ..., A'_n),$$

$$\left[\forall i, \varphi \in A_i \Leftrightarrow \varphi \in A'_i\right] \Rightarrow [\varphi \in A \Leftrightarrow \varphi \in A']$$

Dietrich's (2004) version of Theorem 1 for finite languages employs this very condition.

HENCE THIS IS NOT THE WAY OUT.

An alternative weakening:

Axiom 7 *(Maximally Limited Independence)*
$\forall p \in \mathcal{P}$,

$\forall (A_1, ..., A_n), (A'_1, ..., A'_n)$

$\left[\forall i, p \in A_i \Leftrightarrow p \in A'_i\right]$

$\Rightarrow [p \in A \Leftrightarrow p \in A']$

Propositional variables are the only formulas the acceptance or rejection of which does not have to be justified, since they are basic to the language. Unqualified independence seems unproblematic for this class of formulas.

WITH MAXIMALLY LIMITED INDEPENDENCE, THE IMPOSSIBILITY THEOREM VANISHES.

Define j to be *an antidictator on* $\mathcal{P}$ if for all $p \in \mathcal{P}$ and all $(A_1, ..., A_n) \in D$,

$$p \notin F(A_1, ..., A_n) \Leftrightarrow p \in A_j.$$

This satisfies Maximality Limited Independence, though not Independence. Take $\Phi = \mathcal{L}\{p_1, p_2\}$, and $A_1 = \{p_1, \neg p_2, ...\}$, $A_2 = \{\neg p_1, p_2, ...\}$, $A_1' = \{p_1, p_2, ...\}$, $A_2' = \{\neg p_1, p_2, ...\}$, and assume that 1 is the antidictator. Then, $A = \{\neg p_1, p_2, ...\}$ and $A' = \{\neg p_1, \neg p_2, ...\}$. By maximality and consistency, we have $p_1 \vee p_2 \in A$ and $p_1 \vee p_2 \notin A'$, a contradiction with Independence.

This is not an attractive rule but it reveals a failure of Theorem 1 and an explanation for the impossibility: it is tied with conditions that treat propositional variables and molecular formulas alike.

The *premiss-based procedure* (List, Pettit, Dietrich) relies on majority voting on a limited subset $P \subset \Phi$ of formulas φ and their negations $\neg\varphi$. It completes the social judgment set by drawing logical inferences afterwards. This procedure delivers a consistent set if P is made out of logically independent formulas. With $P = \mathcal{P}$, the PBP clearly satisfies Maximally Limited Independence.

In the presence of a Unanimity Condition, Maximally Limited Independence leads to dictatorship again.

This is a new variant of the impossibility theorem.

(Unanimity) For all $\varphi \in \Phi$, and all profiles $(A_1, ..., A_n)$,

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

Theorem 3 *If F satisfies Universal Domain, Maximally Limited Independence, and Unanimity, it is dictatorial.*

The proof uses a Limited Systematicity condition, i.e., limited to propositional variables and their negations. See Mongin (2005a).

Corollary 1 *If F is satisfies Universal Domain and Nonconstancy, F satisfies Independence iff it satisfies Maximally Limited Independence and Unanimity.*

We have in effect factored out the initial Independence condition into independence proper and a Pareto-like condition. This likens the impossibility theorem to Arrow's and dispels some of its mystery.

TO THE PROBABILISTIC FRAMEWORK

A (*Boolean*) *valuation* $v \in V$ is a mapping from $\mathcal{P}$ to $\{0, 1\}$ satisfying the usual truth-functionality properties:

$v(\neg\varphi) = 1 \Leftrightarrow v(\varphi) = 0$,

$v(\varphi \wedge \psi) = 1 \Leftrightarrow \min(v(\varphi), v(\psi)) = 1$, etc

(It is enough to define v on $\mathcal{P}$ because it can be uniquely extended to $\mathcal{L}$.)

The completeness theorem of the propositional calculus states that φ is a *tautology* (i.e., $v(\varphi) = 1$ for all $v \in V$) iff and only is φ is a *theorem*(i.e., $\varphi \vdash\dashv \top$). This means that valuations can replace judgment sets, and the theory can be developed entirely semantically.

A further step is possible in view of the point that a valuation is a limiting case of a probability measure.

$\mathcal{L}$ quotiented by $\vdash\dashv$ is a Boolean algebra $\mathcal{L}^*$, and the completeness theorem again (abstractly, Stone's representation theorem) makes it possible to replace $\mathcal{L}^*$ and the logical operations $\neg, \wedge, \vee, ...$ by a measurable set $(\Omega\ ,\mathcal{A})$ and the set-theoretic operations $.^c, \cap, \cup, ...$

With this translation, each v can be paired 1-1 with a set-function v^* from $(\Omega\ ,\mathcal{A})$ to $\{0, 1\}$ satisfying the probability axioms.

Take a set Ω with $|\Omega| \geq 3$ and an algebra of subsets $\mathcal{A} \subseteq 2^{\Omega}$ containing at least three mutually distinct A_1, A_2 and A_3. We consider two sets of mappings on $\mathcal{A}$:

- The set $\Delta(\Omega, \mathcal{A})$ of all probability measures on $\mathcal{A}$.
- The set $\Delta^{-}(\Omega, \mathcal{A})$ of all *0-1 probability measures* on $\mathcal{A}$.

A *probabilistic social judgment function* is any mapping

$$G : (\pi_1, ..., \pi_n) \mapsto \pi, \text{ where } \pi_i, \pi \in \Delta(\Omega, \mathcal{A}).$$

A *0-1 social judgment function* is any mapping

$$G^{-} : (\pi_1, ..., \pi_n) \mapsto \pi, \text{ where } \pi_i, \pi \in \Delta^{-}(\Omega, \mathcal{A}).$$

Social judgment functions are *dictatorial* if they are projections.

Axiom 8 *(Universal Domain*)* G *(resp.* G^{-}*) is defined on all* n*-tuples of relevant items.*

Axiom 9 *(Independence*)*

$$\forall A \in \mathcal{A}, \forall(\pi_1, ..., \pi_n), (\pi'_1, ..., \pi'_n), \forall r$$

$$\left[\forall i, \pi_i(A) = r \Leftrightarrow \pi'_i(A) = r\right] \Rightarrow [\pi(A) = r \Leftrightarrow \pi'(A) = r]$$

Equivalently: for all $A \in \mathcal{A}$, there is a mapping $\widetilde{G}_A : [0,1]^n \longmapsto [0,1]$ s.t. for all $(\pi_1, ..., \pi_n)$, $\pi(A) = \widetilde{G}_A(\pi_1(A), ..., \pi_n(A))$; resp., a mapping $\widetilde{G^{-}}_A : \{0,1\}^n \longmapsto \{0,1\}$ s.t. for all $(\pi_1, ..., \pi_n)$, $\pi(A) = \widetilde{G^{-}}_A(\pi_1(A), ..., \pi_n(A))$.

Axiom 10 *(Nonconstancy*) G (resp. G^-) is not a constant mapping.*

Axiom 11 *(Systematicity*)*

$$\forall A, B \in \mathcal{A}, \forall(\pi_1, ..., \pi_n), (\pi'_1, ..., \pi'_n), \forall r$$

$$\left[\forall i, \pi_i(A) = r \Leftrightarrow \pi'_i(B) = r\right] \Rightarrow [\pi(A) = r \Leftrightarrow \pi'(B) = r]$$

Equivalently: there is a mapping $\widetilde{G} : [0, 1]^n \longmapsto [0, 1]$ s.t. for all $(\pi_1, ..., \pi_n)$, $\pi(A) = \widetilde{G}(\pi_1(A), ..., \pi_n(A)$ respectively, a mapping $\widetilde{G^-} : \{0, 1\}^n \longmapsto \{0, 1\}$ s.t. for all $(\pi_1, ..., \pi_n)$, $\pi(A) = \widetilde{G^-}(\pi_1(A), ..., \pi_n(A))$.

Axiom 12 *(r-Unanimity*) For all $A \in \mathcal{A}$, and all profiles $(\pi_1, ..., \pi_n)$,*

$$\pi_i(A) = r, i = 1, ..., n \Rightarrow \pi(A) = r.$$

The theory of probabilistic aggregation explores the effect of imposing these and other conditions on G. We will extend it to the case of G^- and thus recover the theory of logical aggregation as a limiting case.

Theorem 4 *If G satisfies Universal Domain*, Independence*, and 0-Unanimity*, there exist $\lambda_1, ..., \lambda_n \geq 0$ such that $\sum \lambda_i = 1$ and for all $(\pi_1, ..., \pi_n)$ and all $A \in \mathcal{A}$, $\pi(A) = \sum \lambda_i \pi_i(A)$.*

Corollary 2 *If G satisfies Universal Domain*, Systematicity*, and 0-Unanimity*, the same conclusion holds.*

Theorem 5 is due to McConway (1981); see also Genest (1984).

Here are Theorems 1 and 2 in the probabilistic framework:

Theorem 5 *If G^- satisfies Universal Domain*, Independence* and Nonconstancy*,it is dictatorial.*

Theorem 6 *If G^- satisfies Universal Domain* and Systematicity*, it is dictatorial.*

We can derive Theorems 7 and 8 (hence the theory of logical aggregation) from the same probabilistic ideas that drive Theorem 5. The resulting proofs are elegantly simple. See Mongin (2005b) for details.

A COMPARISON OF RESULTS

With the axiomatic conditions kept the same, but the framework enriched, the dictatorship conclusion of Theorems 7 and 8 turns into the well-behaved aggregation rule of Theorem 5. This suggests that the impossibility of logical aggregation is more of an artifact of the propositional calculus than of the conditions themselves.

Another point is that the probabilistic framework *disentangles Unanimity from Independence*. (Independence or Systematicity imply 0-Unanimity (1-Unanimity) only in the case of G^{-}, not G.) It is another artifact of the propositional calculus that one can be hidden under the guise of the other.

Maximallly Limited Independence would amount to reserving independence to a generating subalgebra $\mathcal{A}'$ instead of the whole of $\mathcal{A}$. However, this weakening is not needed anymore to get rid of dictatorship. And is perhaps not as normatively compelling as it was in the other framework. The "linear pooling" rule is widely used in statistics and decision theory. It is roughly plausible in itself and has received plausible derivations elsewhere.

(Mongin, 1995, derives it from 1/2-Unanimity alone in the context of a unique profile of non-atomic probability measures.)

A BRIEF COMPARISON WITH SOCIAL CHOICE THEORY.

This topic is moving fast (Dietrich and List, 2005; Dokow and Holzman, 2005).

Both logically and conceptually, Independence is related to IIA in Arrow, 1963, and Systematicity to Neutrality as in later theorists, e.g., Pollak, 1979. Samuelson, 1977, has emphasized that it is better to base Arrow's theorem on IIA than Neutrality, and the theory of logical aggregation has implicitly followed his hint. But both IIA and Independence remain objectionable on the ground that they do not exploit enough information from the individual items.

A striking difference is that some motivated weakenings of IIA deliver non-dictatorial solutions and even interesting possibilities, but this way out appears not to be available here. (Fleurbaey, Suzumura and Tademuna, 2004, compare weak variants of IIA; some still deliver dictatorship, and others not.)

The only way out is then to introduce a numerical structure, as in the probabilistic or related framework (non-additive probabilities), and exploit the added information that numbers make available. This move parallels the move in social choice theory consisting in introducing interpersonal comparisons of numerical satisfaction.

REFERENCES

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

Dietrich, F. (2004), "Judgment Aggregation: (Impossibility) Theorems", forthcoming in *Journal of Economic Theory*.

Dietrich, F. and C. List (2004), "Strategy-Proof Judgment Aggregation", mimeo.

Dietrich, F. and C. List (2005), "Arrow's Theorem in Judgment Aggregation", mimeo.

Dokow, E. and R. Holzman (2005), "Aggregation of Binary Evaluations", mimeo.

Fleurbaey, M., K. Suzumura, and K. Tadenuma (2002), "Arrovian Aggregation in Economic Environments: How Much Should We Know About Indifference Surfaces?", forthcoming in *Journal of Economic Theory*.

Gärdenfors, P. (2004), "A Representation Theorem for Voting With Logical Consequences", mimeo.

Genest, C. (1984), "Pooling Operators with the Marginalization Property", *Canadian Journal of Statistics*, 12, p. 153-163.

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

McConway, K.J. (1981), "Marginalization and Linear Opinion Pools", *Journal of the American Statistical Association*, 76, p. 410-418.

Mongin, P. (2005a), "Factoring Out the Impossibility of Logical Aggregation", mimeo.

Mongin, P. (2005b), "Logical Aggregation as a Limiting Case of Probabilistic Aggregation", mimeo.

Nehring, K. and C. Puppe (2005), "Consistent Judgment Aggregation: A Characterization", mimeo.

Pauly, M. and M. van Hees (2003), "Logical Constraints on Judgment Aggregation", forthcoming in *Journal of Philosophical Logic*.