Logical Constraints on Judgement Aggregation

Marc Pauly* Martin van Hees†

September 26, 2003

Abstract

Logical puzzles like the doctrinal paradox raise the problem of how to aggregate individual judgements into a collective judgement, or alternatively, how to merge collectively inconsistent knowledge bases. In this paper, we view judgement aggregation as a function on propositional logic valuations, and we investigate how logic constrains judgement aggregation. In particular, we show that there is no non-dictatorial decision method for aggregating sets of judgements in a logically consistent way if the decision method is local, i.e., only depends on the individual judgements on the proposition under consideration.

1 Judges' Dilemmas

Two juridical puzzles can serve as an introduction to our investigation. The first puzzle from [7] is from the artificial intelligence literature and serves to illustrate the problems that may arise when combining multiple knowledge bases.[1] The following description of the puzzle is taken from [1].

> Once upon a time a wise but strictly formal judge questioned two witnesses. They spoke to her on separate occasions. Witness w_1 honestly stated his conviction that proposition p was true. Witness w_2 honestly stated that he believed that the implication $p \rightarrow q$ was true. Nothing else was said or heard. The judge, not noticing any inconsistency, accepted both statements and concluded that q had to be true. However, when the two witnesses heard about her conclusion they were shocked because they both were convinced that q was false. But they were too late to prevent the verdict to be executed...

*Department of Computer Science, University of Liverpool, United Kingdom; e-mail: pauly@csc.liv.ac.uk

†Department of Philosophy, University of Groningen, the Netherlands; e-mail: M.van.Hees@philos.rug.nl

1 We thank Peter van Emde Boas for drawing our attention to this puzzle.

Individually, the witnesses (or knowledge bases) are consistent, witness w_1 believing $\{p, \neg q\}$ and witness w_2 believing $\{p \to q, \neg q\}$. The judge, however, is led to a conclusion which goes against the opinions of both witnesses. Alternatively, the question arises how to combine these two collectively inconsistent knowledge bases into a consistent one. Especially when knowledge bases are distributed, situations like the one described frequently arise.

As a second juridical puzzle, consider the so-called doctrinal paradox.[2] Assume that a three-member court has to decide whether a defendant is liable under a charge of breach of contract. The judges have to make three decisions: whether the contract was valid (p), whether there was a breach (q) and whether the defendant is liable (r). In their decision making they are constrained by the legal doctrine that the defendant is only liable if the contract was valid and if there was indeed a breach ($r \leftrightarrow (p \wedge q)$). Now assume that the members of the court make the following judgements:

1: $p, \neg q, \neg r$
2: $\neg p, q, \neg r$
3: p, q, r

Note that each of these three judgement sets is compatible with the legal doctrine according to which ($r \leftrightarrow (p \wedge q)$). However, and this basically forms the doctrinal paradox, the method of majority voting will run against that doctrine. If a majority vote is held on each proposition, the court will have to decide that there was a breach of a valid contract but that the defendant is not liable, an inconsistency similar to the first juridical puzzle.

Both juridical puzzles can be viewed as instances of the more general problem of judgement aggregation. Given that several individuals make judgements on a set of interconnected propositions, how can one translate these judgements into a collective judgement?

2 Aggregation of Judgements

The doctrinal paradox, and particularly the recent generalisation of it given by List and Pettit [3], has sparked off an interest in the more general issue of how to aggregate individual judgements. Suppose each member of a group of individuals has a certain judgement concerning a set of propositions Ψ. In particular, for each proposition in Ψ an individual either accepts it or rejects it. Suppose furthermore that these individual judgements satisfy certain consistency requirements, viz., if the individual accepts a proposition, then he will reject the negation of that proposition, and he accepts the conjunction of two propositions if and only if he accepts both of its conjuncts. The question now is what kind of aggregation functions assign to each possible configuration of consistent individual judgements a consistent 'group' judgement. The doctrinal

[2] Kornhauser and Sager [2] seem to have been the first to discover the doctrinal paradox. Further contributions to this literature can be found in [3], including the breach-of-contract example used in our paper.

paradox basically shows that majority voting fails to be such an aggregation procedure.

Note that the framework in which the doctrinal paradox is analysed differs from that of social choice theory. The latter assumes that each individual has an ordering over a set of social states, the ordering describing the individual's *preferences*. In the logical framework developed by List and Pettit, and of which we here present a generalisation, the individuals do not express preferences but make statements about their *beliefs*. The logical framework can thus be seen as particularly useful for the analysis of *epistemic decision making*.

Logic plays a much more central role here than it does in standard social choice theory. Not only are certain logical meta-axioms investigated (e.g., anonymity or neutrality), but also on the object level, an aggregation function works with *logical objects*, i.e., sets of formulas or propositional valuations rather than preference relations. Besides being more purely or essentially logical, we also view logical consistency restrictions as more fundamental than constraints on individual preferences. We shall here discard the question to what extent the two frameworks can be related formally, and thus shall also not go into possible logical relationships between the results established here and the results of social choice theory. Remarks regarding the relationship between these two frameworks can be found in [4].

The main objective of this paper is to present a very general impossibility result concerning the aggregation of individual judgements. The structure of the paper is as follows. In section 3 we present our formal framework which makes use of a t-valued logic ($t > 1$). Since t may be larger than 2, we allow individuals as well as the group as a whole to express degrees of acceptance and rejection. In section 4, we present our first main result: given a particular domain assumption, we show that a dictatorship can be characterised in terms of the conditions of independence of irrelevant alternatives and responsiveness. Second, we show that in the particular case of 2-valued logic, the condition of responsiveness can be weakened yet further.

Our third main result is presented in section 5.1. We show that given a somewhat weaker domain assumption than the one used in our first theorem, a dictatorship is characterised by the condition of systematicity. In Sections 5.2 and 5.3 we investigate possible relaxations of some of the assumptions that were used in the derivation of our main results. In section 5.2, we show that possibility results emerge if the set of atomic propositions is a singleton set. In section 5.3 we examine what happens if we restrict the set of logical connectives to conjunction only. Here it is shown that any responsive aggregation procedure satisfying independence of irrelevant alternatives necessarily entails the existence of a veto-dictator, that is, there is at least one individual who is always able to ensure that a proposition is rejected.

3 Formal Framework

Let $N = \{1, 2, \ldots, n\}$ be the finite set of individual decision makers, where $|N| \geq 1$, and let Φ_0 be the (finite or infinite) set of atomic propositions p, q, etc. The set of all propositions Φ is obtained by closing Φ_0 under the standard propositional connectives of conjunction ($\wedge$) and negation ($\neg$):

$$\varphi = p \mid \neg\varphi \mid \varphi \wedge \varphi$$

for $\varphi \in \Phi$, where $p \in \Phi_0$. In general, we shall only be concerned with some nonempty set of *relevant propositions* $\Psi \subseteq \Phi$, the propositions on which the decision makers need to make a judgement. We let $\Psi_0 = \Psi \cap \Phi_0$ be the set of atomic propositions in Ψ. Again, Ψ may be finite or infinite, and Ψ_0 may or may not include Φ_0. A *literal* φ is an atomic proposition under 0 or more negations. Formally, let $\neg^0 p = p$ and $\neg^{k+1} p = \neg\neg^k p$. Then φ is a literal if $\varphi = \neg^k p$ for some $k \geq 0$ and $p \in \Phi_0$. We will consider two literals different if they do not involve the same atomic proposition. Formally, we say that $\varphi \approx \psi$ iff $\varphi = \neg^k \psi$ or $\psi = \neg^k \varphi$ for some $k \geq 0$.

We will be working in the framework of many-valued logic [8], drawing in particular on Post's many valued systems [6]. Let $T = \{0, 1, \ldots, t-1\}$ be the set of truth values, where we assume that $|T| = t > 1$. Intuitively, we may think of $t-1$ as "true", "agree" or "accept", and of 0 as "false", "disagree" or "reject". The values between 0 and $t-1$ (if there are any) then represent degrees of truthfulness (agreement or acceptance).

A *valuation* is a function $v : \Phi_0 \to T$, and we let V be the set of all valuations. We extend a valuation v by induction to a function $\hat{v} : \Phi \to T$ which assigns truth values also to complex propositions:

$$\begin{aligned} \hat{v}(p) &= v(p) \text{ for } p \in \Phi_0 \\ \hat{v}(\neg\varphi) &= \hat{v}(\varphi) + 1 (mod\, t) \\ \hat{v}(\varphi \wedge \psi) &= min(\hat{v}(\varphi), \hat{v}(\psi)) \end{aligned}$$

We shall usually identify v and $\hat{v}$, simply writing $v(\varphi)$ instead of $\hat{v}(\varphi)$. Note that for the particular case of 2-valued logic with $T = \{0, 1\}$, the connectives defined do indeed correspond to standard negation and conjunction. Furthermore, it is well-known that in this setting, $\{\neg, \wedge\}$ is a functionally complete set of connectives, i.e., every boolean truth-function can be defined using only these two connectives. For $|T| > 2$, there are various ways one might define negation and conjunction, and indeed, our definition of negation may not be the most natural one. However, it has been shown in [6] that in general, i.e., also for t-valued logic with $t > 2$, $\{\neg, \wedge\}$ as defined above is functionally complete. Hence, every t-valued truth function can be defined using only these two connectives, and also alternative forms of negation and conjunction can be obtained using the connectives defined above (see [8] for more details).

Valuations naturally give rise to judgement sets. Given valuation v, let $\Psi(v) = \{\varphi \in \Psi | v(\varphi) = t - 1\}$. For $|T| = 2$, in the terminology of [3], any $\Psi(v)$ is a complete, consistent and deductively closed set of propositions. Conversely, for any complete, consistent and deductively closed set of propositions $X \subseteq \Psi$ there will be a valuation v such that $X = \Psi(v)$. Hence, our formal framework formulated in terms of propositional valuations generalises the framework used in [3].

An *aggregation function* $A : V^N \to V$ returns for every profile of valuations $(v_1, \ldots, v_n)$ an aggregated valuation $A(v_1, \ldots, v_n)$. In this paper we shall always assume that aggregation functions have *universal domain*, i.e., they are defined on all possible valuation profiles.

Anonymity *For any permutation $f : N \to N$, any valuation profile $v_1, \ldots, v_n \in V$ and any proposition $\varphi \in \Psi$, $A(v_1, \ldots, v_n)(\varphi) = A(v_{f(1)}, \ldots, v_{f(n)})(\varphi)$.*

A *decision method* $D : T^N \to T$ is an n-ary function which maps a profile of n individual decisions to an aggregated decision. For $x \in T^N$, we usually write x^i for $x(i)$. By treating voting procedures as decision methods, one can obtain elegant equational characterisations of, for instance, democracy, as shown in [5]. Here, we are interested in linking decision methods to the aggregation of sets of judgements.

Systematicity *There is some decision method D such that for all $v_1, \ldots, v_n \in V$ and $\varphi \in \Psi$, $A(v_1, \ldots, v_n)(\varphi) = D(v_1(\varphi), \ldots, v_n(\varphi))$.*

Theorem 1 (List & Pettit, [3]) *Let $\{p, q, p \wedge q, \neg(p \wedge q)\} \subseteq \Psi$ and $|T| = 2$. Then there is no aggregation function satisfying both anonymity and systematicity.*

This result can be seen as a generalisation of the doctrinal paradox. For $|T| = 2$ and $|N|$ odd, consider the decision method D of majority voting, where $D(x) = 1$ if and only if $x^i = 1$ for the majority of $i \in N$. Furthermore, let the aggregated valuation function $A(v_1, \ldots, v_n)$ be defined as majority voting (on atomic propositions), i.e., for $p \in \Phi_0$, $A(v_1, \ldots, v_n)(p) = D(v_1(p), \ldots, v_n(p))$. The doctrinal paradox demonstrates that while this aggregation function A is systematic for atomic propositions, it fails to be systematic (for the given D) for complex propositions like conjunctions. Theorem 1 generalises this observation: Not only majority voting fails to satisfy anonymity and systematicity, *any* aggregation procedure will fail to satisfy at least one of these two properties.

While anonymity might not seem too problematic, systematicity, however, is a rather strong property to demand of aggregation functions. Consequently, one may wonder whether relaxing this requirement can turn the impossibility theorem into a possibility theorem.

4 Characterising Dictatorship

Systematicity is a condition which requires decision making to be uniform across all relevant propositions, i.e., (1) the decision method depends only on the individual judgements regarding the formula under consideration, and (2) the same decision method is used for all relevant formulas. The following notion relaxes the second requirement, as the subsequent lemma shows.

Independence of Irrelevant Alternatives (IIA) *For all* $v_1, \ldots, v_n, v'_1, \ldots, v'_n \in V$ *and* $\varphi \in \Psi$, $A(v_1, \ldots, v_n)(\varphi) = A(v'_1, \ldots, v'_n)(\varphi)$ *whenever for all* $i \in N$ $v_i(\varphi) = v'_i(\varphi)$.

Lemma 1 *An aggregation function* A *is independent of irrelevant alternatives if and only if for every* $\varphi \in \Psi$ *there is some decision method* D_φ *such that for all* $v_1, \ldots, v_n \in V$, $A(v_1, \ldots, v_n)(\varphi) = D_\varphi(v_1(\varphi), \ldots, v_n(\varphi))$.

Proof. Assume that A satisfies IIA. For $\varphi \in \Psi$ and $x_1, \ldots, x_n \in T$, let $D_\varphi(x_1, \ldots, x_n) = A(w_{x_1}, \ldots, w_{x_n})(\varphi)$, where $w_{x_i} \in V$ is some valuation for which $w_{x_i}(\varphi) = x_i$, if such a valuation exists. Otherwise, w_{x_i} can be arbitrary.

Now for any $v_1, \ldots, v_n \in V$, $A(v_1, \ldots, v_n)(\varphi) = A(w_{x_1}, \ldots, w_{x_n})(\varphi)$ for $x_i = v_i(\varphi)$, by IIA, since $v_i(\varphi) = w_{x_i}(\varphi)$. Hence, by definition of D_φ, $A(v_1, \ldots, v_n)(\varphi) = D_\varphi(x_1, \ldots, x_n) = D_\varphi(v_1(\varphi), \ldots, v_n(\varphi))$. This proves the nontrivial direction of the claim. □

The preceding lemma should be compared to the systematicity condition. Whereas systematicity guarantees a single decision method across all propositions, IIA allows for the decision method to depend on the formula under consideration. Hence, systematicity is logically stronger than IIA.

The following lemma summarises some simple consequences of the IIA condition which will be used in the proof of our main result, theorem 2.

Lemma 2 *Consider a decision method* D_φ *and an aggregation function* A *such that for all* $\varphi \in \Psi$ *and* $v_1, \ldots, v_n \in V$, $A(v_1, \ldots, v_n)(\varphi) = D_\varphi(v_1(\varphi), \ldots, v_n(\varphi))$. *Then the following properties hold:*

1. *For every literal* φ *such that* $\varphi, \neg^1\varphi, \ldots, \neg^k\varphi \in \Psi$ *and* $x \in T^N$, *we have* $D_{\neg^k\varphi}(x^1 + k \ (mod\, t), \ldots, x^n + k \ (mod\, t)) = D_\varphi(x^1, \ldots, x^n) + k \ (mod\, t)$.
2. *For all literals* $\varphi \not\approx \psi$ *such that* $\varphi, \psi, \varphi \wedge \psi \in \Psi$ *and* $x, y \in T^N$, *we have* $min(D_\varphi(x^1, \ldots, x^n), D_\psi(y^1, \ldots, y^n)) = D_{\varphi\wedge\psi}(min(x^1, y^1), \ldots, min(x^n, y^n))$.
3. *For all literals* $\varphi \not\approx \psi$ *such that* $\varphi, \psi, \varphi \wedge \psi \in \Psi$ *and* $x \in T^N$, *if* $D_{\varphi\wedge\psi}(x) \neq 0$ *then* $D_\varphi(x) \neq 0$ *and* $D_\psi(x) \neq 0$.

Proof. For the first claim, consider any $x = (x^1, \ldots, x^n) \in T^N$, and let $v_i(\varphi) = x^i$. Then $v_i(\neg^k\varphi) = x^i + k \ (mod\, t)$, and hence $D_{\neg^k\varphi}(x^1 + k \ (mod\, t), \ldots, x^n + k \ (mod\, t)) = D_{\neg^k\varphi}(v_1(\neg^k\varphi), \ldots, v_n(\neg^k\varphi))$ which by our assumption must equal $v(\neg^k\varphi) = A(v_1, \ldots, v_n)(\neg^k\varphi)$. Using the defining properties of valuations, we have $v(\neg^k\varphi) = v(\varphi) + k \ (mod\, t) = D_\varphi(v_1(\varphi), \ldots, v_n(\varphi)) + k \ (mod\, t)$ which in turn equals $D_\varphi(x^1, \ldots, x^n) + k \ (mod\, t)$.

The second claim can be established similarly: Consider any $x = (x^1, \ldots, x^n)$ and $y = (y^1, \ldots, y^n)$, and let $v_i(\varphi) = x^i$ and $v_i(\psi) = y^i$. Then $min(D_\varphi(x), D_\psi(y)) = min(D_\varphi(v_1(\varphi), \ldots, v_n(\varphi)), D_\psi(v_1(\psi), \ldots, v_n(\psi)))$ which by assumption must equal $min(v(\varphi), v(\psi)) = v(\varphi \wedge \psi)$. Again by our assumption, this term must equal $D_{\varphi\wedge\psi}(min(v_1(\varphi), v_1(\psi)), \ldots, min(v_n(\varphi), v_n(\psi)))$ which by definition equals $D_{\varphi\wedge\psi}(min(x^1, y^1), \ldots, min(x^n, y^n))$.

For the third claim, if $D_{\varphi\wedge\psi}(x) \neq 0$ then $D_{\varphi\wedge\psi}(x) = min(D_\varphi(x), D_\psi(x)) \neq 0$ by the second claim, which implies that $D_\varphi(x) \neq 0$ and $D_\psi(x) \neq 0$. □

Dictatorship *There is some $i \in N$ such that for all $v_1, \ldots, v_n \in V$ and $\varphi \in \Psi$, $A(v_1, \ldots, v_n)(\varphi) = v_i(\varphi)$.*

Responsiveness *There are two literals $\varphi_1, \varphi_2 \in \Psi$ with $\varphi_1 \not\approx \varphi_2$ such that for every $i \in \{1, 2\}$ there are $v_1, \ldots, v_n, v'_1, \ldots, v'_n \in V$ such that $A(v_1, \ldots, v_n)(\varphi_i) = 0$ and $A(v'_1, \ldots, v'_n)(\varphi_i) \neq 0$.*

We call the set of formulas Ψ *atomically closed* if the following three conditions are met: (1) If $\varphi \in \Psi$ and $p \in \Phi_0$ occurs in φ, then $p \in \Psi$, (2) if $p \in \Psi_0$ then $\neg^k p \in \Psi$ for all $k < t$, and (3) if two literals $l, l' \in \Psi$ then $l \wedge l' \in \Psi$. As a simple example, Ψ is atomically closed if $\Psi = \Phi$, but it is also easy to construct strict subsets of Φ which are atomically closed.

Lemma 3 *Let Ψ be atomically closed with $|\Psi_0| \geq 2$. If A satisfies IIA and responsiveness, then for all literals $\varphi \in \Psi$, $A(v_1, \ldots, v_n)(\varphi) = 0$ whenever $v_i(\varphi) = 0$ for all i.*

Proof. Suppose to the contrary that for some responsive aggregation function A satisfying IIA and for some literal $\varphi \in \Psi$ we have $A(v_1, \ldots, v_n)(\varphi) \neq 0$ while $v_i(\varphi) = 0$ for all i. By lemma 1, $D_\varphi(0, \ldots, 0) \neq 0$. Furthermore, by responsiveness, there is some literal $l \in \Psi$ with $l \not\approx \varphi$ such that for some $x \in T^N$ we have $D_l(x) = 0$ and for some $x' \in T^N$ we have $D_l(x') \neq 0$. Using lemma 2.2, we can conclude on the one hand that $D_{\varphi\wedge l}(0, \ldots, 0) = 0$ (since $D_l(x) = 0$), and on the other hand that $D_{\varphi\wedge l}(0, \ldots, 0) \neq 0$ (since $D_l(x') \neq 0$), thus yielding a contradiction. □

Theorem 2 *Let Ψ be atomically closed with $|\Psi_0| \geq 2$. Then an aggregation function is responsive and independent of irrelevant alternatives if and only if it is a dictatorship.*

Proof. It is easy to see that a dictatorship satisfies responsiveness and IIA. For the converse direction, assume that the aggregation function A is responsive and independent of irrelevant alternatives, i.e., by lemma 1, for every $\varphi \in \Psi$ there is some decision method D_φ such that for all $v_1, \ldots, v_n \in V$,

$$A(v_1, \ldots, v_n)(\varphi) = D_\varphi(v_1(\varphi), \ldots, v_n(\varphi)), \tag{1}$$

which allows us to apply lemma 2. Suppose by reductio that A is not a dictatorship, i.e., for every $i \in N$ there is some $\varphi_i \in \Psi$ and $v_1, \ldots, v_n \in V$ such that

$A(v_1,\ldots,v_n)(\varphi_i) \neq v_i(\varphi_i)$. Since these two valuations differ on φ_i, they also must differ on some atomic proposition p_i occurring in φ_i. By atomic closure, $p_i \in \Psi$. Using equation (1), this means that for every $i \in N$ there is some D_{p_i} and some $x_i \in T^N$ such that $D_{p_i}(x_i) \neq x_i(i) = x_i^i$, as shown in the table below:

i	p_i	x_i^1	x_i^2	$\cdots$	x_i^n	$D_{p_i}(x_i)$
1	p_1	x_1^1	x_1^2	$\cdots$	x_1^n	$\neq x_1^1$
2	p_2	x_2^1	x_2^2	$\cdots$	x_2^n	$\neq x_2^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	p_n	x_n^1	x_n^2	$\cdots$	x_n^n	$\neq x_n^n$

The condition of non-dictatorship now allows us to apply a diagonalisation argument. For $0 \leq k \leq n$, let $D_\varphi[k]$ abbreviate that there are some $y_1,\ldots,y_{n-k} \in T$ such that

$$D_\varphi(\underbrace{0,\ldots,0}_{k},\underbrace{y_1,\ldots,y_{n-k}}_{n-k}) \neq 0.$$

We will show by induction on k that for all $k \leq n$ there are two literals $l, l' \in \Psi$ for which $l \not\approx l'$ such that $D_l[k]$ and $D_{l'}[k]$. Note that this will suffice to derive a contradiction: $D_l[n]$ states that $D_l(0,\ldots,0) \neq 0$ for some literal l which using equation (1) contradicts lemma 3.

For the base case, if $p, q \in \Psi_0$, we know that $D_p(0,\ldots,0) = D_q(0,\ldots,0) = 0$ by lemma 3, and hence $D_{\neg p}(1,\ldots,1) = D_{\neg q}(1,\ldots,1) = 1$ by lemma 2.1. Consequently, $D_{\neg p}[0]$ and $D_{\neg q}[0]$ hold. To prove the inductive step, assume that $D_l[k]$ and $D_{l'}[k]$ hold. Assume w.l.o.g. that $p_{k+1} \not\approx l$ and that $D_l(0,\ldots,0,y_1,\ldots,y_{n-k})$ $\neq 0$. Suppose that $x_{k+1}^{k+1} + a(mod\, t) = 0$, where $0 \leq a < t$. Then using lemma 2.1 we have

$$\begin{aligned} &D_{\neg^a p_{k+1}}(x_{k+1}^1 + a(mod\, t),\ldots,x_{k+1}^n + a(mod\, t)) \\ &\quad = D_{p_{k+1}}(x_{k+1}^1,\ldots,x_{k+1}^n) + a(mod\, t) \\ &\quad \neq x_{k+1}^{k+1} + a(mod\, t) = 0. \end{aligned}$$

By lemma 2.2, $D_{l\wedge\neg^a p_{k+1}}(\overbrace{0,\ldots,0,0}^{k+1}, min(y_2, x_{k+1}^{k+2} + a(mod\ t)),\ldots,$ $min(y_{n-k}, x_{k+1}^n + a(mod\, t))) \neq 0$, and by lemma 2.3, we can conclude from this that $D_l[k+1]$ and $D_{\neg^a p_{k+1}}[k+1]$ hold, completing the inductive step. □

Inspection of the proof of the previous theorem reveals that responsiveness is only used through lemma 3. Hence, we also have the following result: An aggregation function A satisfying IIA for which $A(v_1,\ldots,v_n)(\varphi) = 0$ whenever $v_i(\varphi) = 0$ for all i must be a dictatorship.

Phrased as an impossibility theorem, the result states that as long as the set of relevant propositions is atomically closed and contains at least two atomic propositions, there is no aggregation function which is responsive, non-dictatorial and independent of irrelevant alternatives. While the domain condition of this result (atomic closure) is more restrictive than the domain condition of theorem 1, the condition of independence of irrelevant alternatives is much weaker than systematicity. Furthermore, our result is formulated for general many-valued logic rather than 2-valued logic. Consequently, even allowing individuals to be undecided about certain propositions (by allowing for a third truth value) will not allow us to escape from an impossibility result. However, the next result shows that for the special case of 2-valued logic, the condition of responsiveness can be replaced by the following weaker condition.

Weak Responsiveness *For at least some* $v_1, \ldots, v_n, v'_1, \ldots, v'_n \in V$ *and some* $\varphi \in \Psi$*, we have* $A(v_1, \ldots, v_n)(\varphi) \neq A(v'_1, \ldots, v'_n)(\varphi)$*.*

Theorem 3 *Let* Ψ *be atomically closed with* $|\Psi_0| \geq 2$*. If* $|T| = 2$*, an aggregation function is weakly responsive and independent of irrelevant alternatives if and only if it is a dictatorship.*

Proof. As was remarked earlier, responsiveness was only used in the proof of theorem 2 through lemma 3. Hence it suffices to show that for $|T| = 2$, responsiveness can be weakened in lemma 3, i.e., we will show that if A satisfies IIA and weak responsiveness, then for all literals $\varphi \in \Psi$, $A(v_1, \ldots, v_n)(\varphi) = 0$ whenever $v_i(\varphi) = 0$ for all i.

If an aggregation function A is weakly responsive, there must be some $\psi \in \Psi$ and some $v_1, \ldots, v_n, v'_1, \ldots, v'_n \in V$ such that $A(v_1, \ldots, v_n)(\psi) \neq A(v'_1, \ldots, v'_n)(\psi)$. Hence, there must also be a detectable difference on some atomic $p \in \Psi_0$ occurring in ψ, so using lemma 1, we can assume that for some $x \in \{0,1\}^N$ we have $D_p(x) = 0$ and for some $x' \in \{0,1\}^N$ we have $D_p(x') = 1$.

Now assume to the contrary that (using lemma 1) there is some literal $\varphi \in \Psi$ such that

$$D_\varphi(0, \ldots, 0) \neq 0. \tag{2}$$

We distinguish two cases. First, if $p \not\approx \varphi$, then using lemma 2.2, from (2) we can conclude on the one hand that $D_{\varphi \wedge p}(0, \ldots, 0) = 0$ (since $D_p(x) = 0$), and on the other hand that $D_{\varphi \wedge p}(0, \ldots, 0) \neq 0$ (since $D_p(x') = 1$), thus yielding a contradiction.

Second, assume that $\varphi = \neg^k p$ for $k \in \{0, 1\}$, and consider some $q \neq p$. We know that there must be some $y \in T^N$ and some $c \in \{0,1\}$ such that $D_{\neg^c q}(y) \neq 0$. Hence, applying lemma 2.2 and (2), we can conclude that $D_{\varphi \wedge \neg^c q}(0, \ldots, 0) \neq 0$. On the other hand, since there is some z such that $D_{\neg^k p}(z) = 0$, we have $D_{\varphi \wedge \neg^c q}(0, \ldots, 0) = min(D_\varphi(z), D_{\neg^c q}(0, \ldots, 0)) = 0$, so again, we have a contradiction. □

This result shows that for 2-valued logic, theorem 2 can be strengthened considerably: Any aggregation function satisfying IIA which depends on individual judgements *in any way whatsoever* will be a dictatorship.

5 Variations and Extensions

In this section, we consider three questions arising from theorem 2. First, we reconsider the notion of systematicity as a strengthening of the IIA assumption. We show that under this stronger assumption, we can relax the domain condition and obtain a result strictly stronger than theorem 1. Second, note that theorem 2 assumes that there are at least two atomic propositions present in the set of relevant propositions. Reconsidering this assumption, we investigate what happens if $|\Psi_0| = 1$. Third, we investigate whether reducing the expressive power of our logical language allows us to escape dictatorship.

5.1 Systematicity: Uniform Decision Methods

Due to the different domain conditions, (i.e., the requirements put on the set of relevant propositions Ψ), our theorem 2 cannot be compared directly to theorem 1. The requirement of atomic closure is stronger than the domain requirement imposed in theorem 1. However, it turns out that atomic closure can be weakened appropriately if one is prepared to replace IIA by the logically stronger notion of systematicity.

Theorem 4 *Let* $\{p, q\} \cup \{\neg^k(p \wedge q) \mid 0 \leq k < t\} \subseteq \Psi$. *Then an aggregation function is systematic if and only if it is a dictatorship.*

Proof. A dictatorship is trivially systematic. Given the domain assumption we can conversely prove (analogous to lemma 2) that for the decision method D provided by systematicity the following two equations hold:

$$D(x^1 + k\ (mod\ t), \ldots, x^n + k\ (mod\ t)) = D(x^1, \ldots, x^n) + k\ (mod\ t) \qquad (3)$$

$$min(D(x^1, \ldots, x^n), D(y^1, \ldots, y^n)) = D(min(x^1, y^1), \ldots, min(x^n, y^n)) \qquad (4)$$

We can then simplify the proof of theorem 2 as follows. The assumption of non-dictatorship will give us the same table as in the proof of theorem 2, but there is only one decision method D involved. By induction, we can establish for every $k \leq n$ that $D[k]$ holds, i.e., that for some $y_1, \ldots, y_{n-k} \in T$, $D(0, \ldots, 0, y_1, \ldots, y_{n-k}) \neq 0$.

For the base case with $k = 0$, we show that $D(t-1, \ldots, t-1) \neq 0$. Suppose that $D(t-1, \ldots, t-1) = x$. By equation (3), $D(0, \ldots, 0) = x + 1(mod\ t)$. On the other hand, $D(0, \ldots, 0) = D(min(t-1, 0), \ldots, min(t-1, 0))$ which by equation (4) equals $min(D(t-1, \ldots, t-1), D(0, \ldots, 0)) = min(x, x+1(mod\ t))$. Hence, $x + 1(mod\ t) = min(x, x + 1(mod\ t))$ which implies that $x = t - 1 \neq 0$. Hence, $D(t-1, \ldots, t-1) = t - 1$ and also $D(0, \ldots, 0) = 0$.

Next, the inductive step can be shown analogous to the original proof: Assuming $D[k]$, i.e. $D(0, \ldots, 0, y_1, \ldots, y_{n-k}) \neq 0$, we consider p_{k+1}. Suppose that $x_{k+1}^{k+1} + a(mod\,t) = 0$, where $0 \leq a < t$. Then by equation (3), we have

$$\begin{aligned} &D(x_{k+1}^1 + a(mod\,t), \ldots, x_{k+1}^n + a(mod\,t)) \\ &\quad = D(x_{k+1}^1, \ldots, x_{k+1}^n) + a(mod\,t) \\ &\quad \neq x_{k+1}^{k+1} + a(mod\,t) = 0. \end{aligned}$$

By equation (4), $D(\overbrace{0,\ldots,0,0}^{k+1}, min(y_2, x_{k+1}^{k+2} + a(mod\, t)), \ldots, min(y_{n-k}, x_{k+1}^{n} + a(mod\, t))) \neq 0$, completing the inductive step. □

Note first that theorem 1 is a corollary of theorem 4 (for the special case where $|T| = t = 2$) since a dictatorship is not anonymous. Second, the proof of theorem 4 shows how by moving from IIA to systematicity we can relax the original domain restriction of atomic closure. The domain restriction of $\{p, q\} \cup \{\neg^k(p \wedge q) \mid 0 \leq k < t\} \subseteq \Psi$ is not the only possible one. It suffices that Ψ contains two literals and their conjunction, as well as a literal with all of its up to $t - 1$ negations.

Since the domain assumption equals the one used in theorem 1, the result strengthens theorem 1 in three ways: (a) it is formulated for any t-valued logic and not just for a two-valued logic, (b) it shows that the condition of anonymity can be replaced by the much weaker condition of non-dictatorship, (c) it not only shows that systematicity is a sufficient condition for dictatorship (and thus for a violation of anonymity) but also makes clear that it is a necessary condition.

5.2 Shortage of Atomic Propositions

The doctrinal paradox assumes that there are at least two basic issues under discussion, and also the proof of theorem 2 does not succeed if $|\Psi_0| = 1$. In fact, we will show here that the theorem does not hold for $|\Psi_0| = 1$. To illustrate this, we will focus on the case where $|T| = 2$.

The following lemma can be proved easily by induction on φ. Essentially, it states that there are only 4 different unary boolean functions.

Lemma 4 *If $|T| = 2$ and $\Phi_0 = \{p\}$ then for all $\varphi \in \Phi$ one of the following four claims holds:*

1. *For all $v \in V$, $v(\varphi) = 1$.*
2. *For all $v \in V$, $v(\varphi) = 0$.*
3. *For all $v \in V$, $v(\varphi) = v(p)$.*
4. *For all $v \in V$, $v(\varphi) = 1 - v(p)$.*

Theorem 5 *There are non-dictatorial aggregation functions satisfying systematicity (and hence IIA) if $|\Phi_0| = 1$.*

Proof. Suppose that $\Phi_0 = \{p\}$. Let $|N|$ be odd, and define $D(x_1, \ldots, x_n) = 1$ if and only if $x_i = 1$ for the majority of x_i. Define $A(v_1, \ldots, v_n)(p) = D(v_1(p), \ldots, v_n(p))$. By the previous lemma, in order to prove that for all $\varphi \in \Psi$ we have $A(v_1, \ldots, v_n)(\varphi) = D(v_1(\varphi), \ldots, v_n(\varphi))$, there are only four cases to consider. (1) If φ is a tautology, the claim holds since $D(1, \ldots, 1) = 1$, (2) if φ is contradictory, the claim holds since $D(0, \ldots, 0) = 0$, (3) if φ is equivalent to p, the claim holds by definition, and (4) if φ is equivalent to $\neg p$,

$A(v_1, \ldots, v_n)(\neg p) = 1 - D(v_1(p), \ldots, v_n(p))$ which equals $D(v_1(\neg p), \ldots, v_n(\neg p))$ by definition of D. $\square$

Dictatorship and majority voting are not the only aggregation functions satisfying systematicity in case $|\Phi_0| = 1$. In general, as the proof of the theorem shows, any decision method D for which $D(0, \ldots, 0) = 0$ and which satisfies equation (3) will do.

The comparison between theorems 5 and 4 shows that for more than two issues under consideration, dictatorship is the only uniform and logically consistent decision method. On the other hand, if there is only one issue under consideration, many different decision methods including majority voting are possible.

5.3 Shortage of Logical Connectives

On yet a different interpretation of theorem 2 which focuses on the domain of relevant propositions Ψ, a dictatorship is characterised as the unique aggregation function which is responsiveness and independent of irrelevant alternatives across both connectives, negation and conjunction. More precisely, both negated formulas and conjunctions need to be present in Ψ for the theorem to hold. This raises the question of whether restricting the set of allowable connectives would allow one to obtain aggregation functions other than dictatorships.

As an example, consider the case where $|T| = 2$ and n is odd. Then majority voting will be a decision method D satisfying the following equation

$$A(v_1, \ldots, v_n)(\varphi) = D(v_1(\varphi), \ldots, v_n(\varphi)) \tag{5}$$

for all $\varphi \in \Phi$ which contain negation as the only connective, whereas consensus voting (i.e., $D(x) = 1$ iff for all $i \in N$ $x^i = 1$) will not. On the other hand, consensus voting will be a non-dictatorial rule that satisfies equation (5) for all $\varphi \in \Phi$ which contain conjunction as the only connective, whereas majority voting will not. It is easy to characterise the decision methods corresponding to various sets of connectives using conditions like those in lemma 2. The more interesting question is whether certain intuitively natural classes of aggregation functions are characterised by intuitively natural sets of connectives. In this light, theorem 2 is an example of such a characterisation result for the set $\{\neg, \wedge\}$.

Call an aggregation function A a *veto-dictatorship* iff there is some $i \in N$ such that for all $v_1, \ldots, v_n \in V$ and $\varphi \in \Psi$, $A(v_1, \ldots, v_n)(\varphi) = 0$ if $v_i(\varphi) = 0$. Note that every dictatorship is also a veto-dictatorship. We call a set of propositions Ψ *conjunctively closed* in case (1) if $\varphi \in \Psi$ and $p \in \Phi_0$ occurs in φ, then $p \in \Psi$, and (2) if two literals $l, l' \in \Psi$ then $l \wedge l' \in \Psi$. Hence, every atomically closed set of propositions is also conjunctively closed, but not vice versa.

Theorem 6 *Let Ψ be conjunctively closed with $|\Psi_0| \geq 2$. Then if an aggregation function is responsive and independent of irrelevant alternatives, it must be a veto-dictatorship.*

Proof. First, note that the equations of lemma 2.2 and 2.3 still hold given conjunctive closure. Second, it can easily be checked that lemma 3 remains valid if Ψ is conjunctively closed rather than atomically closed. We can then simplify the proof of theorem 2 as follows: Assuming by reductio that we do not have veto-dictatorship, we know that $D_{p_i}(x_i) \neq 0$ and $x_i^i = 0$. We show by induction on k that for all $k \leq n$ there are literals $l \not\approx l'$ such that $D_l[k]$ and $D_{l'}[k]$ hold. For the base case, we use responsiveness directly to give us two literals $l \not\approx l'$ for which $D_l[0]$ and $D_{l'}[0]$ hold. □

This result demonstrates that even a relaxation of the domain condition from atomic closure to conjunctive closure still leaves us with a (possibly weaker) form of dictatorship. Hence, restricting the logical connectives to conjunction alone does not seem to gain us much. Note, however, that since this result only establishes a necessary condition for IIA and responsiveness, it may still be the case that restricting the logical connectives to conjunction does not allow us to escape even from a strong form of dictatorship.

6 Conclusion

This paper has addressed the question of how the requirement of logical consistency constrains the aggregation of judgements. Initially, one may investigate what aggregation functions can be characterised by a uniform decision method which only takes into account the individual judgements on the proposition under consideration. This condition of systematicity, expressed by equation (5) was already shown in [3] to be very strong. As demonstrated by theorem 4, it is essentially equivalent to the presence of a dictator. As one way to allow for more flexibility in decision making, one can allow the decision method to depend on the formula under consideration, generalising equation (5) to equation (1). As our main result (theorem 2) has shown, this will still result in a dictatorship under some mild domain conditions. Since this result is formulated in the general setting of $|T|$-valued logic, allowing individuals to remain undecided on propositions does not allow one to escape this result.

As already pointed out in [3], there are other routes to escape dictatorship, giving up universal domain, or the assumption that sets of judgements can always be linked to a valuation function (and thus imposing logical consistency, etc.). Furthermore, our work suggests yet different routes. Our investigation still leaves open the possibility that weakening the domain condition of atomic closure may allow one to avoid dictatorship. In particular, theorem 6 suggests that characterisation results for restricted sets of logical connectives may be of interest.

As mentioned in the introduction, the formal framework for judgement aggregation we used is more logical than the standard framework of social choice theory which uses preferences. For this reason, we expect more fruitful interaction between research in formal logic and social choice theory. In this paper, examples of this interaction can already be seen: sections 5.2 and 5.3 which present typically logical questions, and the use of many-valued logic and its functional completeness result in section 3.

Finally, we return to the juridical puzzles from section 1. A possible solution to the first juridical puzzle is proposed in [1]. We treat witness w_1 as an expert on p and witness w_2 as an expert on $p \to q$. If the witnesses are informed about each other's expertise and accept it, they will retract their belief in $\neg q$ since they will realise that they must be mistaken. In our framework, this would mean dropping the IIA condition for a formula if it is a logical consequence of a combination of other relevant formulas. Hence, we may want to generalise our decision methods yet further, by taking into account more than the individual judgements on the proposition under consideration. From this perspective, one might put the moral of our story informally as follows: when it comes to aggregating individual judgements, sticking to the point may not be wise.

References

[1] Z. Huang. *Logics for Agents with Bounded Rationality.* PhD thesis, ILLC, University of Amsterdam, 1994.

[2] L. Kornhauser and L. Sager. Unpacking the court. *Yale Law Journal*, 96:82–117, 1986.

[3] C. List and P. Pettit. Aggregating sets of judgements: An impossibility result. *Economics and Philosophy*, 18:89–110, 2002.

[4] C. List and P. Pettit. Aggregating sets of judgements: Two impossibility results compared. *Synthese*, forthcoming.

[5] Y. Murakami. *Logic and Social Choice.* Dover, 1968.

[6] E. Post. Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, 43:163–185, 1921.

[7] W. Schoenmakers. A problem in knowledge acquisition. *SIGART Newsletter*, 95:56–57, 1986.

[8] A. Urquhart. Many-valued logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume III, pages 71–116. Reidel, first edition, 1986.