Judgment Aggregation under Issue Dependencies

Marco Costantini
University of Amsterdam
The Netherlands
marcostantini2008@gmail.com

Carla Groenland
University of Amsterdam
The Netherlands
carla.groenland@gmail.com

Ulle Endriss
University of Amsterdam
The Netherlands
ulle.endriss@uva.nl

Abstract

We introduce a new family of judgment aggregation rules, called the binomial rules, designed to account for hidden dependencies between some of the issues being judged. To place them within the landscape of judgment aggregation rules, we analyse both their axiomatic properties and their computational complexity, and we show that they contain both the well-known distance-based rule and the basic rule returning the most frequent overall judgment as special cases. To evaluate the performance of our rules empirically, we apply them to a dataset of crowdsourced judgments regarding the quality of hotels extracted from the travel website TripAdvisor. In our experiments we distinguish between the full dataset and a subset of highly polarised judgments, and we develop a new notion of polarisation for profiles of judgments for this purpose, which may also be of independent interest.

1 Introduction

Judgment aggregation deals with the question of how to best merge the judgments made by several agents regarding a number of issues into a single consensus judgment for each of these issues (List and Puppe 2009; Grossi and Pigozzi 2014). It generalises preference aggregation (Dietrich and List 2007b), and is closely related to a body of work on belief merging in AI (Everaere, Konieczny, and Marquis 2015). While originating in legal theory, philosophy, and economics (List and Pettit 2002), several recent contributions have focused on the algorithmic dimension of judgment aggregation and emphasised its relevance to a variety of applications associated with AI, such as decision making in multiagent systems and the aggregation of information gathered by means of crowdsourcing (Endriss 2016).

In this paper, we introduce a new family of judgment aggregation rules aimed at capturing a phenomenon that hitherto has not received any explicit attention in the literature, namely the fact that when many individuals agree on how to judge certain subsets of issues, then this may suggest that there are hidden dependencies between those issues. To illustrate the idea, consider the following example. A group of 23 gastro-entertainment professionals need to decide on the ideal meal to serve at a party. They propose one dish and one drink each, leading to the following profile:

	Chips	Beer	Caviar	Champagne
11 individuals	1	1	0	0
10 individuals	0	0	1	1
2 individuals	0	1	1	0

How should they decide? Any feasible outcome consists of exactly one dish and exactly one drink, i.e., whatever rule they use to aggregate their judgments should respect this *integrity constraint*. If they use the *majority rule* to decide, they will end up serving beer and caviar. This outcome happens to respect the integrity constraint for this particular profile, but otherwise—intuitively speaking—it does not seem a good choice. Unfortunately, no existing approach to judgment aggregation—including, e.g., premise-based rules (Dietrich and Mongin 2010), quota rules (Dietrich and List 2007a), distance-based rules (Miller and Osherson 2009), rules inspired by voting theory (Lang et al. 2011), scoring rules (Dietrich 2014), and representative-voter rules (Endriss and Grandi 2014)—allows us to capture this intuition.

The aggregation rules we propose, the *binomial rules*, assign scores to patterns of accepted/rejected issues. In our example, one such pattern with a lot of support (and thus a high score) is the joint acceptance of chips and beer. We show that, by varying the range of patterns of interest, we can capture two well-known rules and many interesting rules in between. To better place our rules relative to known rules, we analyse both their axiomatic properties and their computational complexity. Finally, we evaluate some of our rules empirically, by testing them on judgments extracted from a dataset of hotel reviews. Our rules can be expected to work particularly well on highly polarised profiles of judgments. To substantiate this point, we propose a new *definition of polarisation*, which we believe to be of independent interest.

The remainder of this paper is structured as follows. In Section 2 we define the binomial rules and study their theoretical properties. In Section 3 we develop our notion of polarisation, introduce the *compliant-reviewer problem* as a means for evaluating aggregation rules, and compare two of our rules with the majority rule. Section 4 concludes.

2 Binomial Rules

In this section we introduce a class of aggregation rules that try to account for observed dependencies between issues. The formal framework we work in is *binary aggregation*

with integrity constraints (Grandi and Endriss 2011), but our rules can easily be adapted to related frameworks, notably the formula-based framework of List and Pettit (2002).

2.1 Formal Framework

Let $\mathcal{I} = \{1, \ldots, m\}$ be a finite set of binary *issues* on which to take a decision. We may think of each issue as a yes/no-question. A *ballot* is a vector $B = (b_1, \ldots, b_m) \in \{0,1\}^m$, indicating for each issue $j \in \mathcal{I}$ whether it is *accepted* ($b_j = 1$) or *rejected* ($b_j = 0$). We associate a set $\{p_1, \ldots, p_m\}$ of propositional variables with $\mathcal{I}$ and refer to formulas Γ in the corresponding propositional language as *integrity constraints*. For example, the integrity constraint $\Gamma = (p_1 \wedge p_2 \rightarrow p_3)$ expresses that when you accept the first two issues, then you must also accept the third. Note that models for such formulas are isomorphic to ballots. For a given integrity constraint Γ, we say that a ballot B is *rational* if it satisfies Γ, i.e., if $B \in \text{Mod}(\Gamma)$.

Let $\mathcal{N} = \{1, \ldots, n\}$ be a finite set of *agents*. Suppose each of them reports a rational ballot (w.r.t. some integrity constraint Γ), giving rise to a *profile* $\boldsymbol{B} = (B_1, \ldots, B_n)$. We write b_{ij} for the judgment regading the jth issue in B_i, the ballot submitted by agent i.

Using an *aggregation rule*, we want to map every such profile into a single consensus ballot. For example, in a scenario with three agents, three issues, and integrity constraint $\Gamma = (p_1 \wedge p_2 \rightarrow p_3)$, we may get the following profile

$$B_1 = (1,1,1), \quad B_2 = (1,0,0), \quad B_3 = (0,1,0)$$

Each of the three ballots is rational. Nevertheless, if we use the issue-wise *majority rule* to map this into a single ballot, we obtain $(1,1,0)$, which is not rational. This is the well-known *doctrinal paradox* (Kornhauser and Sager 1993; List and Pettit 2002). In this paper, we are only interested in aggregation rules that are *collectively rational*, i.e., where the outcome is rational whenever the input profile is. Ideally, we get a single ballot as the outcome, but in practice there could be ties. Therefore, given an integrity constraint Γ, we define an aggregation rule as a function $F : \text{Mod}(\Gamma)^n \rightarrow \mathcal{P}(\text{Mod}(\Gamma)) \setminus \{\emptyset\}$, mapping every rational profile to a nonempty set of rational ballots.

In the sequel, we will sometimes refer to *preference aggregation* rules and their generalisations to judgment aggregation. The basic idea of embedding preference aggregation into judgment aggregation is that we may think of issues as representing questions such as "is A better than B?" and can ensure that rational ballots correspond to well-formed preference orders by choosing an appropriate integrity constraint, encoding properties of linear orders, such as transitivity (Dietrich and List 2007b; Grandi and Endriss 2011).

2.2 A Family of Rules

Define the *agreement* between two ballots $B, B' \in \{0,1\}^m$ as the number of issues on which they coincide:

$$\text{Agr}(B, B') = \#\{j \in \mathcal{I} \mid b_j = b'_j\}$$

Equivalently, $\text{Agr}(B, B')$ is the difference between m and the Hamming distance between B and B'.

For our rules, the idea is to score each rational ballot $B^\star$ in terms of how well it agrees with the individual ballots in the profile on certain subsets of issues. We parametrise our rules in terms of the sizes of such subsets to consider. Let $K \subseteq \{1, \ldots, m\}$ be such a set of possible sizes. If we award one point for each subset I (the size of which is in K) and each agent $i \in \mathcal{N}$ such that the ballot of i fully agrees with $B^\star$ on all issues in I, we obtain the following score:[1]

$$\sum_{I \subseteq \mathcal{I} \text{ s.t. } \#I \in K} \#\{i \in \mathcal{N} \mid \forall j \in I. b_{ij} = b_j^\star\} = \sum_{B \in \boldsymbol{B}} \sum_{k \in K} \binom{\text{Agr}(B, B^\star)}{k}$$

In fact, we will consider a generalisation of this definition, where we may ascribe different levels of importance to the agreement with issue sets of different sizes.

Definition 1. *For every index set $K \subseteq \{1, \ldots, m\}$ and weight function $w : K \rightarrow \mathbb{R}^+$, the corresponding **binomial rule** is the aggregation rule $F_{K,w}$ mapping any given profile $\boldsymbol{B} \in \text{Mod}(\Gamma)^n$ to the following set:*

$$F_{K,w}(\boldsymbol{B}) = \underset{B^\star \in \text{Mod}(\Gamma)}{\text{argmax}} \sum_{B \in \boldsymbol{B}} \sum_{k \in K} w(k) \cdot \binom{\text{Agr}(B, B^\star)}{k}$$

Thus, for every ballot $B^\star$ we compute a score by adding, for every ballot B in the profile $\boldsymbol{B}$ and every number $k \in K$, $w(k)$ points for every set of k issues that $B^\star$ agrees on with B. $F_{K,w}$ returns all rational ballots that maximise this score. Note that $F_{K,w}$ depends also on the integrity constraint Γ, even if this is not explicit in our notation. We omit w in case $w : k \mapsto 1$ for all $k \in K$. In case K is a singleton $\{k\}$, we call the resulting rule a *binomial-k rule*. Observe that for binomial-k rules the weight function w is irrelevant.

The family of binomial-k rules includes two known rules as special cases. The first is $F_{\{1\}}$, which returns the rational ballots that maximise the sum of agreements with the individual ballots. Miller and Osherson (2009) call this the *prototype rule*, although more often it is simply referred to as "the" *distance-based rule* (Pigozzi 2006; Endriss, Grandi, and Porello 2012), as it is the rule that minimises the sum of the Hamming distances between individual ballots in the profile. $F_{\{1\}}$ thus is the generalisation of *Kemeny's rule* for preference aggregation (Kemeny 1959) to judgment aggregation. In the absence of an integrity constraint (formally: for $\Gamma = \top$), $F_{\{1\}}$ reduces to the issue-wise majority rule.

The second case of a known binomial-k rule is the rule for $k = m$. $F_{\{m\}}$ awards a point to $B^\star$ for every ballot B in the input profile that it agrees with entirely. In other words, it returns the ballot(s) that occur(s) most frequently in the profile. Thus, $F_{\{m\}}$ is a so-called *representative-voter rule* (Endriss and Grandi 2014). These are rules that first determine which of the agents is "most representative" and then return that agent's ballot as the outcome. Other examples for representative-voter ruls are the *average-voter*

---

[1] For the simplified term on the righthand side, we sum over all $B \in \boldsymbol{B}$ rather than over all $i \in \mathcal{N}$ (which would amount to the same thing). The reason we prefer this notation here is that it allows us to apply the same formula for computing scores for varying groups of agents, and not just some fixed $\mathcal{N}$. This will be helpful later on (cf. reinforcement, in Section 2.3).

rule (choosing the ballot from the profile that maximises the sum of agreements with the rest) and the *majority-voter rule* (choosing the ballot from the profile that is closest to the majority outcome). For consistency with this naming scheme, we suggest the name *plurality-voter rule* for $F_{\{m\}}$. Hartmann and Sprenger (2012) have proposed this rule under the name of *situation-based procedure* and argued for it on epistemic grounds, showing that it has good truth-tracking properties. For the special case of preference aggregation, the same rule has been advocated by Caragiannis, Procaccia, and Shah (2014) under the name of *modal ranking rule*, also on the basis of its truth-tracking performance. Thus, the family of binomial-k rules subsumes two important (and very different) aggregation rules from the literature. By varying k, we obtain a spectrum of new rules in between.

For $K = [m] = \{1, \ldots, m\}$, i.e., for rules $F_{[m],w}$ that take agreements of all sizes into account, we will consider two weight functions, intended to normalise the effect each binomial has. The first, $w_1 : k \mapsto 1/\binom{m}{k}$, is chosen so as to ensure that each binomial contributes a term between 0 and 1. We call the resulting rule $F_{[m],w_1}$ the *normalised binomial rule*. It, to some extent, compensates for the fact that larger subsets of agreement have a disproportionally high impact on the outcome. How high is this impact exactly? When no weighting function is used, the identity $\sum_{k=1}^{\ell} \binom{\ell}{k} = 2^{\ell} - 1$, together with $\binom{\ell}{k} = 0$ for $k > \ell$, shows that in fact $F_{[m]}$ selects those ballots $B^\star$ that maximise the sum $\sum_{B \in \boldsymbol{B}} 2^{\mathrm{Agr}(B,B^\star)}$. For this reason, we will also experiment with a second weighting function that discounts large subsets of agreements even more strongly than w_1. We call the rule $F_{[m],w_2}$ with $w_2 : k \mapsto 1/m^k$ the *exponentially normalised binomial rule*.

2.3 Axiomatic Properties

A central question in all of social choice theory, including judgment aggregation, is what normative properties, i.e., what *axioms*, a given aggregation rule satisfies (Arrow, Sen, and Suzumura 2002; List and Pettit 2002; Endriss 2016). In this section, we test our novel rules against a number of standard axioms and find that they stand out for satisfying both *collective rationality* and *reinforcement*.

First, all binomial rules are clearly *anonymous:* all individuals are treated symmetrically, i.e., outcomes are invariant under permutations of the ballots in a profile.

Second, binomial rules do *not* satisfy the standard formulation of the *neutrality* axiom (Endriss 2016), which requires that issues are treated symmetrically: if, for two issues, everyone either accepts them both or rejects them both, then the rule should also either accept or reject both. A counterexample can be constructed by picking an integrity constraint such that only the seven ballots $(0,0,1,0,0)$, $(1,1,1,0,0)$, $(0,0,0,1,0)$, $(1,1,0,1,0)$, $(0,0,0,0,1)$, $(1,1,0,0,1)$, and $(1,0,0,0,0)$ are allowed. For the profile containing the first six ballots once, $F_{\{1\}}$ accepts only $(1,0,0,0,0)$ which contradicts neutrality when looking at the first two issues. We believe that this indicates a deficiency with this standard formulation of neutrality rather than with our rules,[2] as this

---

[2]Slavkovik (2014) also argues against the neutrality axiom.

standard formulation does not account for the asymmetries in the set of rational ballots induced by the integrity constraint. Note that for integrity constraints that are, in some sense, symmetric (such as the one required to model preferences) this problem does not occur. Indeed, Kemeny's rule (the counterpart to $F_{\{1\}}$) is neutral in the context of preference aggregation (Arrow, Sen, and Suzumura 2002).

Going beyond basic symmetry requirements, we may wish to ensure that, when we obtain the same outcome $B^\star$ for two different profiles (possibly tied with other outcomes), then $B^\star$ should continue to win when we join these two profiles together. This is a very natural requirement that, although it has received only little attention in judgment aggregation to date, plays an important role in other areas of social choice theory (Young 1974; Arrow, Sen, and Suzumura 2002). Note that we are now speaking of aggregation rules for possibly varying sets of agents. (In what follows, we use $\oplus$ do denote concatenation of vectors.)

Definition 2. *An aggregation rule F for varying sets of agents is said to satisfy **reinforcement** if $F(\boldsymbol{B} \oplus \boldsymbol{B}') = F(\boldsymbol{B}) \cap F(\boldsymbol{B}')$ whenever $F(\boldsymbol{B}) \cap F(\boldsymbol{B}') \neq \emptyset$.*

While very appealing, reinforcement is in fact violated by most collectively rational judgment aggregation rules that have been proposed for practical use.[3] For example, by the Young-Levenglick Theorem, amongst all preference aggregation rules that are neutral and Condorcet-consistent, Kemeny's rule is the only one that satisfies reinforcement (Young and Levenglick 1978). Hence, the generalisation of any other neutral and Condorcet-consistent preference aggregation rule to judgment aggregation must also violate reinforcement. Thus, we can exclude, e.g., the judgment aggregation rules that generalise *Slater's rule*, Tideman's *ranked-pairs rule*, or *Young's rule*, definitions of which are given by Lang and Slavkovik (2013). Endriss and Grandi (2014) define a weaker version of the reinforcement axiom and show that the majority-voter rule does not satisfy it. The average-voter rule also violates reinforcement.[4]

On the other hand, as we will see next, it is easy to verify that all binomial rules satisfy reinforcement.

Proposition 1. *Every binomial rule satisfies reinforcement.*

Proof. Let us denote the score received by ballot $B^\star$ when the binomial rule $F_{K,w}$ is applied to the profile $\boldsymbol{B}$ by $score_{K,w}(B^\star, \boldsymbol{B}) = \sum_{B \in \boldsymbol{B}} \sum_{k \in K} w(k) \cdot \binom{\mathrm{Agr}(B,B^\star)}{k}$. For any two profiles $\boldsymbol{B}$ and $\boldsymbol{B}'$ over disjoint groups, we have:

$$
\begin{aligned}
score_{K,w}(B^\star, \boldsymbol{B} \oplus \boldsymbol{B}') &= \sum_{B \in \boldsymbol{B} \oplus \boldsymbol{B}'} \sum_{k \in K} w(k) \cdot \binom{\mathrm{Agr}(B,B^\star)}{k} \\
&= \sum_{B \in \boldsymbol{B}} \sum_{k \in K} w(k) \cdot \binom{\mathrm{Agr}(B,B^\star)}{k} + \sum_{B \in \boldsymbol{B}'} \sum_{k \in K} w(k) \cdot \binom{\mathrm{Agr}(B,B^\star)}{k} \\
&= score_{K,w}(B^\star, \boldsymbol{B}) + score_{K,w}(B^\star, \boldsymbol{B}')
\end{aligned}
$$

---

[3]The majority rule and other quota rules, while easily seen to satisfy reinforcement, are collectively rational only for certain (simple) integrity constraints (Dietrich and List 2007a).

[4]To see this, consider a profile in which 10 agents pick $(1,0,0)$, 10 pick $(0,1,0)$, and 11 pick $(0,0,1)$, and a second profile in which 1 voter picks $(0,0,0)$ and 2 pick $(0,0,1)$. In both profiles $(0,0,1)$ wins, but in their union $(0,0,0)$ wins.

Hence, the rule $F_{K,w}$, which simply maximises $score_{K,w}$, must satisfy the reinforcement axiom as claimed. □

The only other judgment aggregation rules proposed in the literature we are aware of that are both collectively rational and satisfy reinforcement are the *scoring rules* of Dietrich (2014). We note that Dietrich discusses neither reinforcement nor the issue of hidden dependencies.

2.4 Computational Complexity

In order to study the complexity of the binomial rules, consider the following winner determination problem.

WINDET(F)
Instance: Integrity constr. Γ, profile $\boldsymbol{B} \in \text{Mod}(\Gamma)^n$, partial ballot $b : I \to \{0, 1\}$ for some $I \subseteq \mathcal{I}$.
Question: Is there a $B^\star \in F(\boldsymbol{B})$ s.t. $b_j^\star = b(j), \forall j \in I$?

If we are able to answer the above question, we can compute winners for F, by successively expanding I. Building on work by Hemaspaandra, Spakowski, and Vogel (2005), Grandi (2012) showed that WINDET is Θ_2^P-complete[5] for the distance-based rule in binary aggregation with integrity constraints, i.e., for our rule $F_{\{1\}}$. On the other hand, WINDET for the plurality-voter rule $F_{\{m\}}$ is immediately seen to be polynomial: one only has to consider the ballots in $\boldsymbol{B}$ and count how often each ballot occurs. The following result generalises this to binomial-k rules with k close to m.

Proposition 2. *The winner determination problem for binomial-k rules $F_{\{k\}}$ is polynomial for $(m - k) \in O(1)$.*

Proof. Let k close to m, integrity constraint Γ, and profile $\boldsymbol{B}$ be given. A ballot B can only receive a positive score if $\text{Agr}(B, B') \geqslant k$ for at least one $B' \in \boldsymbol{B}$. Hence, only ballots that differ on at most $m - k$ issues with a ballot in $\boldsymbol{B}$ can get picked, of which there are at most:

$$n[1 + m + m(m-1) + \cdots + m(m-1) \cdots (k+1)] \sim nm^{m-k}$$

For each ballot, the rule needs to do propositional-logic model checking for $|\Gamma|$ many formulas, compute agreement for m issues in n ballots, and finally compute the binomial coefficients and sum the result. But $O(|\Gamma|n^2 m^{1+(m-k)})$ is polynomial if $m - k$ is constant, so we are done. □

However, when k is a small constant, the corresponding binomial-k rule is as intractable as the Kemeny rule $F_{\{1\}}$:

Proposition 3. *The winner determination problem for binomial-k rules $F_{\{k\}}$ is Θ_2^P-complete for $k \in O(1)$.*

Proof (sketch). Θ_2^P-membership is routine. To prove Θ_2^P-hardness for all $k \in O(1)$, recall how Grandi proved it for $k = 1$ (Grandi 2012, Theorem 7.4.5). The central ingredient of his proof is to show that it is NP-hard to check whether the $F_{\{1\}}$-score of a given ballot exceeds a given number K. If we can show that the same holds for any $F_{\{k\}}$ with $k \in O(1)$, then we are done. Now take $k = 2$ and assume we have an algorithm to compute $F_{\{2\}}$-scores. Recall

that the $F_{\{2\}}$-score of a ballot is equal to the number of times that ballot agrees on a pair of issues with some ballot in the profile, while its $F_{\{1\}}$-score is equal to the number of times it agrees on a single issue with some ballot in the profile. Thus, we could use our algorithm to compute the $F_{\{1\}}$-score of a given profile $\boldsymbol{B}$ as follows: First, add one more issue to the problem and let all voters agree on accepting it, call the resulting profile $\boldsymbol{B'}$, and compute the $F_{\{2\}}$-score for $\boldsymbol{B'}$. Second, compute the $F_{\{2\}}$-score for $\boldsymbol{B}$. Then the $F_{\{1\}}$-score for $\boldsymbol{B}$ is the difference between these two numbers. Hence, computing $F_{\{2\}}$-scores is NP-hard as well. By induction, the same holds for all $F_{\{k\}}$ with $k \in O(1)$. □

The determination of the exact computational complexity of $F_{\{k\}}$ for medium values of k, as well as for $F_{K,w}$ more generally, at this point remains an open question. Our analysis strongly suggests that it will be the smallest index $k \in K$ that influences the complexity of $F_{K,w}$ most significantly.

3 Experiments

In this section we evaluate our new aggregation rules experimentally, using a collection of hotel reviews taken from TripAdvisor. To allow us to investigate how our rules perform specifically on data with strong dependencies between issues, we propose a notion of *polarisation* of a profile.

3.1 Polarisation

Polarisation occurs when there are clusters of ballots that express opposite views on the issues. An example of a polarised profile has already been given in the introduction: one cluster votes for chips and beer and the other for caviar and champagne. One could imagine there is a "latent" integrity constraint reflected by the votes that these are the only "acceptable" combinations, along with some "noise" (the two individuals voting for beer and caviar).

Polarised profiles are characterised by both *correlation* between the issues and *uncertainty* on judgments. Correlation ensures that there is at least one cluster of ballots that all differ only slightly from a particular model for the issues. Uncertainty entails the presence of a second cluster of ballots taking an opposite view on these issues. Therefore, we define the polarisation coefficient of a profile as the product of a correlation coefficient and an uncertainty coefficient.

Definition 3. *The **correlation coefficient** of profile $\boldsymbol{B}$ is:*

$$\rho_{\boldsymbol{B}} = \frac{1}{n \cdot \binom{m}{2}} \sum_{j \neq j' \in \mathcal{I}} |2 \cdot \#\{i \in \mathcal{N} \mid b_{ij} = b_{ij'}\} - n|$$

Thus, correlation is maximal for a completely unanimous profile (where all agents agree on all issues). Pairs of issues on which agents tend to make complementary judgments also increase this coefficient. Correlation is low for random profiles in which we can expect $b_{ij} = b_{ij'}$ to hold for about 50% of all agents $i \in \mathcal{N}$. The normalisation factor $n \cdot \binom{m}{2}$ ensures that the correlation coefficient ranges from 0 to 1.

The uncertainty coefficient we propose measures whether individuals tend to take opposite views on issues.

---

[5] $\Theta_2^P = \text{P}^{\text{NP}}[\log]$ is the class of decision problems that can be solved in polynomial time if we are allowed a logarithmic number of queries to an NP-oracle (Wagner 1990).

Definition 4. *The **uncertainty coefficient** of profile B is:*

$$u_B = 1 - \frac{1}{nm} \sum_{j \in \mathcal{I}} |n - 2 \cdot \#\{i \in \mathcal{N} \mid b_{ij} = 1\}|$$

Again, nm is a normalisation factor that ensures that the coefficient takes values between 0 and 1. Uncertainty is minimal for unanimous profiles and maximal when each issue is accepted by exactly 50% of the population.

Definition 5. *The **polarisation coefficient** of profile B is:*

$$\Psi_B = \rho_B \cdot u_B$$

To see how Ψ_B behaves, consider a simple profile where four individuals have to assign a value to issues 1 and 2. There are four possible ballots. If all the individuals choose the same ballot, there is no uncertainty but maximal correlation. For a profile where the four ballots cover all four possible configurations, i.e., for $B = ((0,0),(0,1),(1,0),(1,1))$, there is no correlation, while the uncertainty coefficient is equal to 1. So in both these cases polarisation is 0. To maximise polarisation, the individuals have to split into two groups of two and pick opposite ballots, either $(0,0)$ against $(1,1)$ or $(0,1)$ against $(1,0)$. Here polarisation can be seen clearly, and the product of correlation and uncertainty coefficient indeed is 1.

In related work, Can, Ozkes, and Storcken (2014) have proposed an axiomatic framework for analysing possible notions of polarisation of a preference profile. As a means of offering further support for our proposed definition, we show that it meets the basic requirements identified by Can, Ozkes, and Storcken when translated to our setting:

- **Regularity:** $\Psi_B \in [0,1]$ for all profiles B; $\Psi_B = 0$ for unanimous profiles B (as $u_B = 0$); and $\Psi_B = 1$ for profiles B in which half of the population pick some ballot B and the other half pick the exact opposite of B.

- **Neutrality:** Polarisation is not affected when we apply the same permutation on issues to all ballots.

- **Replication invariance:** $\Psi_B = \Psi_{kB}$ for all profiles B and $k \in \mathbb{N}$, where kB is the result of k times replicating B. (To see this, note that both ρ and u essentially compute which *proportion* of agents have a certain property.)

Can, Ozkes, and Storcken (2014) also propose a fourth axiom, *support independence*, which in our context would amount to saying that when one agent changes her judgment on one issue from the minority opinion to the majority opinion, then the effect this change has on the polarisation coefficient should not depend on the relative size of the majority. This is a very strong axiom (allowing Can, Ozkes, and Storcken to characterise a single polarisation coefficient as meeting all four axioms) of, arguably, less normative appeal than the other three (e.g., it seems reasonable to also permit polarisation coefficients where small changes have less impact for profiles with already very strong majorities and more impact for relatively balanced profiles).

3.2 Description of the Data

We use a dataset of hotel reviews extracted from TripAdvisor by Wang, Lu, and Zhai (2010), which is available at `PrefLib.org`, an online reference library of preference data (Mattei and Walsh 2013). Users were able to rate each hotel by assigning between 1 and 5 stars for each of a number of features. We only use the part of the original dataset consisting of reviews where the user has rated all of the following six features: value-for-money, rooms, location, cleanliness, front desk, and service. We interpret any rating between 1 and 3 stars as a negative signal ("issue rejected") and any rating of 4 or 5 stars as a positive signal ("issue accepted").[6] A single review is thus transformed into a ballot with six binary issues and these are bundled into profiles, one for each hotel, resulting in 1,850 profiles (hotels) with an average of 68 ballots (reviews) each.

There is no (explicit) integrity constraint. Nevertheless, the judgments made in the reviews can be expected to be fairly correlated, not only because of the relative similarity between some of the features, but also because guests who had a pleasant stay in a hotel will often evaluate all features relatively highly. The mean polarisation coefficient of the 1,850 profiles is 0.26. In order to construct a second dataset of particularly highly polarised profiles, we have collected all profiles with a polarisation coefficient of at least 0.5. This second dataset consists of 31 profiles with 32 ballots on average. Its mean polarisation coefficient is 0.57.

3.3 Evaluation Criteria

We will use three different criteria to assess the quality of the results returned by different aggregation rules. The first two are relatively simply measures that track how well the outcome preserves certain features of the input profile.

Definition 6. *The **agreement score** of outcome $B^\star$ relative to profile B is defined as follows:*

$$\mathrm{agr}_B(B^\star) = \frac{1}{nm} \sum_{j \in \mathcal{I}} \#\{i \in \mathcal{N} \mid b_{ij} = b_j^\star\}$$

Thus, $\mathrm{agr}_B(B^\star)$ measures overall agreement of $B^\star$ with the profile B. In other words, it measures how well the outcome matches the majority outcome (which is the ideal as far as this specific evaluation criterion is concerned).

Definition 7. *The **correlation score** of outcome $B^\star$ relative to profile B is defined as follows:*

$$\mathrm{corr}_B(B^\star) = \frac{1}{2 \cdot \binom{m}{2}} \sum_{j \neq j' \in \mathcal{I}} |\mathrm{corr}_B(j,j') + \mathrm{cmp}(b_j^\star, b_{j'}^\star)|,$$

where $\mathrm{corr}_B(j,j') = \frac{1}{n} \cdot (2 \cdot \#\{i \in \mathcal{N} \mid b_{ij} = b_{ij'}\} - n)$, and $\mathrm{cmp}(x,y) = 1$ if $x = y$ and $\mathrm{cmp}(x,y) = -1$ otherwise.

Both of the terms inside the sum are positive for concordant arguments and negative otherwise. As the absolute value of their sum is maximal when they assume the same sign (with 2 as top value), the more $B^\star$ respects the correlation between issues in B, the higher will be the correlation score. For polarised profiles, the two scores are complementary, as corr_B would evaluate best a ballot in the middle of

---

[6]We chose a cut-off point between 3 and 4 rather than between 2 and 3 to achieve a better balance between 0's and 1's in the resulting dataset. Roughly 24% of all judgments in this dataset are 0's.
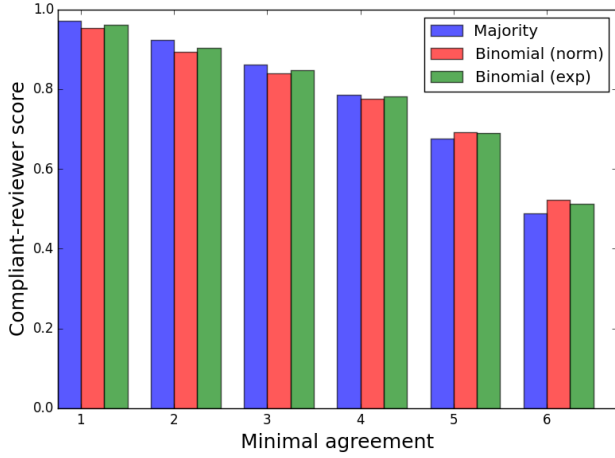
Figure 1: Compliant-reviewer scores of all profiles

Figure 2: Compliant-reviewer scores of polarised profiles

a cluster, while agr_B is usually maximised by ballots that compromise between clusters.

To introduce our third evaluation criterion, consider the following scenario. You want to write a hotel review for an online magazine and you want to please as many of the readers of the magazine as possible (to maximise the number of *like*'s received). Suppose a reader will like your review if she agrees with your judgment on at least k of the issues (for our data, $k \leqslant 6$). Our third evaluation criterion assesses how well the outcome $B^\star$ produced by the rule evaluated solves this *compliant-reviewer problem*.[7]

Definition 8. *The **compliant-reviewer score** with threshold $k \leqslant m$ of outcome $B^\star$ relative to profile $\boldsymbol{B}$ is defined as:*

$$\text{c-r}_B^k(B^\star) \;\; = \;\; \frac{1}{n} \cdot \#\{i \in \mathcal{N} \mid \text{Agr}(B_i, B^\star) \geqslant k\}$$

The assumption underlying this evaluation criterion is that when there are dependencies between issues, agents will want a higher agreement with the outcome of a rule, since even small changes would break the patterns of acceptance satisfying the dependencies, thereby leading to a bad decision as in the example in the introduction.

3.4   Results

We have compared three rules on the two datasets: the majority rule (which is equivalent to $F_{\{1\}}$, as there is no integrity constraint), the normalised binomial rule $F_{[m],w_1}$, and the exponentially normalised binomial rule $F_{[m],w_2}$. For the majority rule, ties have been broken in favour of 0.[8] For the other two rules, we did not observe any ties.

For the full dataset, Figure 1 plots the mean compliant-reviewer score for thresholds $k = 1, \ldots, 6$ for each rule. We can see that the majority rule does slightly better for small k and the binomial rules do slightly better for large k, which is an effect one would have expected to observe. Still, the

main finding here is that all rules perform similarly well. To a large extent, this effect is due to all rules choosing the same outcome: majority and $F_{[m],w_1}$ agree in 75% of the cases, majority and $F_{[m],w_2}$ in 84%, and the two binomial rules in 87%. Also the mean agreement and correlation scores are not affected significantly by the choice of rule:

Majority:	$\text{agr}_B(B^\star) = 1.00$	$\text{corr}_B(B^\star) = 0.75$
Binomial (norm):	$\text{agr}_B(B^\star) = 0.99$	$\text{corr}_B(B^\star) = 0.79$
Binomial (exp):	$\text{agr}_B(B^\star) = 1.00$	$\text{corr}_B(B^\star) = 0.78$

The situation changes for the second dataset, which includes only highly polarised profiles. Figure 2 shows the mean compliant-reviewer scores for $k = 1, \ldots, 6$. As we can see, when the readers' agreement with the reviewer does not need to be too high, e.g., at most 3, then the majority rule performs best in getting as many *like*'s as possible. But when readers want a high agreement with the reviewer to be satisfied, then the normalised binomial rule works much better. This is because it chooses ballots in the middle of a cluster, thereby achieving high agreement with the individuals in that cluster, but disregarding all the ballots belonging to the other cluster. In typical examples in the dataset, the majority rule accepts half of the issues and rejects the other half, thereby arriving at a compromise that is too weak for most readers, while the normalised binomial rule tends to pick either a clearly positive outcome (with all issues accepted) or a clearly negative outcome (with all issues rejected). The exponentially normalised binomial rule scores even better.

For the restricted dataset, the agreement between the majority rule and the two binomial rules drops significantly, from 75% to 13% and from 84% to 16%, respectively. The agreement between the two binomial rules stays high, at 84% (down from 87%). The mean agreement and correlation scores for the restricted dataset are as follows:

Majority:	$\text{agr}_B(B^\star) = 1.00$	$\text{corr}_B(B^\star) = 0.58$
Binomial (norm):	$\text{agr}_B(B^\star) = 0.95$	$\text{corr}_B(B^\star) = 0.83$
Binomial (exp):	$\text{agr}_B(B^\star) = 0.97$	$\text{corr}_B(B^\star) = 0.81$

As we can see, moving outcomes towards the centre of a cluster, as is the case for the two binomial rules, increases

---

[7]As pointed out by an anonymous reviewer, there are interesting connections between the *Ostrogorski Paradox* (Kelly 1989) and (failures to solve) our compliant-reviewer problem.

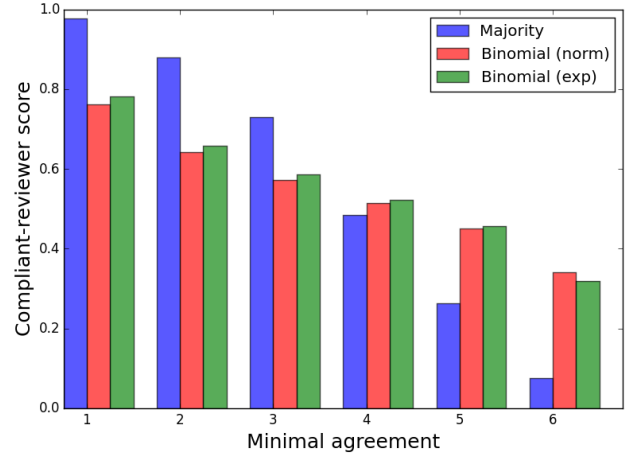[8]Breaking ties in favour of 1 does not significantly affect results.

the correlation score, but that improvement is paid for with a (very modest) decrease in the agreement score.

4 Conclusion

We have introduced the binomial rules for judgment aggregation to account for hidden dependencies in the input and placed them into the larger landscape of rules proposed in the literature. They stand out as satisfying the reinforcement axiom as well as being collectively rational, and they include both intractable and computationally easy rules. Our experiments, performed on real-world data, show that, indeed, binomial rules capture dependencies better than the majority rule, with only a very small loss in total agreement. The exponentially normalised rule performs particularly well.

To be able to carry out our experiments, we have developed both a novel notion of polarisation for profiles of judgments and several evaluation criteria for judgment aggregation rules. Both of these contributions may be of independent interest to others.

Our work suggests multiple avenues for future work. First, a better understanding of how to choose the weight function w in $F_{K,w}$ is required (so far, our experiments merely suggest that fast-decreasing weight functions are useful to compensate for the fast-growing binomial coefficients). Second, our evaluation criteria may be applied to other rules and other data. Third, these criteria themselves suggest new approaches to designing aggregation rules.

References

Arrow, K. J.; Sen, A. K.; and Suzumura, K., eds. 2002. *Handbook of Social Choice and Welfare*. North-Holland.

Can, B.; Ozkes, A. I.; and Storcken, T. 2014. Measuring polarization in preferences. Technical Report RM/14/018, Maastricht University.

Caragiannis, I.; Procaccia, A. D.; and Shah, N. 2014. Modal ranking: A uniquely robust voting rule. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI-2014)*.

Dietrich, F., and List, C. 2007a. Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics* 19(4):391–424.

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

Dietrich, F., and Mongin, P. 2010. The premiss-based approach to judgment aggregation. *Journal of Economic Theory* 145(2):562–582.

Dietrich, F. 2014. Scoring rules for judgment aggregation. *Social Choice and Welfare* 42(4):873–911.

Endriss, U., and Grandi, U. 2014. Binary aggregation by selection of the most representative voter. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI-2014)*.

Endriss, U.; Grandi, U.; and Porello, D. 2012. Complexity of judgment aggregation. *Journal of Artificial Intelligence Research (JAIR)* 45:481–514.

Endriss, U. 2016. Judgment aggregation. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds., *Handbook of Computational Social Choice*. Cambridge University Press.

Everaere, P.; Konieczny, S.; and Marquis, P. 2015. Belief merging versus judgment aggregation. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2015)*. IFAAMAS.

Grandi, U., and Endriss, U. 2011. Binary aggregation with integrity constraints. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-2011)*.

Grandi, U. 2012. *Binary Aggregation with Integrity Constraints*. Ph.D. Dissertation, ILLC, University of Amsterdam.

Grossi, D., and Pigozzi, G. 2014. *Judgment Aggregation: A Primer*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers.

Hartmann, S., and Sprenger, J. 2012. Judgment aggregation and the problem of tracking the truth. *Synthese* 187(1):209–221.

Hemaspaandra, E.; Spakowski, H.; and Vogel, J. 2005. The complexity of Kemeny elections. *Theoretical Computer Science* 349(3):382–391.

Kelly, J. S. 1989. The Ostrogorski paradox. *Social Choice and Welfare* 6(1):71–76.

Kemeny, J. 1959. Mathematics without numbers. *Daedalus* 88(4):577–591.

Kornhauser, L. A., and Sager, L. G. 1993. The one and the many: Adjudication in collegial courts. *California Law Review* 81(1):1–59.

Lang, J., and Slavkovik, M. 2013. Judgment aggregation rules and voting rules. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT-2013)*. Springer-Verlag.

Lang, J.; Pigozzi, G.; Slavkovik, M.; and van der Torre, L. 2011. Judgment aggregation rules based on minimization. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2011)*.

List, C., and Pettit, P. 2002. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy* 18(1):89–110.

List, C., and Puppe, C. 2009. Judgment aggregation: A survey. In Anand, P.; Pattanaik, P.; and Puppe, C., eds., *Handbook of Rational and Social Choice*. Oxford University Press.

Mattei, N., and Walsh, T. 2013. PrefLib: A library of preference data. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT-2013)*. Springer-Verlag. http://www.preflib.org.

Miller, M. K., and Osherson, D. 2009. Methods for distance-based judgment aggregation. *Social Choice and Welfare* 32(4):575–601.

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

Slavkovik, M. 2014. Not all judgment aggregation should be neutral. In *Proceedings of the European Conference on Social Intelligence (ECSI-2014)*. CEUR.

Wagner, K. W. 1990. Bounded query classes. *SIAM Journal on Computing* 19(5):833–846.

Wang, H.; Lu, Y.; and Zhai, C. 2010. Latent aspect rating analysis on review text data: A rating regression approach. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2010)*. ACM.

Young, H. P., and Levenglick, A. 1978. A consistent extension of Condorcet's election principle. *SIAM Journal on Applied Mathematics* 35(2):285–300.

Young, H. P. 1974. An axiomatization of Borda's rule. *Journal of Economic Theory* 9(1):43–52.