Mathematical Social Sciences 76 (2015) 19–30

Contents lists available at ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/econbase

Complexity of manipulation and bribery in judgment aggregation for uniform premise-based quota rules☆

Dorothea Baumeister [a,*], Gábor Erdélyi [b], Olivia J. Erdélyi [c], Jörg Rothe [a]

[a] Institut für Informatik, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany
[b] School of Economic Disciplines, University of Siegen, 57076 Siegen, Germany
[c] Institut für Versicherungsrecht, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

HIGHLIGHTS

- We extend previous work on the complexity of manipulation in judgment aggregation.
- We consider incomplete judgment sets and various notions of preferences on them.
- We introduce bribery in judgment aggregation and study its complexity.

ARTICLE INFO

Article history:
Received 28 January 2014
Received in revised form
17 March 2015
Accepted 23 March 2015
Available online 3 April 2015

ABSTRACT

Endriss et al. (2012) initiated the complexity-theoretic study of problems related to judgment aggregation. We extend their results on the manipulation of two specific judgment aggregation procedures to a whole class of such procedures, namely to uniform premise-based quota rules. In addition, we consider incomplete judgment sets and the notions of top-respecting and closeness-respecting preferences introduced by Dietrich and List (2007). This complements previous work on the complexity of manipulation in judgment aggregation that focused on Hamming-distance-respecting preferences only, which we also study here. Furthermore, inspired by work on bribery in voting (Faliszewski and Rothe, in press), we introduce and study the closely related issue of bribery in judgment aggregation.

1. Introduction

Judgment Aggregation is the task of aggregating individual judgment sets of possibly interconnected logical propositions (see the surveys by List and Puppe (2009) and List (2012), and the book chapter by Endriss (in press)) and can therefore be seen as an important framework for collective decision making. Decision-making processes are often susceptible to various types of interference, be it internal or external. In social choice theory and in computational social choice, ways of influencing the outcome of elections – such as manipulation and bribery – have been studied intensely, with a particular focus on the complexity of the related problems. In particular, (coalitional) *manipulation* refers to (a group of) strategic voters casting their votes insincerely to reach their desired outcome. In *bribery* (see, e.g., Faliszewski et al., 2009a,b and the book chapter by Faliszewski and Rothe, in press) an external agent seeks to reach her desired outcome by bribing – within a given budget – some voters to alter their votes. Strategic behavior has been studied to a far lesser extent in judgment aggregation than in voting so far.

Mechanisms for collective decision making that are susceptible to strategic behavior, be it from the agents involved as in manipulation or from external authorities or actors as in bribery, are obviously not desirable, as that undermines the trust we have in them. We therefore have a strong interest in accurately assessing how vulnerable a mechanism for collective decision making is to these internal or external influences. Unfortunately, in many concrete settings of social choice, "perfect" mechanisms do not exist. For example, the Gibbard–Satterthwaite theorem says that no

☆ This paper combines and extends the results from preliminary conference versions that appear in the proceedings of the *2nd* and the *3rd International Conference on Algorithmic Decision Theory* (ADT'11 and ADT'13) (Baumeister et al., 2011, 2013) and of the *4th* and the *5th International Workshop on Computational Social Choice* (COMSOC'12 and COMSOC'14) (Baumeister et al., 2012, 2014), and was presented at the *ESSLLI Workshop on Logical Models of Group Decision Making* (ESSLLI-LMGD'13).

* Corresponding author. Tel.: +49 211 8111634.
E-mail address: baumeister@cs.uni-duesseldorf.de (D. Baumeister).
URLs: http://ccc.cs.uni-duesseldorf.de/~baumeister (D. Baumeister), http://www.wiwi.uni-siegen.de/dt/team/erdelyi/ (G. Erdélyi), http://ccc.cs.uni-duesseldorf.de/~rothe (J. Rothe).

http://dx.doi.org/10.1016/j.mathsocsci.2015.03.006
0165-4896/

reasonable voting system can be "strategy-proof" (Gibbard, 1973; Satterthwaite, 1975) (see also the generalization by Duggan and Schwartz, 2000), and Dietrich and List (2007a) give an analogue of the Gibbard–Satterthwaite theorem in judgment aggregation.

To avoid this obstacle, a common approach in computational social choice is to apply methods from theoretical computer science to show that undesirable strategic behavior is blocked, or at least hindered, by the corresponding task being a computationally intractable problem. While this approach evidently makes sense in voting, one may wonder whether it is also applicable to judgment aggregation scenarios. After all, these scenarios originate from and are motivated by juridical issues in court proceedings (Kornhauser and Sager, 1986) where usually only a few judges collectively make decisions. However, there are also large-scale examples of real-world judgment aggregation scenarios. Suppose that a community is going to make a decision on whether or not to build a new prison and that all citizens are invited to participate in this decision making via an online platform.[1] After discussing the pros and cons, the question boils down to their judgment of the following two propositions: (A) The community's income from municipal taxes is large enough to afford building a prison, and (B) the crime rate in the community has raised so much that this indeed is needed. The prison will be built if both (A) and (B) are affirmed by the citizens. This example is very similar to other examples motivating judgment aggregation (see, e.g., Example 1 in Section 2 on a decision by three judges in a soccer match), but its point is that it can be realistic to have a large number of judges, and it therefore makes sense to apply computational complexity also to judgment aggregation problems. Besides these examples judgment aggregation is also used in artificial intelligence, notably as a tool for collective decision making in systems with interacting autonomous agents, which also may involve a very large number of participating judges. For further applications of judgment aggregation in computer science, see the book chapter by Endriss (in press).

1.1. Our contributions

As mentioned above, much work on the complexity of manipulation and bribery problems has been done in voting, but only a few results are known for these problems in judgment aggregation. Most notably, Endriss et al. (2012) recently initiated the algorithmic and complexity-theoretic study of the winner determination problem and the manipulation problem in judgment aggregation, and we here extend their work for manipulation to the class of uniform premise-based quota rules and to other notions of preference that have been introduced by Dietrich and List (2007a). In particular, we will study incomplete judgment sets and top-respecting and closeness-respecting preferences in addition to Hamming-distance-respecting preferences. We also introduce exact variants of manipulation where the manipulator's goal is to achieve not only a better, but a *best* outcome for a given subset of her desired set. This gives rise to a number of manipulation problems for each judgment aggregation rule, which is why we here focus on only one class of such rules, the uniform premise-based quota rules. Extending this work to other judgment aggregation rules (such as those mentioned in Section 1.2 below), to allow a comparison of these rules in terms of their resistance to manipulation, is left for future work.

A main result of this paper is presented in Theorem 10, which says that for each rational quota and for any fixed number of at least three judges, the uniform premise-based quota rule is hard to manipulate for Hamming-distance-respecting preferences in terms of the parameterized complexity class W[2] (see Section 2.2) when parameterized by the maximum number of changes in the premises needed in the manipulator's desired set. We also provide many complexity results for manipulation with respect to top-respecting and closeness-respecting preferences (see Table 6 in Section 5 for an overview).

In addition, we here initiate the algorithmic and complexity-theoretic study of bribery problems in judgment aggregation. Again, these problems are each closely related to the corresponding bribery problems in voting, yet are specifically tailored to judgment aggregation scenarios. Table 7 in Section 5 gives an overview of our results on the complexity of bribery problems for judgment aggregation with the premise-based procedure.

This paper combines and extends previous work by Baumeister et al. (2011, 2012, 2013, 2014b) that appeared in the proceedings of ADT'11, COMSOC'12, ADT'13, and COMSOC'14. The present version contains some additional results and it provides more discussion and a number of notational improvements.

1.2. Related work

Manipulation and bribery are two forms of strategic actions that have been studied extensively in voting (see the references below), yet much less so in judgment aggregation. Endriss et al. (2012) were the first to study manipulation in judgment aggregation from a computational point of view, and we here extend their work as described in Section 1.1. In voting theory, another way of tampering with elections is control, and Baumeister et al. (2013, 2012) have studied certain types of control in judgment aggregation where an external agent seeks to influence the outcome by altering the structure of the judgment aggregation process by adding, deleting, or replacing judges. Dietrich (2014) studied the agenda manipulation problem, where one tries to influence the outcome by carefully choosing the formulas in the agenda. In the case of sequential judgment aggregation procedures, the order of the formulas in the agenda is important and may give the opportunity to rule manipulation, see the work of List (2004) and Dietrich and List (2007b).

For our complexity-theoretic analysis of manipulation and bribery in judgment aggregation, we will focus on the *uniform premise-based quota rules* (see Section 2.1 for a formal definition). In these rules, the agenda is divided into *premises* and *conclusions*, the outcome for each of the premises is determined by a given quota (just one – a *uniform* quota assigned to each of the premises), and the outcome for each conclusion is then derived from the outcome for the premises in a consistent way. That is, under a uniform premise-based quota rule we collectively accept those conclusions that logically follow from the premises we collectively accept according to the given quota. By contrast, in *conclusion-based procedures* (see, e.g., Kornhauser and Sager, 1986; List and Pettit, 2002; Dietrich, 2006), the collective decision is made for the conclusions of the agenda only. Another approach are the *distance-based procedures* (see the work of Miller and Osherson, 2009) where a collective outcome minimizes the distance (according to a certain predetermined metric) to the given individual judgment sets. Lang et al. (2011) define and study judgment aggregation procedures based on minimization, which are inspired by voting

[1] Implementing real-world, large-scale judgment aggregation scenarios such as this one on an online platform is not fictitious. For example, the interdisciplinary graduate college "Online Partizipation" (www.fortschrittskolleg.de) of HHU Düsseldorf and other institutions (including as practice partners registered societies such as Liquid Democracy e.V., limited liability companies such as Zebralog GmbH & Co KG and polidia GmbH as well as the municipal councils of Bonn, Köln, and Münster, among others) investigates such settings to explore the technical, societal, economical, and juridical aspects of a self-organizing society. A central goal in this project, which is funded by the NRW Ministry for Innovation, Science, and Research, is to build an internet platform (www.normsetting.org) that can be used for online discussions and deliberations, and hundreds of participants have been involved in previous pilot projects.