

Strategic voting when aggregating partially ordered preferences

M. S. Pini, F. Rossi, K. B. Venable
Department of Pure and Applied Mathematics
University of Padova, Italy
{mpini,frossi,kvenable}@math.unipd.it

T. Walsh
NICTA and School of CSE, UNSW
Sydney, Australia.
tw@cse.unsw.edu.au

ABSTRACT

Preferences of a single agent are often partially ordered. For example, it may be hard to compare a novel with a biography. In such a situation, the agent may want the novel and the biography to be considered incomparable. We consider here how to aggregate the partially ordered preferences of multiple agents in order to return a set of most preferred outcomes. We define the notion of strategy-proofness for such a scenario. This is when preference aggregation cannot be manipulated. We prove that if there is no dictator, agents can manipulate the result by voting strategically to determine the most preferred outcomes. This extends the well-known theorem by Gibbard and Satterthwaite for total orders.

Categories and Subject Descriptors

I.2.11 [Computing Methodologies]: Artificial Intelligence—*Distributed Artificial Intelligence*

General Terms

Theory

Keywords

Preference aggregation, strategy-proofness, partial orders, Gibbard Satterthwaite theorem

1. INTRODUCTION

Many problems require us to combine the preferences of different agents. For example, when planning a wedding, we must combine the preferences of the bride, the groom and possibly some or all of the in-laws. Typically such preferences are partial orders over the outcomes. We can view many different mechanisms for preference aggregation in terms of voting. Each agent is voting for their preferred outcomes. When aggregating preference orders, one may be

interested in obtaining a combined ordering among the outcomes, or just the set of most preferred outcomes (also called the winners). The first choice, in which the aggregation function is called a *social welfare* function, provides more information about the combined preference ordering since it tells us the ordering between any pair of outcomes. The second scenario, in which the aggregation function, called a *social choice* function, returns a set, is less informative but often enough when we are just interested in choosing one of the most preferred outcomes.

In this paper we will consider social choice functions. Thus the result of aggregating preferences will be just a set of outcomes. A very desirable property of preference aggregation is non-manipulability (also called strategy proofness). It should not be possible for agents to manipulate the election by voting strategically. Strategic voting is when agents express preferences which are different from their real ones, to get the result they want. If this is possible, then the preference aggregation rule is said to be manipulable. For social choice rules on *totally* ordered preferences, the Gibbard Satterthwaite theorem [3] proves that it is not possible to be at same time non-manipulable and have no dictators. Either there is a dictator (that is, an agent who gets what he wants by voting sincerely) or a manipulator (that is, an agent who gets what he wants by lying). In either case, there is an agent who gets what he wants no matter what the other agents say. In this paper we extend this result to *partially* ordered preferences. Even in this more general case, we prove that it is impossible for a social choice function to have no dictator and be non-manipulable at the same time. As with total orders, we conjecture that there will be ways around this negative result. For example, it may be that certain social choice functions on partial orders are computationally hard to manipulate. As another example, it may be that certain restrictions on the way agents vote (like single-peaked preferences for total orders) guarantee strategy-proofness.

2. SOCIAL CHOICE ON PARTIAL ORDERS

In some situations, the result of aggregating the preferences of a number of agents might not need to be an order over outcomes. It might be enough to know the “most preferred” outcomes. Social choice functions identify such most preferred outcomes.

A **profile** p is a sequence of n orderings p_1, \dots, p_n over outcomes, one for each agent $i \in \{1, \dots, n\}$. A social choice function on total orders is a mapping from a profile to the optimal outcome, or winner. With partial orders, (i.e. binary relation which are reflexive, transitive, antisymmetric

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.
Copyright 2006 ACM 1-59593-303-4/06/0005 ...\$5.00.

and not always complete), there can be several outcomes which are incomparable and optimal. We can therefore consider a generalization in which a social choice function is a mapping from a profile to a non-empty set of outcomes, called the optimal outcomes, or the winners.

Let us now consider some properties of such social choice functions. Our definition of these properties is a generalization of the corresponding definition for social choice function on totally ordered profiles [1]. We say that a social choice function f is **unanimous** iff

- for any profile p with $a \in \text{top}(p_i)$ for every agent i , then $a \in f(p)$;
- for any profile p with $\{a\} = \text{top}(p_i)$ for every agent i , then $f(p) = \{a\}$.

It is instead **monotonic** iff, given two profiles p and p' ,

- if $a \in f(p)$ and, for any other alternative b , $a >_{p_i} b$ implies $a >_{p'_i} b$, $a \bowtie_{p_i} b$ implies $a \bowtie_{p'_i} b$ or $a >_{p'_i} b$, for all agents i , then $a \in f(p')$;
- if $f(p) = A$ and for all $a \in A$, for all b , $a >_{p_i} b$ implies $a >_{p'_i} b$ and $a \bowtie_{p_i} b$ implies $a \bowtie_{p'_i} b$ or $a >_{p'_i} b$, for all agents i , then $f(p') = A$.

We define three notions of dictators for social choice functions: a **strong dictator** is an agent i such that, for all profiles p , $f(p) = \text{top}(p_i)$; a **dictator** is an agent i such that, for all profiles p , $f(p) \subseteq \text{top}(p_i)$; a **weak dictator** is an agent i such that, for all profiles p , $f(p) \cap \text{top}(p_i) \neq \emptyset$. Notice that, in any profile p , if a is the unique top of a weak dictator i , then $a \in f(p)$. However, this is not true if a is not the unique top of i .

We say a social choice function is **strongly fair**, **fair** or **weakly fair** if it is unanimous, monotonic, and does not have a strong dictator, dictator or weak dictator, respectively. Since a strong dictator is a dictator, and a dictator is a weak dictator, weak fairness implies fairness which itself implies strong fairness. In [6], we generalize the Muller-Satterthwaite theorem [5] to social choice functions over partial orders without ties, for weak dictators. That is, weak dictators are inevitable if we have two or more agents, three or more outcomes and the social choices function is unanimous and monotonic. We rewrite here this theorem formally since it will be used in the rest of the paper.

Theorem 1 ([6]) *If we have at least two agents and at least three outcomes, and a social choice function on partial order without ties is unanimous and monotonic, then there is at least one weak dictator.*

3. STRATEGY PROOFNESS

The Gibbard Satterthwaite theorem on totally ordered preferences [3, 7] proves that either we have a dictator or the social choice function can be manipulated. That is, agents can manipulate the result using tactical voting. We may now wonder if a similar relationship holds between weak dictators and non-manipulability. To answer this question, we must generalize the notion of non-manipulability (also called strategy proofness) to social choice functions on partially ordered preferences, and then consider the Gibbard Satterthwaite theorem [3, 7] in this more general context.

We will show that weak dictators are inevitable if we have at least two agents and three outcomes, and the social choice function is strategy proof and onto. These conditions are identical to those in the Gibbard Satterthwaite theorem for total orders. A social choice function is strategy-proof if it is best for each agent to order outcomes as she really prefers and not to try to order them tactically, with the hope of getting a better result. More precisely, the social choice function must never allow an agent to get a preferred outcome among the winners by ordering outcomes in a way that contradicts his or her true preferences [1].

Formally, a social choice function is **strategy proof** if, for every agent i , for every pair of profiles p and p' , which differ only for the agent i 's ranking, that is, p_i , we have that

- $\forall a \in f(p) - f(p'), \forall b \in f(p')$, if $a \bowtie_{p_i} b$, then $a \bowtie_{p'_i} b$ or $a <_{p'_i} b$; if $a <_{p_i} b$, then $a <_{p'_i} b$;
- $\forall a \in f(p) - f(p'), \exists b \in f(p')$ such that, if $a >_{p_i} b$, then $a \bowtie_{p'_i} b$ or $a <_{p'_i} b$; if $a \bowtie_{p_i} b$, then $a <_{p'_i} b$.

In other words, a social choice function is strategy proof if an agent can remove an element a from the set of winners only by worsening its rank with respect to at least one the new winners b , and not improving it with respect to any other new winner. This means that it is not possible for an agent to make a disappear from the set of winners by improving its ranking in her preference ordering. In fact, this would be tactical voting.

In general, even in the totally ordered case, most voting procedures involving three or more alternatives are not strategy proof [1]. This is true also in the partially ordered case. Let us consider the following two social choice functions: f_1 is such that $f_1(p) = \bigcup_i \text{top}(p_i)$, that is, this function returns the union of the sets of optimal elements of each agent and f_2 is the Pareto function, which returns the optimal elements of the ordering returned by the Pareto social welfare function (where $a > b$ if all agents say $a > b$, otherwise $a \bowtie b$).

Let us now consider two profiles p and p' on three alternatives a, b and c such that $p = (p_1 = (c > a \wedge c \bowtie b \wedge a \bowtie b), p_2 = (c \bowtie a \wedge c \bowtie b \wedge a \bowtie b), p_3 = p_1)$ and $p' = (p'_1 = p_1, p'_2 = (c > a > b), p'_3 = p_3)$. Then, for both such social choice functions, the set of winners in profile p is $\{a, b, c\}$, while in profile p' is $\{b, c\}$. Thus a has disappeared by passing from p to p' but its ranking has improved with respect to b in agent 2. Thus both social choice functions are not strategy-proof.

In order to prove the generalization of the Gibbard Satterthwaite theorem to the partial order case we first prove the following result, which relates strategy proofness to unanimity and monotonicity.

Theorem 2 *If a social choice function is strategy proof and onto, then it is unanimous and monotonic.*

PROOF. The proof is composed by two parts. Part 1 shows that if a social choice function f is strategy proof then it is monotonic, while part 2 shows that if f is onto and monotonic then it is unanimous.

Part 1. Consider two profiles p and p' , which differ only for the ranking of agent i .

Assume that $a \in f(p)$ and that for any other alternative b , $a >_{p_i} b$ implies $a >_{p'_i} b$ and $a \bowtie_{p_i} b$ implies $a \bowtie_{p'_i} b$ or $a >_{p'_i} b$. We want to show that $a \in f(p')$. For the sake of

contradiction, assume that $a \notin f(p')$. Since f is strategy proof, then $\exists c \in f(p')$ such that one of the following holds: (i) if $a >_{p_i} c$, then $a \bowtie_{p'_i} c$ or $a <_{p'_i} c$, (ii) if $a \bowtie_{p_i} c$, then $a <_{p'_i} c$. If the first holds then there is an element c which is worse than a in p_i and that becomes strictly better than or incomparable to a in p'_i . This contradicts the fact that for any other alternative b , $a >_{p_i} b$ implies $a >_{p'_i} b$. If the second holds then there is an element c which is incomparable with a in p_i and becomes strictly better than a in p'_i . This contradicts the fact that for any other alternative b , $a \bowtie_{p_i} b$ implies $a \bowtie_{p'_i} b$ or $a >_{p'_i} b$.

Assume now that $\forall a \in f(p)$, for any other alternative b , $a >_{p_i} b$ implies $a >_{p'_i} b$ and $a \bowtie_{p_i} b$ implies $a \bowtie_{p_i} b$ or $a >_{p'_i} b$. We want to show that $f(p) = f(p')$. For the sake of contradiction, we can assume that $\exists a$ such that $a \in f(p)$ and $a \notin f(p')$ or that $\exists a$ such that $a \in f(p')$ and $a \notin f(p)$. If $\exists a$ such that $a \in f(p)$ and $a \notin f(p')$ then, since f is strategy proof, the same reasoning above leads to the same contradictions. If instead $\exists a$ such that $a \in f(p')$ and $a \notin f(p)$, then since f is strategy proof, then $\exists c \in f(p)$ such that one of the following holds: (i) if $a >_{p'_i} c$, then $a \bowtie_{p_i} c$ or $a <_{p_i} c$, (ii) if $a \bowtie_{p'_i} c$, then $a <_{p_i} c$. In the first case there is an element a which is strictly smaller than or incomparable to c in p_i that becomes strictly greater than c in p'_i . This is in contradiction either with the fact that for any other alternative b , $c \bowtie_{p_i} b$ implies $c \bowtie_{p'_i} b$ or $c >_{p'_i} b$ or with the fact that $c >_{p_i} b$ implies $c >_{p'_i} b$. If the second case holds then there is an element a that is smaller than c in p_i and that becomes incomparable with c in p'_i . This contradicts the assumption that for any other alternative b , $c >_{p_i} b$ implies $c >_{p'_i} b$.

Consider two profiles q such that $f(q) = \{a\}$ and a profile q' such that for every agents i and for every alternative b , $a >_{q_i} b$ implies $a >_{q'_i} b$. We want prove that $f(q') = \{a\}$, that is the first part of the definition of the monotonicity for social choice functions. Since we can move from $q = (q_1, \dots, q_n)$ to $q' = (q'_1, \dots, q'_n)$, passing from $q = (q_1, q_2, \dots, q_n)$ to (q'_1, q_2, \dots, q_n) , and (q'_1, q'_2, \dots, q_n) to (q'_1, q'_2, \dots, q_n) and so on, and we have shown above that at each step a remains in the set of winners, $a \in f(q')$. The same reasoning holds for profiles q such that $f(q) = A$ and q' such that for every agent i , for $\forall a \in A$, for every other alternative b , such that $a >_{q_i} b$ implies $a >_{q'_i} b$. In this case we conclude that $f(q') = A$. This is the second part of the definition of monotonicity for social choice function. We have thus shown that f is monotonic.

Part 2. Since f is onto, then for every subset S of alternatives there is a profile p such that $f(p) = S$. If $S = \{a\}$, where a is an alternative, since f is onto, there is a profile p such that $f(p) = \{a\}$. If we consider the profile p' , obtained from profile p bringing a to the very top of every agent, then for strict monotonicity on profiles p and p' , that we have just proved, $f(p') = a$. Therefore, whenever a is the unique top of every agent's ranking in a profile \bar{p} , then $f(\bar{p}) = \{a\}$. Because a is arbitrary then f satisfies Pareto efficiency in the case of unique top for every agent.

If $S \supset \{a\}$, since f is onto, then there is a profile p_1 such that $f(p_1) = S \supset \{a\}$. If we consider a profile p'_1 where a is at the top (not unique) for every agent, then, for monotonicity on profiles p_1 and p'_1 , $f(p'_1)$ must contain a . Therefore, whenever a is one of the tops of every agent's ranking in a profile \bar{p} , then $f(\bar{p}) = \{a\}$. Because a is arbitrary, then f satisfies Pareto efficiency in the case of not unique top for

every agent. Q.E.D.

We now use this result, together with the extension of the Muller-Satterthwaite result (Theorem 1), to prove the main theorem stating that it is not possible for a social choice function to be at the same time strategy proof and onto, and have no weak dictators.

Theorem 3 *If there are at least two agents and at least three outcomes and the social choice function is strategy-proof and onto then there is at least one weak dictator.*

PROOF. By Theorem 2, if f is onto and strategy proof then it is monotonic and unanimous. Moreover, Theorem 1 states that if f is monotonic and unanimous then there is at least one weak dictator. Q.E.D.

4. RELATED WORK

Efforts have been made to weaken the conditions of the Gibbard Satterthwaite theorem. For example, in [2], the domain of the social choice function has been generalized to preferences over sets of outcomes and an impossibility result proved. The Gibbard Satterthwaite theorem has been shown to be robust to several other restrictions of the domain of the social choice function. Here we have shown that this theorem is robust also to including incomparability in the preference ordering.

5. REFERENCES

- [1] K. J. Arrow and A. K. Sen and K. Suzumara. *Handbook of Social Choice and Welfare*. North-Holland, Elsevier, 2002.
- [2] S. Barbera. Strategy-Proofness and Pivotal Voters: A Direct Proof of the Gibbard-Satterthwaite Theorem. *International Economic Review*, Vol.24, No.2, pages 413–417, 1983.
- [3] A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41, 1973.
- [4] K. Konczak and J. Lang. Voting procedures with incomplete preferences. IJCAI 05 Workshop on Advances in Preference Handling, 2005.
- [5] E. Muller and M. A. Satterthwaite. The equivalence of strong positive association and strategy-proofness. *Economic Theory*, 14, 1977.
- [6] M. S. Pini, F. Rossi, K. B. Venable, T. Walsh. Aggregating partially ordered preferences: impossibility and possibility results. *Proc. TARK X*, ACM Digital Library, June 2005.
- [7] M. A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.