Non-manipulability in Set-valued and Probabilistic Social Choice Theory

Doctoral Consortium

Patrick Lederer
Technische Universität München
München, Germany
ledererp@in.tum.de

ABSTRACT

A fundamental requirement in social choice theory is non-manipulability, i.e., voters should not be able to benefit by voting dishonestly. Unfortunately, a seminal result by Gibbard [12] and Satterthwaite [15] states that only extremely unattractive voting rules can be strategyproof if it is required to choose a single winner deterministically. Two common approaches for circumventing this impossibility are to allow for sets of winners and to allow for randomization. It is for both approaches possible to define various strategyproofness notions based on different assumptions on how voters compare sets of alternatives or lotteries on alternatives, and consequently, both positive and negative results can be obtained. The goal of this PhD project is to analyze for both models the boundary between possibility and impossibility results for various strategyproofness notions.

KEYWORDS

Non-manipulability; Strategyproofness; Social Choice Theory; Probabilistic Social Choice; Social Choice Correspondences; Social Choice Functions; Social Decision Schemes

ACM Reference Format:
Patrick Lederer. 2021. Non-manipulability in Set-valued and Probabilistic Social Choice Theory: Doctoral Consortium. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

1 INTRODUCTION

When multiple agents try to agree on a common decision, they need to aggregate their preferences. However, a large number of mechanisms are used in practice and it is difficult to name the best one. This question is central in social choice theory, which analyzes aggregation mechanisms and tries to find arguments for using particular methods by investigating desirable axioms. A particularly important axiom is non-manipulability, also called strategyproofness, which describes that agents should not be able to improve the outcome of an election from their individual perspective by voting dishonestly. Unfortunately, Gibbard [12] and Satterthwaite [15] have shown that every strategyproof aggregation mechanism that deterministically chooses a single winner is dictatorial if it has at least three different outcomes. As a consequence of this result, only extremely undesirable voting rules can be strategyproof.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

The goal of my PhD project is to investigate novel approaches to circumvent the Gibbard-Satterthwaite theorem by relaxing single-valuedness. It is a common sentiment among social choice theorists that this condition is unnatural and too restrictive [see, e.g., 1, 3, 9, 11, 14, 16]. The reason for this sentiment is that single-valuedness is in variance with basic fairness conditions: consider, for instance, an election with two alternatives, each of which is favored by exactly half of the voters. Then, both alternatives are equally acceptable, but single-valuedness forces us to decide on one of them. Two possible solutions to this problem are to either allow for sets of winners, which leads to (set-valued) social choice functions (SCFs), or to allow for randomization, which leads to social decision schemes (SDSs). In the sequel, I explain results and ideas for future work for both approaches.

2 SOCIAL CHOICE FUNCTIONS

The first escape route to the Gibbard-Satterthwaite theorem is to allow for a set of winners from which the final winner is extracted by a tie-breaking mechanism. For instance, this tie-breaking mechanism can be a lottery that determines the final winner by chance, or a chairperson who picks the final winner according to its preferences. In particular, we assume that voters cannot compute the outcome of the tie-breaking mechanism if they know the set of possible winners. Therefore, we are interested in the first step in which a set of alternatives is chosen based on the voters' preferences. This idea is formalized by social choice functions (SCFs) which map every preference profile to a non-empty set of alternatives.

Unfortunately, it is not obvious how to define strategyproofness for SCFs as it is unclear how voters compare sets of alternatives. For instance, if a voter prefers a to b to c, it is not clear whether he prefers the set $\{a, c\}$ to the set $\{b\}$. As a consequence, various strategyproofness notions based on different assumptions on how voters compare sets of alternatives have been investigated [see, e.g., 4, 9–11, 14, 16]. The results of these authors are of mixed nature for strict preferences: weak strategyproofness notions, such as the ones by Kelly [14] or Fishburn [10], allow for positive results, whereas stronger notions, such as the ones by Duggan and Schwartz [9] and Benoît [4], lead to impossibility results similar to the Gibbard-Satterthwaite theorem.

In my PhD project, I focus particularly on weak notions of strategyproofness as they allow for positive results in general. For instance, Brandt [5] has shown that attractive SCFs such as the top cycle, the uncovered set, and the bipartisan set are Kelly-strategyproof (see [14] for a definition). Even more, the top cycle is known to satisfy the slightly stronger notions of Fishburn-strategyproofness

[10] and Gärdenfors-strategyproofness [11] if preferences are strict. In particular, the last two strategyproofness notions are not well understood as there is neither a strong impossibility result nor a significant possibility result.

Note that most of these positive results break down once we enrich the domain by allowing voters to report indifferences between alternatives. In a recent result, Brandt et al. [8] have shown that Fishburn-strategyproofness is for weak preferences incompatible with Pareto-optimality and anonymity, which are a weak efficiency notion and a weak fairness notion, respectively. It is known that this result does not hold for Kelly-strategyproofness as the Pareto rule, which chooses all Pareto-optimal alternatives, satisfies this strategyproofness notion as well as anonymity and Pareto-optimality. Nevertheless, all known SCFs that satisfy Kelly-strategyproofness for weak preferences are rather indecisive. One of my first results formalizes this intuition by showing for important classes of SCFs that Kelly-strategyproofness is only possible if a large number of alternatives is chosen [7]. For instance, we have shown that for every Kelly-strategyproof SCF that satisfies the Condorcet loser property, there is an alternative that cannot be returned as single winner, not even if it is unanimously top-ranked. As the Condorcet loser property only requires that a single alternative should not be chosen for some profiles, this strong negative result affects a large number of SCFs.

3 SOCIAL DECISION SCHEMES

A second escape route to the Gibbard-Satterthwaite theorem is to allow for randomization in choosing the winner. This approach investigates social decision schemes (SDSs) which return lotteries over the alternatives. The outcome of an SDS states for every alternative its winning chance and the final winner is decided randomly according to these probabilities. Unfortunately, we face the same problem as for SCFs when defining strategyproofness for SDSs: it is not clear how voters compare lotteries on alternatives. Perhaps the most famous approach is to assume that each voter is endowed with a utility function consistent with his preference relation and tries to maximize his expected utility. Nevertheless, SDSs are still assumed to get ordinal preferences as input and consequently, strategyproofness is defined by demanding that no voter is able to manipulate, regardless of his utility function. This strategyproofness notion, often called *SD*-strategyproofness, has been investigated by Gibbard [13] and Barberà [2] who characterize the set of *SD*-strategyproof SDSs. Unfortunately, these results entail that only rather unattractive SDSs satisfy strategyproofness.

On the other hand, *SD*-strategyproofness seems too demanding because in many situations, not all possible utility functions are plausible. For instance, similar alternatives should have similar utilities for the agents, and thus, we might ignore utility functions where the difference between such alternatives is too large. Consequently, it is appealing to consider weaker strategyproofness notions, many of which are surveyed by Brandt [6]. One of the notions discussed by this author is *PC*-strategyproofness which is motivated by the idea that voters prefer a lottery p to another lottery q if p returns more likely a better outcome than q. While the intuitive appeal of *PC*-strategyproofness is clear, it is rarely considered in the literature and there are many open problems related to it.

Another interesting idea is to consider strategyproofness for a specific set of utility functions U by requiring that a voter is not able to increase his expected utility by voting dishonestly if his utility function is in U. Clearly, this strategyproofness notion is equal to *SD*-strategyproofness if U contains all utility functions, and becomes weaker if we restrict U to a smaller set. This approach allows in principle to investigate for every SDS the set U for which it is strategyproof and hence, it leads to much more detailed insights than classic strategyproofness notions. In particular, it might be possible to find an attractive SDS that is strategyproof for a large set of utility functions U, which means that it is strategyproof for many types of voters. Moreover, U-strategyproofness can be extended in multiple appealing way as we can now compare lotteries based on their expected utility. For instance, it is possible to formalize that voters are only willing to manipulate if the manipulation has a large benefit by requiring that a successful manipulation has to increase the expected utility by some minimal value.

4 CONCLUSION AND OUTLOOK

The goal of this PhD project is to investigate the boundary between possibility and impossibility results in set-valued and probabilistic social choice. First results have been obtained for set-valued social choice functions by proving that, under mild conditions, Kelly-strategyproofness entails indecisiveness if preferences are weak. Moreover, it is planned to investigate the stronger strategyproofness notions due to Fishburn [10] and Gärdenfors [11] for strict preferences as they are not well understood yet. Also, work on strategyproofness in randomized social choice is planned in the near future by investigating strategyproofness notions weaker than *SD*-strategyproofness.

In the long term, I will also consider strategyproofness in different social choice problems, for instance in the assignment domain. The assignment problem asks for the assignment of alternatives to agents according to their preferences, and can be viewed as a subproblem of the general social choice problem. Consequently, we can use many concepts from voting also in the assignment domain. In particular, it might even be possible to transfer results on strategyproofness from the voting domain to the assignment domain or vice versa, which could lead to new results. On the other hand, the differences between voting and assignment make it easier to find strategyproof assignment mechanism. In particular, a standard assumption for the assignment problem is that agents only care about their own share and not about the shares of others. This assumption simplifies the construction of strategyproof assignment mechanisms and therefore, it might be possible to find more positive results for the assignment problem.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft under grant BR 2312/12-1. I thank the anonymous reviewers for helpful comments and Felix Brandt for his constant guidance and support as supervisor.

REFERENCES

[1] S. Barberà. 1977. The Manipulation of Social Choice Mechanisms That Do Not Leave “Too Much” to Chance. *Econometrica* 45, 7 (1977), 1573–1588.

[2] S. Barberà. 1979. Majority and Positional Voting in a Probabilistic Framework. *Review of Economic Studies* 46, 2 (1979), 379–389.
[3] S. Barberà, B. Dutta, and A. Sen. 2001. Strategy-proof social choice correspondences. *Journal of Economic Theory* 101, 2 (2001), 374–394.
[4] J.-P. Benoît. 2002. Strategic Manipulation in Voting Games When Lotteries and Ties Are Permitted. *Journal of Economic Theory* 102, 2 (2002), 421–436.
[5] F. Brandt. 2015. Set-Monotonicity Implies Kelly-Strategyproofness. *Social Choice and Welfare* 45, 4 (2015), 793–804.
[6] F. Brandt. 2017. Rolling the Dice: Recent Results in Probabilistic Social Choice. In *Trends in Computational Social Choice*, U. Endriss (Ed.). AI Access, Chapter 1, 3–26.
[7] F. Brandt, M. Bullinger, and P. Lederer. 2021. On the Indecisiveness of Kelly-Strategyproof Social Choice Functions. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. Forthcoming.
[8] F. Brandt, C. Saile, and C. Stricker. 2018. Voting with Ties: Strong Impossibilities via SAT Solving. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 1285–1293.
[9] J. Duggan and T. Schwartz. 2000. Strategic Manipulability without Resoluteness or Shared Beliefs: Gibbard-Satterthwaite Generalized. *Social Choice and Welfare* 17, 1 (2000), 85–93.
[10] P. C. Fishburn. 1972. Even-chance lotteries in social choice theory. *Theory and Decision* 3, 1 (1972), 18–40.
[11] P. Gärdenfors. 1976. Manipulation of Social Choice Functions. *Journal of Economic Theory* 13, 2 (1976), 217–228.
[12] A. Gibbard. 1973. Manipulation of Voting Schemes: A General Result. *Econometrica* 41, 4 (1973), 587–601.
[13] A. Gibbard. 1977. Manipulation of schemes that mix voting with chance. *Econometrica* 45, 3 (1977), 665–681.
[14] J. S. Kelly. 1977. Strategy-Proofness and Social Choice Functions Without Single-Valuedness. *Econometrica* 45, 2 (1977), 439–446.
[15] M. A. Satterthwaite. 1975. Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory* 10, 2 (1975), 187–217.
[16] A. D. Taylor. 2005. *Social Choice and the Mathematics of Manipulation*. Cambridge University Press.