Proving the Incompatibility of Efficiency and Strategyproofness via SMT Solving

Florian Brandl Felix Brandt Christian Geist

Abstract

Two important requirements when aggregating the preferences of multiple agents are that the outcome should be economically efficient and the aggregation mechanism should not be manipulable. In this paper, we provide a computer-aided proof of a sweeping impossibility using these two conditions for randomized aggregation mechanisms. More precisely, we show that every efficient aggregation mechanism can be manipulated for *all* expected utility representations of the agents' preferences. This settles a conjecture by Aziz *et al.* [2013b] and strengthens a number of existing theorems, including statements that were shown within the special domain of assignment. Our proof is obtained by formulating the claim as a satisfiability problem over predicates from real-valued arithmetic, which is then checked using an SMT (satisfiability modulo theories) solver. To the best of our knowledge, this is the first application of SMT solvers in computational social choice.

1 Introduction

Models and results from microeconomic theory, in particular from game theory and social choice, have proven to be very valuable when reasoning about computational multiagent systems. Two fundamental notions in this context are efficiency—no agent can be made better off without making another one worse off—and strategyproofness—no agent can obtain a more preferred outcome by manipulating his preferences. Gibbard [1973] and Satterthwaite [1975] have shown that every strategyproof social choice function is either dictatorial or imposing. Hence, strategyproofness can only be achieved at the cost of discriminating among the agents or among the alternatives. One natural possibility to restore fairness, which is particularly popular in computer science, is to allow for randomization. Functions that map a profile of individual preferences to a probability distribution over alternatives (a so-called *lottery*) are known as *social decision schemes (SDSs)*.

Generalizing his previous result, Gibbard [1977] proved that the only strategyproof and *ex post* efficient social decision schemes are randomizations over dictatorships. Gibbard's notion of strategyproofness requires that no agent is better off by manipulating his preferences for *some* expected utility representation of the agents' ordinal preferences. This condition is quite demanding because an SDS may be deemed manipulable just because it can be manipulated for a contrived and highly unlikely utility representation. In this paper, we adopt a weaker notion of strategyproofness, first used by Postlewaite and Schmeidler [1986] and popularized by Bogomolnaia and Moulin [2001]. This notion requires that no agent should be better off by manipulating his preferences for *all* expected utility representations of the agents' preferences. At the same time, we use a stronger notion of efficiency than Gibbard [1977]. This notion is defined in analogy to our notion of strategyproofness and requires that no agent can be made better off for *all* utility representations of the agents' preferences, without making another one worse off for *some* utility representation. This

A shortened version of this paper appears in the *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, AAAI Press, 2016.

type of efficiency was introduced by Bogomolnaia and Moulin [2001] and is also known as ordinal efficiency or *SD*-efficiency where *SD* stands for stochastic dominance.

Our main result establishes that no anonymous and neutral SDS satisfies efficiency and strategyproofness. This settles a conjecture by Aziz *et al.* [2013b] and generalizes theorems by Aziz *et al.* [2013b], Aziz *et al.* [2014], and Brandl *et al.* [2016b]. It also strengthens related statements by Zhou [1990], Bogomolnaia and Moulin [2001], and Katta and Sethuraman [2006], which were shown within the special domain of assignment.

Our proof of this theorem heavily relies on computer-aided solving techniques. Some of these have already been applied in computational social choice, where, due to the rigorous axiomatic foundation, computer-aided theorem proving appears to be a particularly promising line of research. Perhaps the best known result in this context stems from Tang and Lin [2009], who reduce well-known impossibility results, such as Arrow's theorem, to finite instances, which can then be checked by a Boolean satisfiability (SAT) solver. Their work has sparked a number of contributions which, besides using this general idea for more complex settings or axioms, focus on proving *novel* results [Geist and Endriss, 2011; Brandl *et al.*, 2015; Brandt *et al.*, 2016; Brandt and Geist, 2016].

In this paper, we go beyond the SAT-based techniques of previous contributions by designing an SMT (satisfiability modulo theories) encoding that captures axioms for *randomized* social choice. SMT can be viewed as an enriched form of the satisfiability problem (SAT) where Boolean variables are replaced by statements from a *theory*, such as specific data types or arithmetics. Similar to SAT, there is a range of SMT solvers developed by an active community that runs annual competitions [Barrett *et al.*, 2013]. Typically, SMT solvers are used as backends for verification tasks such as the verification of software. To capture axioms about lotteries, we use the theory of (quantifier-free) linear real arithmetic. Solving this version of SMT can be seen as an extension to *linear programming* in which arbitrary Boolean operators are allowed to connect (in-)equalities.

We follow the idea of Brandt and Geist [2016] and extract a *minimal unsatisfiable set (MUS)* of constraints in order to verify our result. Despite its relatively complex 94 (non-trivial) constraints, which operate on 47 canonical preference profiles, the MUS enables manual and computer-aided verification of the encoding, and, hence, releases any need to verify our program for generating it.

2 The Model

Let A be a finite set of m alternatives and $N = \{1, \dots, n\}$ a set of agents. A *(weak) preference relation* is a complete and transitive binary relation on A. The preference relation reported by agent i is denoted by $\succsim_i$, and the set of all preference relations by $\mathcal{R}$. In accordance with conventional notation, we write $\succ_i$ for the strict part of $\succsim_i$, i.e., $x \succ_i y$ if $x \succsim_i y$ but not $y \succsim_i x$, and $\sim_i$ for the indifference part of $\succsim_i$, i.e., $x \sim_i y$ if $x \succsim_i y$ and $y \succsim_i x$. A preference relation $\succsim_i$ is *linear* if $x \succ_i y$ or $y \succ_i x$ for all distinct alternatives $x, y \in A$. We will compactly represent a preference relation as a comma-separated list with all alternatives among which an agent is indifferent placed in a set. For example, $x \succ_i y \sim_i z$ is represented by $\succsim_i \colon x, \{y, z\}$. A *preference profile* $R = (\succsim_1, \dots, \succsim_n)$ is an n-tuple containing a preference relation $\succsim_i$ for each agent $i \in N$. The set of all preference profiles is thus given by $\mathcal{R}^N$. For a given $R \in \mathcal{R}^N$ and $\succsim \in \mathcal{R}$, $R^{i \mapsto \succsim}$ denotes a preference profile identical to R except that $\succsim_i$ is replaced with $\succsim$, i.e., $R^{i \mapsto \succsim} = R \setminus \{(i, \succsim_i)\} \cup \{(i, \succsim)\}$.

2.1 Social Decision Schemes

Our central objects of study are social decision schemes: functions that map a preference profile to a *lottery (or probability distribution)* over the alternatives. The set of all lotteries

over A is denoted by $\Delta(A)$, i.e., $\Delta(A) = \{p \in \mathbb{R}^A_{\geq 0} \colon \sum_{x \in A} p(x) = 1\}$, where $p(x)$ is the probability that p assigns to x. Then, formally, a *social decision scheme (SDS)* is a function $f \colon \mathcal{R}^N \to \Delta(A)$. By $\mathrm{supp}(p)$ we denote the *support* of a lottery $p \in \Delta(A)$, i.e., the set of all alternatives to which p assigns positive probability. Two common minimal fairness conditions for SDSs are anonymity and neutrality, i.e., symmetry with respect to agents and alternatives, respectively. Formally, *anonymity* requires that $f(R) = f(R \circ \sigma)$ for all $R \in \mathcal{R}^N$ and permutations $\sigma \colon N \to N$ over agents. *Neutrality*, on the other hand, is defined via permutations over alternatives. An SDS f is *neutral* if $f(R)(x) = f(\pi(R))(\pi(x))$ for all $R \in \mathcal{R}^N$, permutations $\pi \colon A \to A$, and $x \in A$.[1]

2.2 Efficiency and Strategyproofness

Many important properties of SDSs, such as efficiency and strategyproofness, require us to reason about the preferences that agents have over lotteries. This is commonly achieved by assuming that in a preference profile R every agent i, in addition to this preference relation $\succsim_i$, is equipped with a von Neumann-Morgenstern (vNM) *utility function* $u_i^R \colon A \to \mathbb{R}$. By definition, a utility function u_i^R has to be consistent with the ordinal preferences, i.e., for all $x, y \in A$, $u_i^R(x) \geq u_i^R(y)$ iff $x \succsim_i y$. A *utility representation* u then associates with each preference profile R an n-tuple $(u_1^R, \ldots, u_n^R)$ of such utility functions. Whenever the preference profile R is clear from the context, the superscript will be omitted and we write u_i instead of the more cumbersome u_i^R.

Given a utility function u_i, agent i prefers lottery p to lottery q iff the expected utility for p is at least as high as that of q. With slight abuse of notation the domain of utility functions can be extended in the canonical way to $\Delta(A)$ by letting

$$u_i(p) = \sum_{x \in A} p(x) u_i(x).$$

It is straightforward to define efficiency and strategyproofness using expected utility. For a given utility representation u and a preference profile R, a lottery p *u-(Pareto-)dominates* another lottery q if

$$u_i(p) \geq u_i(q) \text{ for all } i \in N, \text{ and}$$
$$u_i(p) > u_i(q) \text{ for some } i \in N.$$

An SDS f is *u-efficient* if it never returns u-dominated lotteries, i.e., for all $R \in \mathcal{R}^N$, $f(R)$ is not u-dominated. The notion of u-strategyproofness can be defined analogously: for a given utility representation u, an SDS can be *u-manipulated* if there are $R \in \mathcal{R}^N$, $i \in N$, and $\succsim \in \mathcal{R}$ such that

$$u_i^R(f(R^{i \mapsto \succsim})) > u_i^R(f(R)).$$

An SDS is *u-strategyproof* if it cannot be u-manipulated.

The assumption that the vNM utility functions of all agents (and thus their complete preferences *over lotteries*) are known is quite unrealistic. Often even the agents themselves are uncertain about their preferences over lotteries and only know their ordinal preferences over alternatives.[2] A natural way to model this uncertainty is to leave the utility functions unspecified and instead *quantify over all utility functions* that are consistent with

[1] $\pi(R)$ is the preference profile obtained from π by replacing $\succsim_i$ with $\succsim_i^\pi$ for every $i \in N$, where $\pi(x) \succsim_i^\pi \pi(y)$ if and only if $x \succsim_i y$.

[2] When assuming that all agents possess vNM utility functions, these utility functions could be taken as inputs for the aggregation function. Such aggregation functions are called *cardinal decision schemes* [see, e.g., Dutta *et al.*, 2007]. In addition to the fact that concrete vNM utility functions are typically unavailable, their representation may require infinite space.

the agents' ordinal preferences. This model leads to much weaker notions of efficiency and strategyproofness.

Definition 1. An SDS is *efficient* if it never returns a lottery that is u-dominated for all utility representations u.

As mentioned in the introduction, this notion of efficiency is also known as *ordinal efficiency* or *SD-efficiency* [see, e.g., Bogomolnaia and Moulin, 2001; Aziz *et al.*, 2014, 2015]. The relationship to stochastic dominance will be discussed in more detail in Section 4.2.

Example 1. For illustration consider $A = \{a, b, c, d\}$ and the preference profile $R = (\succsim_1, \ldots, \succsim_4)$,

$$\succsim_1 \colon \{a, c\}, \{b, d\}, \qquad \succsim_2 \colon \{b, d\}, \{a, c\},$$
$$\succsim_3 \colon \{a, d\}, b, c, \qquad \succsim_4 \colon \{b, c\}, a, d$$

Observe that the lottery $^7\!/_{24}\, a + {}^7\!/_{24}\, b + {}^5\!/_{24}\, c + {}^5\!/_{24}\, d$, which is returned by the well-known SDS *random serial dictatorship (RSD)*, is u-dominated by $^1\!/_2\, a + {}^1\!/_2\, b$ for every utility representation u. Hence, any SDS that returns this lottery for the profile R would not be efficient. On the other hand, the lottery $^1\!/_2\, a + {}^1\!/_2\, b$ is not u-dominated, which can, for instance, be checked via linear programming (see Lemma 4).

We can also define a weak notion of strategyproofness in analogy to our notion of efficiency.

Definition 2. An SDS is *strategyproof* if it cannot be u-manipulated for all utility representations u.

Alternatively, there is a stronger version of strategyproofness by Gibbard [1977], in which an SDS should not be u-manipulable for *some* utility representation u.

For more information concerning the relationship between sets of possible utility functions and preference extensions, such as stochastic dominance, the reader is referred to Aziz *et al.* [2015].

3 The Result

Our main result shows that efficiency and strategyproofness are incompatible with basic fairness properties. Aziz *et al.* [2013b] raised the question whether there exists an anonymous, efficient, and strategyproof SDS. When additionally requiring neutrality, we can answer this question in the negative.

Theorem 1. *If $m \geq 4$ and $n \geq 4$, there is no anonymous and neutral SDS that satisfies efficiency and strategyproofness.*

The proof of Theorem 1, which heavily relies on computer-aided solving techniques, is discussed in Section 4. Let us first discuss the independence of the axioms and relate the result to existing theorems. *RSD* satisfies all axioms *except efficiency*; another SDS known as *maximal lotteries* satisfies all axioms *except strategyproofness* [cf. Aziz *et al.*, 2013b]. Serial dictatorship, the deterministic version of *RSD*, satisfies neutrality, efficiency, and strategyproofness *but violates anonymity*. It is unknown whether Theorem 1 still holds when dropping the assumption of neutrality. Our proof, however, only requires a technical weakening of neutrality (cf. Section 4.1).

3.1 Related Results for Social Choice

Our result generalizes several existing results and is closely related to a number of results in subdomains of social choice. Aziz *et al.* [2013b] proved a weak version of Theorem 1 for the rather restricted class of majoritarian SDSs, i.e., SDSs whose outcome may only depend on the pairwise majority relation. This statement has later been generalized by Aziz *et al.* [2014] to all SDSs whose outcome only depends on the *weighted* majority relation. More recently, Brandl *et al.* [2016b] have shown that while random dictatorship is efficient and strategyproof on the domain of linear preferences, it cannot be extended to the full domain of weak preferences without violating at least one of these properties. Their theorem, which also assumes anonymity and neutrality, is a direct consequence of Theorem 1. Other impossibility results have been obtained for stronger notions of efficiency and strategyproofness, which weakens the corresponding statements. Aziz *et al.* [2014] have shown that there is no anonymous and neutral SDS that satisfies efficiency and strategyproofness with respect to the *pairwise comparison* lottery extension and with respect to the *upward lexicographic* extension.[3] Both of these notions of efficiency and strategyproofness are stronger than the ones used in Theorem 1.

3.2 Related Results for Assignment

A subdomain of social choice that has been thoroughly studied in the literature is the assignment (aka house allocation or two-sided matching with one-sided preferences) domain. An assignment problem can be associated with a social choice problem by letting the set of alternatives be the set of deterministic allocations and postulating that agents are indifferent among all allocations in which they receive the same object [see, e.g., Aziz *et al.*, 2013a].[4] Thus, impossibility results for the assignment setting can be interpreted as impossibility results for the social choice setting because they even hold in a smaller domain.

In the following we discuss impossibility results in the assignment domain which, if interpreted for the social choice domain, can be seen as weaker versions of Theorem 1 because they are based on stronger notions of efficiency or strategyproofness or require additional properties. In a very influential paper, Bogomolnaia and Moulin [2001] have shown that no randomized assignment mechanism satisfies both efficiency and a strong notion of strategyproofness while treating all agents equally. The underlying notion of strategyproofness is identical to the one used by Gibbard [1977] and prescribes that the SDS cannot be u-manipulated for *some* utility representation u. The result by Bogomolnaia and Moulin even holds when preferences over objects are linear. (Nevertheless, when transferred to the social choice domain, the preferences over allocations will contain ties.) In a related paper, Katta and Sethuraman [2006] proved that no assignment mechanism satisfies efficiency, strategyproofness, and envy-freeness for the full domain of preferences.[5]

Settling a conjecture by Gale [1987], Zhou [1990] showed that no cardinal assignment mechanism satisfies u-efficiency and u-strategyproofness while treating all agents equally.[6] The relationship between Zhou's result and Theorem 1 is not obvious because Zhou's the-

[3]The statement for the pairwise comparison extension holds for at least three agents and three alternatives, whereas Theorem 1 does not hold for less then four alternatives since *RSD* satisfies all properties for up to three alternatives. In contrast to Theorem 1, the statement for the upward lexicographic extension does not require neutrality and also holds for linear preferences.

[4]Note that this transformation turns assignment problems with linear preferences over k objects into social choice problems with non-linear preferences over $k!$ allocations.

[5]Envy-freeness is a fairness property that is stronger than *equal treatment of equals* as used by Bogomolnaia and Moulin [2001].

[6]The theorem by Zhou only requires that agents with the same utility function receive the same amount of utility but not necessarily the same assignment. Gale's original conjecture assumed equal treatment of equals.

orem concerns cardinal mechanisms, i.e., functions that take a utility profile rather than a preference profile as input. However, every cardinal assignment mechanism can be associated with an ordinal assignment mechanism by choosing the outcome for some consistent utility profile for every preference profile. This transformation turns a u-efficient and u-strategyproof cardinal mechanism into an efficient and strategyproof ordinal mechanism as these properties are purely ordinal. Hence, Theorem 1 implies that there is no anonymous, neutral, u-efficient, and u-strategyproof cardinal decision scheme.

4 Proving the Result

In this section, we first reduce the statement of Theorem 1 to the case of $m = 4$ and $n = 4$, which we then prove via SMT solving. We present an encoding for any finite instance of Theorem 1 as an SMT problem in the logic of (quantifier-free) linear real arithmetic (`QF_LRA`). For compatibility with different SMT solvers our encoding adheres to the SMT-LIB standard [Barrett *et al.*, 2010]. In total, we are going to design the following four types of SMT constraints:

- lottery definitions (Lottery),
- the orbit condition[7] (Orbit),
- strategyproofness (SP), and
- efficiency (Efficiency).

Other conditions such as anonymity are taken care of by the representation of preference profiles.

We then, first, apply an SMT solver to show that this set of constraints for the case of $m = 4$ and $n = 4$ is unsatisfiable, i.e., no SDS f with the desired properties exists. Second, we explain how the output of the solver can be used to obtain a human-verifiable proof of this result.

But let us start with the reduction lemma before we turn to the concrete encoding in the following subsections.

Lemma 1. *If there is an anonymous and neutral SDS f that satisfies efficiency and strategyproofness for $|A| = m$ alternatives and $|N| = n$ agents then we can also find an SDS f' defined for $m' \leq m$ alternatives and $n' \leq n$ agents that satisfies the same properties.*

Proof. Let f be an anonymous and neutral SDS that satisfies efficiency and strategyproofness for m alternatives and n agents. We define a projection f' of f onto $A' \subseteq A, |A'| = m' \leq m$ and $N' = \{1, 2, \ldots, n'\} \subseteq N, n' \leq n$ that satisfies all required properties:

For every preference profile R' on A' and N', let $f'(R') = f(R)$, where R is defined by the following conditions:

$$\succsim_i \cap (A' \times A') = \succsim'_i \text{ for all } i \in N', \tag{1}$$

$$x \succ_i y \text{ for all } x \in A', y \in A \setminus A' \text{ and } i \in N', \tag{2}$$

$$y \sim_i z \text{ for all } y, z \in A \setminus A' \text{ and } i \in N', \text{ and} \tag{3}$$

$$y \sim_i z \text{ for all } y, z \in A \text{ and } i \in N \setminus N'. \tag{4}$$

Informally, by (1) agents in N' have the same preferences over alternatives from A' in R and R'. Moreover, by (2) they like every alternative in A' strictly better than every alternative

[7] The orbit condition models a part of neutrality.

not in A' and by (3) they are indifferent between all alternatives not in A'. Finally, by (4) all agents in $N \setminus N'$ are completely indifferent. With these conditions, R is uniquely specified given R', and only lotteries p with $\text{supp}(p) \subseteq A'$ are efficient in R. Thus, f' is well-defined and it is left to show that f' inherits the relevant properties from f. The SDS f' is anonymous since f is anonymous and agents in N can only differ by their preferences over A'. Neutrality follows as f is neutral and all agents are indifferent between all alternatives not in A'. Efficiency is satisfied by f' since f is efficient and the same set of lotteries is efficient in R and R'. Finally, f' is strategyproof because f is strategyproof and the outcomes of f' under the two profiles R' and $(R')^{i \mapsto \succsim'}$ are equal to the outcomes of f under the two (extended) profiles R and $R^{i \mapsto \succsim}$, respectively. □

4.1 Framework, Anonymity, and Neutrality

For a given number of agents n and set of alternatives A, we encode an arbitrary SDS $f \colon \mathcal{R}^N \to \Delta(A)$ by a set of real-valued variables $p_{R,x}$ with $R \in \mathcal{R}^N$ and $x \in A$. Each $p_{R,x}$ then represents the probability with which alternative x is selected for profile R, i.e., $p_{R,x} = f(R)(x)$.

This encoding of lotteries leads to the first simple constraints for our SMT encoding, which ensure that for each preference profile R the corresponding variables $p_{R,x}$, $x \in A$ indeed encode a lottery:

$$\begin{aligned} \sum_{x \in A} p_{R,x} = 1 \text{ for all } R \in \mathcal{R}^N, \text{ and} \\ p_{R,x} \geq 0 \text{ for all } R \in \mathcal{R}^N \text{ and } x \in A. \end{aligned} \tag{Lottery}$$

We are now going to argue that, in conjunction with anonymity and neutrality (see Section 2), it suffices to consider these constraints for a subset of preference profiles. This is because, in contrast to the other axioms, we directly incorporate anonymity and neutrality into the structure of the encoding rather than formulating them as actual constraints. Similar to the construction involving canonical tournament representations by Brandt and Geist [2016], we model anonymity and neutrality by computing for each preference profile $R \in \mathcal{R}^N$ a *canonical representation* $R_{\mathfrak{c}} \in \mathcal{R}^N$ with respect to these properties. In this representation, two preference profiles R and R' are equal (i.e., $R_{\mathfrak{c}} = R'_{\mathfrak{c}}$) iff one can be transformed into the other by renaming the agents and alternatives. Equivalently, $R_{\mathfrak{c}} = R'_{\mathfrak{c}}$ iff, for every anonymous and neutral SDS f, the lotteries $f(R)$ and $f(R')$ are equal (modulo the renaming of the alternatives).

The SMT constraints and SMT variables are then instantiated only for these canonical representations $\mathcal{R}^N_{\mathfrak{c}} \subseteq \mathcal{R}^N$. Apart from enabling an encoding of anonymous and neutral SDSs without any explicit reference to permutations, this also offers a substantial performance gain compared to considering the full domain $\mathcal{R}^N$ of (non-anonymous and non-neutral) preference profiles.

Technically, we compute the canonical representation $R_{\mathfrak{c}}$ as follows: Let $R = (\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^N$ be a preference profile. First, we identify R with a function $r : \mathcal{R} \to \mathbb{N}$, which we call *anonymous preference profile*, and which counts the number of agents with a certain preference relation, i.e., $r(\succsim) = |\{i \in N \mid \succsim_i = \succsim\}|$, thereby ignoring the identity of the agents. This representation fully captures anonymity.

To additionally enforce neutrality, we had to resort to a computationally demanding, naive solution: given r, we compute all anonymous preference profiles $\pi(r)$ that can be achieved via a permutation $\pi \colon A \to A$, and, among those profiles, choose the one $\pi_{\text{lexmin}}(r)$ with lexicographically minimal values (for some fixed ordering of preference relations). For the canonical representation $R_{\mathfrak{c}}$ we then pick any preference profile $R \in \mathcal{R}^N$ which agrees

with $\pi_{\text{lexmin}}(r)$, for instance, by again using the same fixed ordering of preference relations. Fortunately, this approach is still feasible for the small numbers of alternatives with which we are dealing.

While this representation of preference profiles does not completely capture neutrality—the *orbit condition* [see Brandt and Geist, 2016] is missing—this weaker version suffices to prove the impossibility. In favor of simpler proofs we, however, include the simple constraints corresponding to a randomized version of the orbit condition.

In our context, an *orbit* O of a preference profile R is an equivalence class of alternatives. Two alternatives $x, y \in A$ are considered equivalent if $\pi(x) = y$ for some permutation $\pi\colon A \to A$ that maps the anonymous preference profile associated with R to itself (i.e., π is an automorphism of the anonymous preference profile). In such a situation, every anonymous and neutral SDS has to assign equal probabilities to x and y. We hence require that, for each orbit $O \in \mathcal{O}_R$ of a (canonical) profile R, the probabilities $p_{R,x}$ are equal for all alternatives $x \in O$. As an SMT constraint, this reads

$$p_{R,x} = p_{R,y} \tag{Orbit}$$

for all $R \in \mathcal{R}_{\mathfrak{c}}^N$, $O \in \mathcal{O}_R$, and $x, y \in O$.

Example 2. Consider the anonymous preference profile r based on R from Example 1 and the permutation

$$\pi = \begin{pmatrix} a & b & c & d \\ b & a & d & c \end{pmatrix}.$$

As $\pi(r) = r$ (and since no other non-trivial permutation has this property) the set of orbits of R is $\mathcal{O}_R = \{\{a, b\}, \{c, d\}\}$.

4.2 Stochastic Dominance

In order to avoid quantifying over utility functions, we leverage well-known representations of efficiency and strategyproofness via *stochastic dominance (SD)* [cf. Bogomolnaia and Moulin, 2001; McLennan, 2002; Aziz *et al.*, 2015]. A lottery p *stochastically dominates* a lottery q for an agent i (short: $p \succsim_i^{SD} q$) if for every alternative x, lottery p is at least as likely as lottery q to yield an alternative at least as good as x. Formally,

$$p \succsim_i^{SD} q \text{ iff } \sum_{y \succsim_i x} p(y) \geq \sum_{y \succsim_i x} q(y) \text{ for all } x \in A.$$

When $p \succsim_i^{SD} q$ and not $q \succsim_i^{SD} p$ we write $p \succ_i^{SD} q$.

As an example, consider the preference relation $\succsim_i\colon a, b, c$. We then have that

$$(2/3\, a + 1/3\, c) \succ_i^{SD} (1/3\, a + 1/3\, b + 1/3\, c)$$

while $2/3\, a + 1/3\, c$ and b are incomparable according to stochastic dominance.

Lemma 2. *Let $\succsim_i \in \mathcal{R}$. A lottery p SD-dominates another lottery q for an agent i iff $u_i(p) \geq u_i(q)$ for every utility function u_i consistent with $\succsim_i$. As a consequence,*

1. *an SDS f is efficient iff, for all $R \in \mathcal{R}^N$, there is no lottery p such that $p \succsim_i^{SD} f(R)$ for all $i \in N$ and $p \succ_i^{SD} f(R)$ for some $i \in N$, and*
2. *an SDS f is manipulable iff there exist a preference profile R, an agent i, and a preference relation $\succsim$ such that $f(R^{i \mapsto \succsim}) \succ_i^{SD} f(R)$.*

Proof. For the direction from left to right, assume that $p \succsim_i^{SD} q$. Let $A = \{x_1, \ldots, x_m\}$ and $x_j \succsim_i x_k$ if and only if $j \leq k$ for all $j, k \in \{1, \ldots, m\}$. Then, by definition, for all $j \in \{1, \ldots, m\}$, $\sum_{k=1}^{j} p(x_k) \geq \sum_{k=1}^{j} q(x_k)$. Let u_i be a utility function consistent with $\succsim_i$, i.e., $u_i(x_j) \geq u_i(x_k)$ if and only if $j \leq k$. Then,

$$u_i(p) - u_i(q) = \sum_{j=1}^{m} (p(x_j) - q(x_j)) u_i(x_j) = \sum_{j=1}^{m} \underbrace{(u_i(x_j) - u_i(x_{j+1}))}_{\geq 0} \underbrace{\sum_{k=1}^{j} (p(x_k) - q(x_k))}_{\geq 0} \geq 0,$$

where $u_i(x_{m+1})$ is set to 0. Hence, $u_i(p) \geq u_i(q)$.

For the direction from right to left, assume that $u_i(p) \geq u_i(q)$ for all utility functions u_i consistent with $\succsim_i$. Assume for contradiction that $p \not\succsim_i^{SD} q$, i.e., there is $x \in A$ such that $\sum_{y \succsim_i x} q(x) - \sum_{y \succ_i x} p(x) = \epsilon > 0$. Let u_i be a utility function consistent with $\succsim_i$ such that $u_i(y) \in [1 - \epsilon/2, 1]$ for all $y \succsim_i x$ and $u_i(y) \in [0, \epsilon/2]$ for all $x \succ_i y$. Such a u_i exists, since $\epsilon > 0$. Then,

$$u_i(q) \geq (1 - \epsilon/2) \sum_{y \succsim_i x} q(y) > \sum_{y \succsim_i x} p(y) + \epsilon/2 \geq u_i(p),$$

which contradicts the assumption. □

In words, Lemma 2 shows that an SDS f is efficient if and only if $f(R)$ is Pareto-efficient with respect to stochastic dominance for all preference profiles R. Secondly, f is manipulable if and only if some agent can misrepresent his preferences to obtain a lottery that he prefers to the lottery obtained by sincere voting with respect to stochastic dominance.

4.2.1 Encoding Strategyproofness

Starting from the above equivalence, encoding strategyproofness as an SMT constraint is now a much simpler task. For each (canonical) preference profile $R \in \mathcal{R}_{\mathfrak{c}}^N$, agent $i \in N$,[8] and preference relation $\succsim \in \mathcal{R}$, we encode that the manipulated outcome $f(R^{i \mapsto \succsim})$ is not *SD*-preferred to the truthful outcome $f(R)$ by agent i:

$$\begin{aligned}
&\neg \left(f(R^{i \mapsto \succsim}) \succ_i^{SD} f(R) \right) \\
\equiv\ & f(R^{i \mapsto \succsim}) \not\succsim_i^{SD} f(R) \vee f(R) \succsim_i^{SD} f(R^{i \mapsto \succsim}) \\
\equiv\ & \left((\exists x \in A) \sum_{y \succsim_i x} f(R^{i \mapsto \succsim})(y) < \sum_{y \succsim_i x} f(R)(y) \right) \vee \\
& \left((\forall x \in A) \sum_{y \succsim_i x} f(R^{i \mapsto \succsim})(y) \overset{(*)}{\leq} \sum_{y \succsim_i x} f(R)(y) \right) \\
\equiv\ & \left(\bigvee_{x \in A} \sum_{y \succsim_i x} p_{(R^{i \mapsto \succsim})_{\mathfrak{c}}, \pi_{\mathfrak{c}}^{R^{i \mapsto \succsim}}(y)} < \sum_{y \succsim_i x} p_{R,y} \right) \vee \\
& \left(\bigwedge_{x \in A} \sum_{y \succsim_i x} p_{(R^{i \mapsto \succsim})_{\mathfrak{c}}, \pi_{\mathfrak{c}}^{R^{i \mapsto \succsim}}(y)} \overset{(**)}{=} \sum_{y \succsim_i x} p_{R,y} \right),
\end{aligned} \tag{SP}$$

[8] Note that, due to anonymity, it is not necessary to iterate over all agents i. Rather it suffices to pick one agent per unique preference relation contained in R.

where $\pi_{\mathfrak{c}}^{R^{i\mapsto\succsim}}$ stands for a permutation of alternatives that (together with a potential renaming of alternatives) leads from $R^{i\mapsto\succsim}$ to $(R^{i\mapsto\succsim})_{\mathfrak{c}}$. The inequality $(*)$ can be replaced by the equality $(**)$ since the case of at least one strict inequality is captured by the corresponding disjunctive condition one line above.

4.2.2 Encoding Efficiency

While Lemma 2 helps to formulate efficiency as an SMT axiom it is not yet sufficient since a quantification over the set of all lotteries $\Delta(A)$ remains. In order to get rid of this quantifier, we apply two lemmas by Aziz *et al.* [2015], for which we include (slightly simplified) proofs in favor of a self-contained presentation. The first lemma states that efficiency of a lottery only depends on its support. The second lemma shows that deciding whether a lottery is efficient reduces to solving a linear program.

Lemma 3 (Aziz *et al.*, 2015)**.** *Let $R \in \mathcal{R}^N$. A lottery $p \in \Delta(A)$ is efficient iff every lottery $p' \in \Delta(A)$ with $\operatorname{supp}(p') \subseteq \operatorname{supp}(p)$ is efficient.*

Proof. We prove the statement by contraposition: if $p' \in \Delta(A)$ is not efficient, then no lottery p with $\operatorname{supp}(p') \subseteq \operatorname{supp}(p)$ is efficient. If p' is not efficient, there is $q' \in \Delta(A)$ such that q' u-dominates p' for all utility representations u^R, i.e., for all agents $i \in N$ and all utility functions u_i consistent with $\succsim_i$, $u_i(q') - u_i(p') \geq 0$ and $u_{i'}(q') - u_{i'}(p') > 0$ for some agent $i' \in N$ and all utility functions $u_{i'}$ consistent with $\succsim_{i'}$. Let $v = q' - p' \in \mathbb{R}^A$. Note that, for all $x \in A$, $v(x) < 0$ implies $x \in \operatorname{supp}(p')$. Now let $q = p + \epsilon v$ for $\epsilon > 0$ small enough such that $q \in \Delta(A)$. This is possible because $\operatorname{supp}(p') \subseteq \operatorname{supp}(p)$. By definition of q, we have that, for all $i \in N$ and all u_i consistent with $\succsim_i$, $u_i(q) - u_i(p) = \epsilon u_i(v) = \epsilon(u_i(q') - u_i(p')) \geq 0$ and $u_{i'}(q) - u_{i'}(p) > 0$ for all $u_{i'}$ consistent with $\succsim_{i'}$. Thus, p is not efficient. □

Lemma 4 (Aziz *et al.*, 2015)**.** *Whether a lottery $p \in \Delta(A)$ is efficient for a given preference profile R can be computed in polynomial time by solving a linear program.*

Proof. Given the equivalence from Lemma 2, a lottery p is easily seen to be efficient iff the optimal objective value of the following linear program is zero (since then there is no lottery q that *SD*-dominates p):

$$\begin{aligned} \max_{q,r} \sum_{i\in N}\sum_{x\in A} r_{i,x} \quad & \text{subject to} \\ \sum_{y \succsim_i x} q_y - r_{i,x} &= \sum_{y \succsim_i x} p_y \text{ for all } x \in A,\, i \in N, \\ \sum_{x\in A} q_x &= 1,\ q_x \geq 0 \text{ for all } x \in A, \\ r_{i,x} &\geq 0 \text{ for all } x \in A,\, i \in N. \end{aligned}$$

□

Recall that an SDS is efficient if it never returns a dominated lottery. By Lemma 3, this is equivalent to never returning a lottery with *inefficient support*. To capture this, we encode, for each (canonical) preference profile $R \in \mathcal{R}_{\mathfrak{c}}^N$, that the probability for at least one alternative in every (inclusion-minimal) inefficient support $I_R \subseteq A$ is zero:

$$\bigvee_{x\in I_R} p_{R,x} = 0. \qquad \text{(Efficiency)}$$

4.3 Restricted Domains

Since *RSD* (cf. Section 3) is known to satisfy both strategyproofness as well as efficiency for up to 3 alternatives, the search for an impossibility has to start at $m = 4$ alternatives. For $n = 3$ agents, the encoding is solved as satisfiable; for $n = 4$, an encoding of the full domain, unfortunately, becomes prohibitively large. Hence, for $m = 4$ and $n = 4$, one has to carefully optimize the domain under consideration, on the one hand, to include a sufficient number of profiles for a successful proof, and, on the other hand, not to include too many profiles, which would prevent the solver from terminating within a reasonable amount of time.

The following incremental strategy was found to be successful. We start with a specific profile R, from which we only consider sequences of potential manipulations as long as (in each step) the manipulated individual preferences are not too distinct from the truthful preferences. To this end, we measure the magnitude of manipulations by the Kendall tau distance τ, which counts pairwise disagreements between R_i and R'_i [see also Sato, 2013]. A change in the individual preferences of an agent will be called a *k-manipulation* if $\tau(R_i, R'_i) \leq k$. Then, for example, strategically swapping two alternatives is a 2-manipulation, and breaking or introducing a tie between two alternatives is a 1-manipulation.

On the domain which starts from the preference profile R from Example 1 and from there allows sequences of $(1, 2, 1, 2)$-manipulations[9] we were able to prove the result within a few minutes of running-time.[10] On smaller domains (e.g., considering $(1, 2, 2)$-manipulations from R) the axioms are still compatible.

4.4 Verification of Correctness

For verification of the result, one would ideally construct a human-readable proof from the output of the SMT solver. While the approach described by Brandt and Geist [2016] for SAT solving—of finding a *minimal unsatisfiable set (MUS)* of constraints, i.e., an inclusion-minimal set of constraints such that this set is still unsatisfiable—is theoretically also applicable to SMT solving, it is less clear how these "proof ingredients" have to be combined.[11] The proof object that z3 can produce, which also contains information of how the MUS constraints have to be combined, unfortunately, is too long and complicated for humans to parse.

Hence two aspects of our approach still deserve verification: the correctness of the constraints in the MUS and the unsatisfiability of the MUS. In addition to manual inspection of the constraints and some sanity-checks,[12] we have certified in Isabelle/HOL that all constraints logically follow from the original axioms presented in Section 2. This also releases any need to verify our program for generating the constraints. The unsatisfiability of the MUS, on the other hand, has been verified by the solvers CVC4, MathSAT, Yices2, z3, and even by the Isabelle/HOL kernel.

[9] I.e., first we allow any 1-manipulation from R, then, from every resulting profile, any 2-manipulation is allowed (not necessarily by the same agent), and so forth. Showing the result on this domain implies a slightly stronger statement where strategyproofness only has to hold for "small" lies (of at most Kendall tau distance 2).

[10] The SMT solver MathSAT [Cimatti *et al.*, 2013] terminates quickly within less than 3 minutes with the suggested competition settings, whereas z3 [de Moura and Bjørner, 2008] requires some additional configuration, but then also supports core extraction within the same time frame.

[11] Here we have an MUS of 94 constraints, not counting the (trivial) lottery definitions. This MUS, annotated with e.g., the 47 required canonical preference profiles, is available as part of an arXiv version of this paper [Brandl *et al.*, 2016a].

[12] Such as running solvers on multiple variants of the encoding which represent known theorems. This way, we reproduced (amongst others) the results by Bogomolnaia and Moulin [2001] and Katta and Sethuraman [2006], as well as the possibility result for $m < 4$.

Statement	Number of canonical preference profiles
Theorem 1	47
Brandl *et al.* [2016b, Theorem 1]	13
Aziz *et al.* [2014, Theorem 2]	7
Aziz *et al.* [2014, Theorem 4]	7
Aziz *et al.* [2013b, Theorem 1]	5
Aziz *et al.* [2014, Theorem 3]	3
Bogomolnaia and Moulin [2001, Theorem 2]	11
Zhou [1990, Theorem 1]	5
Katta and Sethuraman [2006, Section 4]	2

Table 1: Proof complexity comparison of impossibility statements using efficiency and strategyproofness in terms of the number of canonical preference profiles used in the proof. The last three statements have been proven for the assignment domain.

Furthermore, based on the MUS, a proof of Theorem 1 which no longer relies on SMT solving has been created in Isabelle/HOL. This proof, however, is tedious to verify by hand since it is rather large (more than 500 lines of code) and offers little insight.

5 Conclusion

In this paper, we have leveraged computer-aided solving techniques to prove a sweeping impossibility for randomized aggregation mechanisms.

It seems unlikely that this proof would have been found without the help of computers because manual proofs of significantly weaker statements already turned out to be quite complex (see Table 1 for a comparison of the proof complexity of related statements). Nevertheless, now that the theorem has been established, our computer-aided methods may guide the search for related, perhaps even stronger statements that allow for more intuitive proofs and provide more insights into randomized social choice.

Generally speaking, we believe that SMT solving is applicable to a wide range of problems in randomized social choice. In particular, extending our result to the special domain of assignment (see Section 3.2) is desirable as this would strengthen a number of existing theorems. Other interesting questions are whether the impossibility still holds when weakening strategyproofness even further to *BD*-strategyproofness [see, e.g., Aziz *et al.*, 2014] or when omitting neutrality.

6 Acknowledgments

This material is based upon work supported by the Deutsche Forschungsgemeinschaft under grants BR 2312/7-2 and BR 2312/10-1 and the TUM Institute for Advanced Study through a Hans Fischer Senior Fellowship. The authors also thank Manuel Eberl for the extensive verification work in Isabelle/HOL, Alberto Griggio and Mohammad Mehdi Pourhashem Kallehbasti for guidance on how to most effectively use MathSAT and z3, respectively, and three anonymous reviewers for their helpful comments.

References

H. Aziz, F. Brandt, and M. Brill. The computational complexity of random serial dictatorship. *Economics Letters*, 121(3):341–345, 2013.

H. Aziz, F. Brandt, and M. Brill. On the tradeoff between economic efficiency and strategyproofness in randomized social choice. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 455–462. IFAAMAS, 2013.

H. Aziz, F. Brandl, and F. Brandt. On the incompatibility of efficiency and strategyproofness in randomized social choice. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, pages 545–551. AAAI Press, 2014.

H. Aziz, F. Brandl, and F. Brandt. Universal Pareto dominance and welfare for plausible utility functions. *Journal of Mathematical Economics*, 60:123–133, 2015.

C. Barrett, A. Stump, and C. Tinelli. The SMT-LIB Standard: Version 2.0. In *Proceedings of the 8th International Workshop on Satisfiability Modulo Theories*, 2010.

C. Barrett, M. Deters, L. de Moura, A. Oliveras, and A. Stump. 6 years of SMT-COMP. *Journal of Automated Reasoning*, 50(3):243–277, 2013.

A. Bogomolnaia and H. Moulin. A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2):295–328, 2001.

F. Brandl, F. Brandt, C. Geist, and J. Hofbauer. Strategic abstention based on preference extensions: Positive results and computer-generated impossibilities. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 18–24. AAAI Press, 2015.

F. Brandl, F. Brandt, and C. Geist. Proving the incompatibility of efficiency and strategyproofness via SMT solving. Technical report, http://arxiv.org/abs/1604.05692, 2016.

F. Brandl, F. Brandt, and W. Suksompong. The impossibility of extending random dictatorship to weak preferences. *Economics Letters*, 141:44–47, 2016.

F. Brandt and C. Geist. Finding strategyproof social choice functions via SAT solving. *Journal of Artificial Intelligence Research*, 55:565–602, 2016.

F. Brandt, C. Geist, and D. Peters. Optimal bounds for the no-show paradox via SAT solving. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 314–322. IFAAMAS, 2016.

A. Cimatti, A. Griggio, B. Schaafsma, and R. Sebastiani. The MathSAT5 SMT Solver. In *Proceedings of TACAS*, volume 7795 of *Lecture Notes in Computer Science (LNCS)*, pages 93–107. Springer-Verlag, 2013.

L. de Moura and N. Bjørner. Z3: An efficient SMT solver. In *Proceedings of TACAS*, volume 7795 of *Lecture Notes in Computer Science (LNCS)*, pages 337–340. Springer-Verlag, 2008.

B. Dutta, H. Peters, and A. Sen. Strategy-proof cardinal decision schemes. *Social Choice and Welfare*, 28(1):163–179, 2007.

D. Gale. College course assignments and optimal lotteries, 1987. Mimeo.

C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.

A. Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45(3):665–681, 1977.

A.-K. Katta and J. Sethuraman. A solution to the random assignment problem on the full preference domain. *Journal of Economic Theory*, 131(1):231–250, 2006.

A. McLennan. Ordinal efficiency and the polyhedral separating hyperplane theorem. *Journal of Economic Theory*, 105(2):435–449, 2002.

A. Postlewaite and D. Schmeidler. Strategic behaviour and a notion of ex ante efficiency in a voting model. *Social Choice and Welfare*, 3(1):37–49, 1986.

S. Sato. Strategy-proofness and the reluctance to make large lies: the case of weak orders. *Social Choice and Welfare*, 40(2):479–494, 2013.

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(2):187–217, 1975.

P. Tang and F. Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

L. Zhou. On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory*, 52(1):123–135, 1990.

Florian Brandl
Department of Informatics
Technical University of Munich
Munich, Germany
Email: `brandlfl@in.tum.de`

Felix Brandt
Department of Informatics
Technical University of Munich
Munich, Germany
Email: `brandtf@in.tum.de`

Christian Geist
Department of Informatics
Technical University of Munich
Munich, Germany
Email: `geist@in.tum.de`