# Finding Strategyproof Social Choice Functions via SAT Solving

Felix Brandt and Christian Geist

**Abstract**

A promising direction in computational social choice is to address open research problems using computer-aided proving techniques. In conjunction with SAT solving, this approach has been shown to be viable in the context of classic impossibility theorems such as Arrow's impossibility as well as axiomatizations of preference extensions. In this paper, we demonstrate that it can also be applied to the more complex notion of strategyproofness for irresolute social choice functions. These types of problems, however, require a more evolved encoding as otherwise the search space rapidly becomes much too large. We present an efficient encoding for translating such problems to SAT and leverage this encoding to prove new results about strategyproofness with respect to Kelly's and Fishburn's preference extensions. For example, we show that no Pareto-optimal majoritarian social choice function satisfies Fishburn-strategyproofness.

## 1 Introduction

Ever since the famous Four Color Problem was solved using a computer-assisted approach, it is clear that computers can contribute significantly to finding and proving formal statements. Due to its rigorous axiomatic foundation, *social choice theory* appears to be a field in which computer-aided theorem proving is a particularly promising line of research. Perhaps the best known result in this context is due to Tang and Lin [21], who reduce well-known impossibility results such as Arrow's theorem to finite instances, which can then be checked by a satisfiability (SAT) solver [3]. Geist and Endriss [14] were able to extend this method to a fully-automatic search algorithm for impossibility theorems in the context of preference relations over sets of alternatives. In this paper, we apply these techniques to improve our understanding of strategyproofness in the context of set-valued (or so-called irresolute) social choice functions. These types of problems, however, are more complex and require an evolved encoding as otherwise the search space rapidly becomes too large. Table 1 illustrates how quickly the number of involved objects grows and that, therefore, an exhaustive search is doomed to fail.

Formally, a social choice function (SCF) is defined as a function that maps individual preferences over a set of alternatives to a set of socially most preferred alternatives. An SCF is strategyproof if no agent can obtain a more preferred outcome by misrepresenting her preferences. It is well-known from the Gibbard-Satterthwaite theorem that, when restricting attention to SCFs that always return a single alternative, only trivial SCFs can be strategyproof.[1] A proper definition of strategyproofness for the more general setting of irresolute SCFs requires the specification of preferences over *sets* of alternatives. Rather than asking the agents to specify their preferences over all sets (which would be bound to various rationality constraints), it is typically assumed that the preferences over single alternatives can be extended to preferences over sets. Of course, there are various ways how to extend preferences to sets (see, e.g., [13, 2, 22]), each of which leads to a different class of strategyproof SCFs. A function that yields a preference relation over subsets of alternatives

[1] The assumption of single-valuedness has been criticized for being restrictive and unreasonable (see, e.g., [12, 15, 22]).

| Alternatives | 4 | 5 | 6 | 7 |
|---|---|---|---|---|
| Choice sets | 15 | 31 | 63 | 127 |
| Tournaments | 64 | 1,024 | 32,768 | $\sim 2 \cdot 10^6$ |
| Canonical tourn. | 4 | 12 | 56 | 456 |
| **Maj. SCFs** | $\mathbf{50,625}$ | $\mathbf{\sim 10^{18}}$ | $\mathbf{\sim 10^{101}}$ | $\mathbf{\sim 10^{959}}$ |

Table 1: Number of objects involved in problems with irresolute majoritarian SCFs

when given a preference relation over single alternatives is called a *set extension* or *preference extension*. In this paper, we focus on two very natural set extensions due to Kelly [15] and Fishburn [10][2].

While strategyproofness for Kelly's extension (henceforth *Kelly-strategyproofness*) is known to be a rather restrictive condition [15, 1, 19], some SCFs such as the Pareto rule, the omninomination rule, the top cycle, the minimal covering set, and the bipartisan set were shown to be Kelly-strategyproof [6]. Interestingly, the more discriminating of these SCFs are *majoritarian*, i.e., they are based on the pairwise majority relation only, and, moreover, can be ordered with respect to set inclusion. In particular, these results suggest that the bipartisan set may be the finest Kelly-strategyproof majoritarian SCF. In this paper, we show that this not the case by automatically generating a Kelly-strategyproof SCF which is strictly contained in the bipartisan set.

For the more demanding notion of *Fishburn-strategyproofness*, existing results suggested that it may only be satisfied by rather indiscriminating SCFs such as the top cycle [20, 7, 9].[3] Using our computer-aided proving technique we were able to confirm this suspicion by proving that, within the domain of majoritarian SCFs, Fishburn-strategyproofness is incompatible with Pareto-optimality. In order to achieve this impossibility, we manually proved a novel characterization of Pareto-optimal majoritarian SCFs and an induction step, which allows us to generalize the computer-verified impossibility to larger numbers of alternatives.

The universality of our method and its ease of adaptation suggests that it could be applied to similar open questions in the future.

# 2 Mathematical Framework of Strategyproofness

In this section, we provide the terminology and notation required for our results and introduce notions of strategyproofness for majoritarian SCFs that allow us to abstract away any reference to preference profiles.

## 2.1 Social Choice Functions

Let $N = \{1, \dots, n\}$ be a set of at least 3 voters with preferences over a finite set $A$ of $m$ alternatives. For convenience, we assume that $n$ is odd.[4] The preferences of each voter $i \in N$ are represented by a complete, anti-symmetric, and transitive *preference relation* $R_i \subseteq A \times A$. The interpretation of $(a, b) \in R_i$, usually denoted by $a \; R_i \; b$, is that voter $i$ values alternative $a$ at least as much as alternative $b$. The set of all preference relations over $A$ will be denoted by $\mathcal{R}(A)$. The set of *preference profiles*, i.e., finite vectors of preference

[2]Gärdenfors [13] attributed this extension to Fishburn because it is the weakest extension that satisfies a certain set of axioms proposed by Fishburn [10].

[3]The negative result by Ching and Zhou [8] uses Fishburn's extension but a much stronger notion of strategyproofness.

[4]This ensures that the majority relation is anti-symmetric and we can restrict our attention to tournament solutions.

relations, is then given by $\mathcal{R}^*(A)$. The typical element of $\mathcal{R}^*(A)$ will be $R = (R_1, \ldots, R_n)$. In accordance with conventional notation, we write $P_i$ for the strict part of $R_i$, i.e., $a\ P_i\ b$ if $a\ R_i\ b$ but not $b\ R_i\ a$. In a preference profile, the *weight* of an ordered pair of alternatives $w(a, b)$ is defined as the majority margin $|\{i \in N \mid a\ R_i\ b\}| - |\{i \in N \mid b\ R_i\ a\}|$.

Our central object of study are *social choice functions*, i.e., functions that map the individual preferences of the voters to a nonempty set of socially preferred alternatives.

**Definition 1.** A *social choice function (SCF)* is a function $f : \mathcal{R}^*(A) \to 2^A \setminus \emptyset$.

An SCF is *resolute* if $|f(R)| = 1$ for all $R \in \mathcal{R}^*(A)$, otherwise it is *irresolute*.

We restrict our attention to *majoritarian* SCFs (or tournament solutions), which are defined using the *majority relation*. The majority relation (or: *majority graph*) $R_M$ of a preference profile $R$ is the relation on $A \times A$ defined by

$$(a, b) \in R_M \text{ iff } w(a, b) \geq 0, \text{ for all alternatives } a, b \in A.$$

An SCF $f$ is said to be *majoritarian* if it is neutral[5] and its outcome only depends on the (unweighted) majority comparisons between pairs of alternatives, i.e., $f(R) = f(R')$ whenever $R_M = R'_M$.

We will now introduce the majoritarian SCFs considered in this paper (see [16, 5], for more information).

**Top Cycle** The top cycle rule *TC* (also known as *weak closure maximality*, *GETCHA*, or the *Smith set*) returns the maximal elements of $R_M$.

**Uncovered Set** Let $C$ denote the *covering relation* on $A \times A$, i.e., $a\ C\ b$ ("$a$ covers $b$") if and only if $a\ R_M\ b$ and $b\ R_M\ x$ implies $a\ R_M\ x$ for all $x \in A$. The *uncovered set UC* returns the maximal elements of $C$, i.e., those alternatives that are not covered by any other alternative.

**Bipartisan Set** Consider the symmetric two-player zero-sum game in which the set of actions for both players is given by $A$ and payoffs are defined as follows. If the first player chooses $a$ and the second player chooses $b$, the payoff for the first player is 1 if $a\ R_M\ b$, $-1$ if $b\ R_M\ a$, and 0 otherwise. The *bipartisan set BP* contains all alternatives that are played with positive probability in some Nash equilibrium of this game.

An SCF $f$ is called a *refinement* of another SCF $g$ if $f(R) \subseteq g(R)$ for all preference profiles $R$. In short, we write $f \subseteq g$ in this case. It can be shown for the above that $BP \subseteq UC \subseteq TC$ (see, e.g., Laslier [16]).

## 2.2 Strategyproofness

Even though our investigation of strategyproof SCFs is universal in the sense that it can be applied to any set extension, in this paper we will concentrate on two well-known set extensions due to Kelly [15] and Fishburn [10], respectively. They are defined as follows: Let $R_i$ be a preference relation over $A$ and $X, Y \subseteq A$ two nonempty subsets of $A$.

$\mathbf{X\ R_i^K\ Y}$ iff $x\ R_i\ y$ for all $x \in X$ and all $y \in Y$ [15]
One interpretation of this extension is that voters are completely unaware of the mechanism (e.g., a lottery) that will be used to pick the winning alternative [13].

[5] Neutrality postulates that for any permutation $\pi$ of the alternatives $A$ the choice function produces the "same" outcome (modulo the permutation). See also Section 3.1.1.

$\mathbf{X\ R_i^F\ Y}$ iff all of the following three conditions are satisfied [10]:

(*i*) $x\ R_i\ y$ for all $x \in X \setminus Y$ and $y \in X \cap Y$

(*ii*) $y\ R_i\ z$ for all $y \in X \cap Y$ and $z \in Y \setminus X$

(*iii*) $x\ R_i\ z$ for all $x \in X \setminus Y$ and $z \in Y \setminus X$

One interpretation of this extension is that the winning alternative is picked by a lottery according to some underlying *a priori* distribution and that voters are unaware of this distribution [8]. Alternatively, one may assume the existence of a chairman who breaks ties according to a linear, but unknown, preference relation.

It is easy to see that $X\ R_i^K\ Y$ implies $X\ R_i^F\ Y$ for any pair of sets $X, Y \subseteq A$ [13].

Based on these or any other set extension, we can now define different notions of strategyproofness for irresolute SCFs. Note that, in contrast to some related papers, we interpret preference extensions as fully specified (incomplete) preference relations rather than minimal conditions on set preferences.

Again, we write $P_i^{\mathcal{E}}$ for the asymmetric part of $R_i^{\mathcal{E}}$, for any set extension $\mathcal{E}$.

**Definition 2.** Let $\mathcal{E}$ be a set extension. An SCF $f$ is $P^{\mathcal{E}}$*-manipulable* by voter $i$ if there exist preference profiles $R$ and $R'$ with $R_j = R'_j$ for all $j \neq i$ such that

$$f(R')\ P_i^{\mathcal{E}}\ f(R),$$

i.e., $f(R')$ is $\mathcal{E}$-preferred to $f(R)$ by voter $i$.

An SCF is called $P^{\mathcal{E}}$*-strategyproof* if it is not $P^{\mathcal{E}}$-manipulable.

It follows from the observation on set extensions above that $P^F$-strategyproofness implies $P^K$-strategyproofness.

Of the above SCFs, *TC* has been shown to be $P^F$-strategyproof, *BP* is only $P^K$-strategyproof whereas *UC* fails to satisfy a variant of $P^K$-strategyproofness for weak preferences [6, 7].[6]

## 2.3 Strategyproofness with Tournaments

For reasons of efficiency we would like to omit references to preference profiles in our encodings and replace them by a more succinct representation of the same expressive power. As we will see, the notion of a *tournament* fulfills exactly this requirement:

A *tournament* is an asymmetric and complete binary relation on the set of alternatives $A$. Since, for an odd number of voters, tournaments correspond to majority graphs, we can alternatively view majoritarian SCFs as functions defined on tournaments rather than preference profiles, and write $f(T)$ instead of $f(R)$ with $T = R_M$ being the majority graph of the preference profile $R$.

For any two tournaments $T$ and $T'$ we denote the *edge difference* $T - T' := \{e \in T : e \notin T'\}$ by $\Delta_{T,T'}$.

**Definition 3.** A majoritarian SCF $f$ is said to be $P^{\mathcal{E}}$*-tournament-manipulable* if there exist tournaments $T, T'$ and a preference relation $R_\mu \supseteq \Delta_{T,T'}$ such that

$$f(T')\ P_\mu^{\mathcal{E}}\ f(T).$$

A majoritarian SCF is called $P^{\mathcal{E}}$*-tournament-strategyproof* if it is not $P^{\mathcal{E}}$-tournament-manipulable.

[6] Another natural and well-known set extension, called Gärdenfors' extension, leads to an even stronger notion of strategyproofness, which cannot be satisfied by any interesting majoritarian SCF [7]. Note that our negative result for Fishburn-strategyproofness trivially carries over to such stronger versions of strategyproofness.

As we show in the following theorem, for majoritarian SCFs it suffices to check this alternative definition of strategyproofness, and therefore any open problem involving strategyproofness for majoritarian SCFs can be reduced to an equivalent one involving tournament-strategyproofness only. In our case, this enables the efficient encoding described in the following section.

**Theorem 1.** *A majoritarian SCF is $P^{\mathcal{E}}$-strategyproof iff it is $P^{\mathcal{E}}$-tournament-strategyproof.*

*Proof.* We show that a majoritarian SCF is $P^{\mathcal{E}}$-manipulable iff it is $P^{\mathcal{E}}$-tournament-manipulable.

For the direction from left to right, let $f$ be a $P^{\mathcal{E}}$-manipulable majoritarian SCF. Then there exist preference profiles $R, R'$ and an integer $j$ with $R_i = R'_i$ for all $i \neq j$ such that $f(R')\ P_j^{\mathcal{E}}\ f(R)$. Define tournaments $T := R_M$ and $T' := R'_M$ as the majority graphs of $R$ and $R'$, respectively. Since $R$ and $R'$ only differ for voter $j$, it follows that $\Delta_{T,T'} \subseteq R_j$, i.e., all edges that are reversed from $T$ to $T'$ must have been in $R_j$. Thus, with $R_\mu := R_j$, we get that $f$ is tournament-manipulable.

For the converse, let $f$ be a $P^{\mathcal{E}}$-tournament-manipulable majoritarian SCF. It then admits a manipulation instance, i.e., there are two tournaments $T, T'$ and a preference relation $R_\mu \supseteq \Delta_{T,T'}$ such that $f(T')\ P_\mu^{\mathcal{E}}\ f(T)$. Using Debord's construction (see, e.g., [17]), we define a preference profile $R^- = (R_1, \ldots, R_{n-1})$ which has $T$ as its majority graph with weights

$$w_{R^-}(a,b) = \begin{cases} 0 & \text{if } (a,b) \in \Delta_{T,T'}, \\ 2 & \text{otherwise.} \end{cases}$$

Note that with this construction the number of voters $n-1$ so far is even. By adding $R_\mu$ as the $n$-th voter, we get to a profile $R := (R^-, R_\mu)$ with an odd number of voters as required. Then $w_R(a,b) = 1$ for all edges $(a,b) \in \Delta_{T,T'}$ and the weights of all other edges are greater than or equal to 1. The second profile $R'$ can be defined to contain the same first $n-1$ voters from $R$ and the reversed preference $\overline{R_\mu}$ as the $n$-th voter. The profile $R'$ then has $T'$ as its majority graph (since $w(b,a) = 1$ for all edges $(a,b) \in \Delta_{T,T'}$ and the weights of all other edges in the original tournament $T$ are at least 1 again), which completes the manipulation instance. I.e., we have found preference profiles $R, R'$ which only differ for voter $n$ (who has "truthful" preferences $R_\mu$) and for which it holds that $f(R') = f(T')\ P_\mu^{\mathcal{E}}\ f(T) = f(R)$. □

# 3 Methodology

The method applied here to solve open problems in social choice theory is similar to, and yet more powerful than the one presented in Tang and Lin [21] and Geist and Endriss [14]. Rather than translating the whole problem naïvely to SAT, a more evolved approach is used, which resolves a large degree of freedom already during the encoding of the problem. It is comparable to the way SMT (satisfiability modulo theories) solving works: at the core there is also a SAT solver; certain aspects of the problem, however, are dealt with in a separate theory solving unit which accepts a richer language and makes use of specific domain knowledge (Chapter 26, [3]). The general idea, however, remains: to encode the problem into a language suitable for SAT solving and to apply a SAT solver as a highly efficient, universal problem solving machine.

Using existing tools for higher-order formalizations directly rather than our specific approach, unfortunately, is not an option. For instance, a formalization of strategyproof majoritarian SCFs in higher-order logic as accepted by Nitpick [4] is straightforward, highly flexible, and well-readable, but only successful for proofs and counterexamples involving up

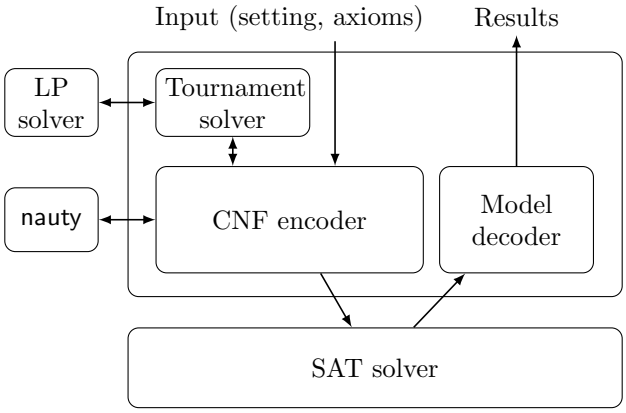


Figure 1: High-level system architecture

to 3 alternatives before the search space is exceeded.[7] An optimized formalization, which we derived together with the author of Nitpick (at the cost of reduced readability and flexibility), extends the performance to 4 alternatives, which is still below the requirements for our results.

In more detail, our approach is the following: for a given domain size $n$ we want to check whether there exists a majoritarian SCF $f$ which satisfies a set of properties (e.g., $P^F$-strategyproofness and Pareto-optimality).[8] For this specific domain size, we then encode the requirements of the majoritarian SCF $f$ as a propositional formula and let a SAT solver decide whether this formula has a satisfying assignment. If it has, we can translate the satisfying assignment back to a concrete instance of a majoritarian SCF $f$ which satisfies the required properties. If the formula is unsatisfiable, we know that no such function $f$ exists.

The high-level architecture of our implementation is depicted in Figure 1. The user selects the setting and the axioms, which are then encoded as a SAT instance. Depending on the problem, some preparatory tasks have to be solved before the actual encoding:

- sets, tournaments, and preference relations are enumerated,
- tournament isomorphisms are determined using the tool nauty [18], and
- choice sets for specific SCFs are computed (e.g., through matrix multiplication for $UC$ and linear programming for $BP$).

After the SAT solver has found a solution for the generated SAT instance, this solution is decoded back into a human-readable format.

In the following, we are going to describe in more detail how the general setting of majoritarian SCFs as well as their properties like strategyproofness can be encoded into CNF (conjunctive normal form). Firstly, we describe our initial encoding, which is expressive enough to encode all required properties, but allows for small domain sizes of (depending on the axioms) at most 4 to 5 alternatives only. Secondly, we explain how this encoding can be optimized to increase the overall performance by orders of magnitude such that larger instances of up to 7 alternatives are solvable.

[7] On the other hand, the strict formalization required for Nitpick helped identifying a formally inaccurate definition of Fishburn-strategyproofness by Gärdenfors [13] (which had later been repeated by several authors).

[8] Note that this usually is a substantially more complex task than checking certain properties for a *given* majoritarian SCF or simply computing the choice sets for a given function.

## 3.1 Initial Encoding

By design SAT solvers operate on propositional logic. A direct and naïve propositional encoding of the problem would, however, require a huge number of propositional variables since many higher-order concepts are involved (e.g., sets of alternatives, preference relations on sets and alternatives, and functions from tuples of such relations to sets). In our approach, only one type of variable is required, which we use to encode SCFs. The variables are of the form $c_{T,X}$ with $T$ being a tournament and $X$ being a set of alternatives.[9] The semantics of these variables are that $c_{T,X}$ if and only if $f(T) = X$, i.e., the majoritarian SCF $f$ selects the set of alternatives $X$ as the choice set for any preference profile with majority graph $T$. In total, this gives us a high but manageable number of $2^{\frac{m(m-1)}{2}} \cdot 2^m = 2^{\frac{m(m+1)}{2}}$ variables in the initial encoding.

The following two subsections demonstrate the initial encoding of both contextual as well as explicit axioms to CNF.

### 3.1.1 Context Axioms

Apart from the explicit axioms, which we are going to describe in the next subsection, there are further axioms that need to be considered in order to fully model the context of majoritarian SCFs. For this purpose, an arbitrary function that maps tournaments to non-empty sets of its vertices will be called a *choice function*. Using our initial encoding three axioms are introduced, which will ensure that functionality of the choice function and neutrality are respected: (1) functionality, (2) canonical isomorphism equality, and (3) the orbit condition.

The first axiom ensures that the relational encoding of $f$ by variables $c_{T,X}$ indeed models a function rather than an arbitrary relation, i.e., for each tournament $T$ there is exactly one set $X$ such that the variable $c_{T,X}$ is set to true. In formal terms this can be written as

$$\begin{aligned} & (\forall T)\,((\exists X)\ c_{T,X} \wedge (\forall Y, Z)\ Y \neq Z \rightarrow \neg(c_{T,Y} \wedge c_{T,Z})) \\ \equiv \quad & \bigwedge_{T} \left( \left( \bigvee_{X} c_{T,X} \right) \wedge \bigwedge_{Y \neq Z} (\neg c_{T,Y} \vee \neg c_{T,Z}) \right), \end{aligned} \tag{1}$$

which then translates to the pseudo code in Algorithm 1 for generating the CNF file.

```
foreach Tournament T do
    foreach Set X do
        variable(c(T, X));
    newClause;
    foreach Set Y do
        foreach Set Z ≠ Y do
            variable_not(c(T, Y));
            variable_not(c(T, Z));
            newClause;
```

**Algorithm 1:** Functionality of the choice function

In all algorithms, the subroutine $\mathrm{c}(T, X)$ takes care of the compact enumeration of variables.

[9] An encoding with variables $c_{T,x}$ for alternatives $x$ rather than sets would require less variable symbols. It, however, leads to much more complexity in the generated clauses, which more than offsets these savings.

The second and third axiom together constitute neutrality of the choice function $f$, which, formally, can be written as

$$\pi(f(T)) = f(\pi(T)) \quad \text{for all tournaments } T \text{ and permutations } \pi.$$

A direct encoding of this neutrality axiom, however, would be tedious (quantification over all permutations); in addition, our reformulation as canonical isomorphism equality and orbit condition enables a substantial optimization of the encoding as we will see in Section 3.2. In order to precisely state these two axioms we require some further observations:

We are going to use the well-known fact that isomorphisms define an equivalence relation on the set of all tournaments. For each equivalence class, pick a representative as the *canonical tournament* of this class. For any tournament $T$, we then have a unique canonical representation (denoted by $T_{\mathfrak{C}}$). Furthermore, we can pick one of the potentially many isomorphisms from $T_{\mathfrak{C}}$ to $T$ as the *canonical isomorphism* of $T$, and denote it by $\pi_T$.[10]

A choice function $f$ then satisfies *canonical isomorphism equality* if

$$f(T) = \pi_T(f(T_{\mathfrak{C}})) \text{ for all tournaments } T. \tag{2}$$

For the last of the three context axioms, the definition of an orbit should be clarified. The *orbits* of a tournament $T$ are the equivalence classes of alternatives, where two alternatives $a, b$ are considered equivalent iff there is an automorphism $\alpha : T \to T$ which maps $a$ to $b$, i.e., for which $\alpha(a) = b$. The set of orbits of a tournament $T$ is denoted by $\mathcal{O}_T$.

A choice function $f$ satisfies the *orbit condition* if

$$O \subseteq f(T_{\mathfrak{C}}) \;\text{ or }\; O \cap f(T_{\mathfrak{C}}) = \emptyset \tag{3}$$

for all canonical tournaments $T_{\mathfrak{C}}$ and their orbits $O \in \mathcal{O}_{T_{\mathfrak{C}}}$.

It can be shown that, for any choice function, neutrality is equivalent to the conjunction of the orbit condition and canonical isomorphism equality. The corresponding proof can be found in Appendix A.1.

### 3.1.2 Explicit Axioms

Many axioms can be efficiently encoded in our proposed encoding language. In this section we present the two main conditions that were required to achieve the results in Section 4. Clearly, the most important one is strategyproofness. In formal terms, $P^{\mathcal{E}}$-tournament-strategyproofness can be written as

$$\begin{aligned}&(\forall T, T', R_\mu \supseteq \Delta_{T,T'})\ \neg\left(f(T')\ P_\mu^{\mathcal{E}}\ f(T)\right)\\ \equiv &\bigwedge_{T}\bigwedge_{T'}\bigwedge_{R_\mu \supseteq \Delta_{T,T'}}\bigwedge_{Y P_\mu^{\mathcal{E}} X}(\neg c_{T,X} \vee \neg c_{T,Y})\end{aligned}$$

where $T, T'$ are tournaments, $R_\mu$ is a preference relation, and $X, Y$ are non-empty subsets of $A$. The algorithmic encoding of strategyproofness is omitted here since an optimized version is presented in Section 3.2.

Another property of SCFs that will play an important role in our results is the one of being a refinement of another SCF $g$. Fortunately, this can easily be encoded using our framework:

$$\begin{aligned}&(\forall T)(\exists X \subseteq g(T))\ f(T) = X\\ \equiv &\bigwedge_{T}\bigvee_{X \subseteq g(T)} c_{T,X}.\end{aligned}$$

[10] In practice, the tool nauty will automatically compute canonical representations for both graphs and isomorphisms.

Since the size of problems we could treat with this initial encoding remains limited, a selection of optimization techniques is described in the following section. These techniques enable a significant speedup of the computation and therefore make larger problem instances feasible.

## 3.2 Optimized Encoding for Improved Performance

In order to efficiently encode instances of more than four alternatives, we need to streamline our initial encoding without losing its universality or weakening it. In this section, we present the three optimization techniques we found most effective:

**Obvious redundancy elimination.** A straightforward first step is to reduce the obvious redundancy within the axioms. As an example consider the axiom of strategyproofness, where—in order to determine whether an outcome $Y = f(T')$ is preferred to an outcome $X = f(T)$—we consider *all* preference relations $R_\mu \supseteq \Delta_{T,T'}$. It suffices, however, if we stop after finding the first such preference relation with $Y \; P_\mu^{\mathcal{E}} \; X$, because then we already know that not both $Y = f(T')$ and $X = f(T)$ can be true.

Similarly, in many axioms, we can exclude considering symmetric pairs of objects (e.g., for functionality of the choice function, there is no need to consider both pairs of sets $(X, Y)$ and $(Y, X)$).

**Canonical tournaments.** The main efficiency gain can be achieved by making use of the canonical isomorphism equality (see Section 3.1.1) during encoding. Recall that this condition states that for any tournament $T$ the choice set $f(T)$ can be determined from the choice set $f(T_{\mathcal{C}})$ of the corresponding *canonical tournament* $T_{\mathcal{C}}$ by applying the respective canonical isomorphism $\pi_T$. Therefore, it suffices to formulate the axioms on a single representative of each equivalence class of tournaments, in our case, the canonical tournament. The magnitudes in Table 1 illustrate that this dramatically reduces the required number of variables, the size of the CNF formula and the time required for encoding it.

In particular, in all axioms we can replace any outer quantifier $\forall T$ by a quantifier $\forall T_{\mathcal{C}}$ that ranges over canonical tournaments only. In the case of strategyproofness, however, there is a second tournament $T'$ for which the restriction to canonical tournaments is potentially not strong enough anymore. We therefore keep it as an arbitrary tournament but make sure that we only need variable symbols $c_{T'_{\mathcal{C}},Y}$ for canonical tournaments in our CNF encoding. This can be achieved through the canonical isomorphism $\pi_{T'}$, since by Condition (2), $f(T') = Y$ if and only if $f(T'_{\mathcal{C}}) = \pi_{T'}^{-1}(Y)$.[11] The optimized encoding is shown in Algorithm 2.

Furthermore, since within the CNF formula we no longer make any statements about non-canonical tournaments, the canonical isomorphism equality condition becomes an "empty" condition and, thus, can be dropped from the encoding.

**Approximation through logically related properties.** Approximation is a standard tool in SAT/SMT which can speed up the solving process. For instance, over-approximation can help find unsatisfiable instances faster by only solving on parts of the full problem description in CNF. If then this partial CNF formula is found to be unsatisfiable, any superset will trivially be unsatisfiable, too. Since we are not aware of manipulation instances in the literature that require more than one edge in a tournament to be reversed, we, for instance, use over-approximation in the form of *single-edge tournament-strategyproofness*, a slightly weaker variant of strategyproofness with $|\Delta_{T,T'}| = 1$. If the solver returns that there is no single-edge tournament-strategyproof SCF that satisfies some properties $\Gamma$, we know immediately that there is also no strategyproof SCF that satisfies $\Gamma$.

[11] The inverse canonical isomorphisms are computed during preprocessing using nauty.

```
foreach Canonical Tournament T_C do
    foreach Tournament T' do
        Δ_{T_C,T'} ← T \ T';
        R_{Δ_{T_C,T'}} ← {R_μ | R_μ is a preference relation and R_μ ⊇ Δ_{T_C,T'}};
        foreach Set X do
            foreach Set Y do
                boolean found ← false;
                foreach R_μ ∈ R_{Δ_{T_C,T'}} do
                    p ← setExt(R_μ,E).prefers(Y,X);
                    if !found ∧ p then
                        variable_not(c(T_C, X));
                        variable_not(c(T'_C, π^{-1}_{T'}(Y)));
                        newClause;
                        found ← true;
```

**Algorithm 2:** $P^{\mathcal{E}}$-tournament-strategyproofness (optimized)

In a similar fashion we have applied various *logically simpler* conditions by Brandt and Brill [7] that are slightly stronger (or weaker, respectively) than $P^{\mathcal{E}}$-strategyproofness for specific set extensions $\mathcal{E}$ in order to *logically* over- or under-approximate problems and thus reduce encoding and solving time.

### 3.3 Finding Refinements through Incremental Solving

Generally, since the task of a SAT solver is to generate only one satisfying assignment, it does not necessarily output the most refined SCF that satisfies a given set of properties. Through iterated or incremental solving, however, we can force the SAT solver to generate finer and finer or simply different SCFs that satisfy a set of desired properties. [12] For refinements, this can be achieved by adding clauses which encode that the desired SCF must be (strictly) finer than previously found solution (see, e.g., the formulation in Section 3.1.2). When the most refined SCF with the desired properties has been found, adding these clauses leads to an unsatisfiable formula, which the SAT solver detects and therefore verifies the minimality of the solution.

With this final solving step, we have all the tools at hand which were required for our results, the main ones of which we describe in the next section.

## 4 Results and Discussion

Because of the universality of the proof method, many results can be generated in short time. Here we present our two main findings:

- There exists a refinement of $BP$ which is still $P^K$-strategyproof (Theorem 2).
- For majoritarian SCFs, $P^F$-strategyproofness and Pareto-optimality are incompatible for $m \geq 5$ (Theorem 3). For $m < 5$, $UC$ satisfies $P^F$-strategyproofness and Pareto-optimality.

[12] Finding a refinement of an SCF is *not* equivalent to finding a smaller/minimal model in the SAT sense; in our encoding all assignments have the same number of satisfied variables.

Where appropriate, further results are alluded to in the discussions proceeding the proofs. We start with our result on $P^K$-strategyproofness:

**Theorem 2.** *There exists a refinement of BP which is still $P^K$-strategyproof. As a consequence, BP is not the smallest majoritarian SCF satisfying $P^K$-strategyproofness.*

*Proof.* Within seconds our implementation finds a satisfying assignment for $m = 5$ and the encoding of the explicit axioms *refinement of BP* and *$P^K$-strategyproofness*. The corresponding majoritarian SCF can be decoded from the assignment and is defined like *BP* with the exception depicted in Figure 2. □

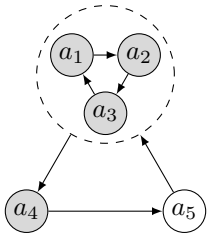


Figure 2: Tournament on which a $P^K$-strategyproof refinement of *BP* is possible. $\{a_1, a_2, a_3\}$ represents a component and of all its elements dominate $a_4$ and are dominated by $a_5$. While *BP* would choose the whole set $A$, the refined solution selects $\{a_1, a_2, a_3, a_4\}$ only.

Using the technique described in Section 3.3, we could furthermore confirm that this is the only refinement of *BP* on 5 alternatives which is still $P^K$-strategyproof. Note, however, that it does not satisfy the (strong, but natural) property of *composition-consistency* (see, e.g., [16]), which is one of the properties that one might want to address as future work (see Section 5).

In order to prove our main result on the incompatibility of Pareto-optimality and $P^F$-strategyproofness we first show the following important lemma, which establishes that, for majoritarian SCFs, the notion of Pareto-optimality is equivalent to being a refinement of the uncovered set ($UC$).

**Lemma 1.** *A majoritarian SCF $f$ is Pareto-optimal iff it is a refinement of $UC$.*

*Proof.* It is well-known (and was actually already observed by Fishburn [11]) that $UC$ is Pareto-optimal, from which it trivially follows that all its refinements are Pareto-optimal, too.

For the direction from left to right, let $f$ be a Pareto-optimal majoritarian SCF and $T$ an arbitrary tournament. It suffices to show that $f(T)$ can never contain a covered alternative (since then $f(T) \subseteq UC(T)$ contains uncovered alternatives only). So let $b$ be a covered alternative (say, it is covered by alternative $a$). We are going to construct a preference profile $R$ which has $T$ as its majority graph and in which $b$ is Pareto-dominated by $a$. Together with the Pareto-optimality of $f$ this implies that $b \notin f(T)$. We use a variant of the well-known construction by McGarvey, but for triples rather than pairs of alternatives. Note that for each voter we need to ensure that he strictly prefers $a$ to $b$ in order to obtain the desired Pareto-dominance of $a$ over $b$. Starting with an empty profile, for each alternative $x \notin \{a, b\}$ we add two voters $R_{x_1}, R_{x_2}$ to the profile. These two voters are defined depending on how $x$ is ranked relative to $a$ and $b$ in order to establish the edges between $a, x$ and $b, x$. Note that since $x\ T\ a$ implies $x\ T\ b$ (because of $a\ C\ b$), edge $(a, b)$ cannot be contained in a 3-cycle with $x$ and, thus, always forms a transitive triple with $x$.

- Case 1: $x\ T\ a$ (implies $x\ T\ b$)
  $R_{x_1}$ : $x, a, b, v_1, \ldots, v_{m-3}$
  $R_{x_2}$ : $v_{m-3}, \ldots, v_1, x, a, b$
- Case 2a: $a\ T\ x$ and $x\ T\ b$
  $R_{x_1}$ : $a, x, b, v_1, \ldots, v_{m-3}$
  $R_{x_2}$ : $v_{m-3}, \ldots, v_1, a, x, b$
- Case 2b: $a\ T\ x$ and $b\ T\ x$
  $R_{x_1}$ : $a, b, x, v_1, \ldots, v_{m-3}$
  $R_{x_2}$ : $v_{m-3}, \ldots, v_1, a, b, x$

Here $v_1, \ldots, v_{m-3}$ denotes an arbitrary enumeration of the $m-3$ alternatives in $X \setminus \{a, b, x\}$. The comma separated lists above are a shorthand notation in the sense that $R_i : v_1, v_2, v_3$ stands for the preference relation $v_1\ R_i\ v_2\ R_i\ v_3$.

In all cases, the two voters cancel out each other for all pairwise comparisons other than $(a, b)$, $(x, a)$ and $(x, b)$. For each of the remaining edges $(y, z) \in T$ (with $\{y, z\} \cap \{a, b\} = \emptyset$) we further add two voters (now even closer to the construction by McGarvey)

$$\begin{aligned} R_{(y,z)_1} &: \quad y, z, a, b, v_1, \ldots, v_{m-4} \quad \text{and} \\ R_{(y,z)_2} &: \quad v_{m-4}, \ldots, v_1, a, b, y, z, \end{aligned}$$

which together establish edge $(y, z)$, reinforce $(a, b)$ and cancel otherwise. Note that in order to achieve an odd number of voters, an arbitrary voter can be added without changing the majority relation (as all edges had a weight of at least 2 so far). This completes the construction of a preference profile $R$ which has $T$ as its majority graph and in which $b$ is Pareto-dominated by $a$. □

To establish the full result (which does not admit a proof by counterexample, like for Theorem 2) we—similarly to previous approaches—make use of an inductive argument.

**Lemma 2.** *If there exists a majoritarian SCF $f$ for $m + 1$ alternatives that is $P^{\mathcal{E}}$-strategyproof and is a refinement of UC, then there also exists a majoritarian SCF $f'$ for just $m$ alternatives that satisfies these two properties.*

*Proof.* Let $f \subseteq UC$ be a majoritarian SCF for $m + 1 \geq 2$ alternatives that is $P^{\mathcal{E}}$-strategyproof. Then we define $f_e$ to be the restriction of $f$ to $m$ alternatives based on tournaments in which alternative $e$ is a Condorcet loser, i.e., an alternative that is dominated by all other alternatives. In formal terms, define

$$f_e(T) := f(T^{+e}),$$

where $T^{+e}$ is the tournament obtained from $T$ by adding an alternative $e$ as a Condorcet loser (i.e., adding it below all previous $m$ alternatives). This restriction of $f$ is a well-defined choice function since alternative $e$ cannot be contained in $f(T^{+e}) \subseteq UC(T^{+e}) = UC(T)$, where the last equation follows from the fact that $UC$ is composition-consistent, or, alternatively, by observing that the covering relation is unaffected by deleting Condorcet losers.

We now need to show that for some alternative $e$ the restriction $f_e$ is a majoritarian SCF that is $P^{\mathcal{E}}$-strategyproof and a refinement of $UC$. Since this will hold for any $e \in X$, we just pick one $e$ arbitrarily.

**Majoritarian:** The fact that $f_e$ is a majoritarian SCF carries over trivially from $f$.

**$P^{\mathcal{E}}$-strategyproofness:** Assume for a contradiction that $f_e$ is not $P^{\mathcal{E}}$-strategyproof. Then there exist tournaments $T$ and $T'$ on $m$ alternatives such that $f_e(T')\ P_\mu^{\mathcal{E}}\ f_e(T)$ with $R_\mu \supseteq \Delta_{T,T'}$. But since $f_e(T) = f(T^{+e})$ and $f_e(T') = f(T'^{+e})$, we get

$$f(T'^{+e}) = f_e(T')\ P_\mu^{\mathcal{E}}\ f_e(T) = f(T^{+e}),$$

which contradicts $P^{\mathcal{E}}$-strategyproofness of $f$ (as the two tournaments $T'^{+e}$ and $T^{+e}$ form a manipulation instance).

**Refinement of $UC$:** Let $T$ be an arbitrary tournament on $m$ alternatives and consider the following chain of set inclusions, which proves that $f_e \subseteq UC$:

$$f_e(T) = f(T^{+e}) \subseteq UC(T^{+e}) = UC(T).$$

□

Note that the proofs of the individual properties within the inductive proof above do only rely on the definition of $f_e$ and stand independently of each other. Furthermore, it may be noted that Lemma 2 can even be shown for refinements of arbitrary SCFs $g$ whose choice set $g(T)$ does not shrink when all Condorcet losers are removed from $T$.

Finally, we are now in the position to state and prove this paper's main result regarding $P^F$-strategyproofness.

**Theorem 3.** *For any number of alternatives $m \geq 5$ there is no majoritarian SCF $f$ that satisfies $P^F$-strategyproofness and Pareto-optimality.*

*Proof.* By Lemma 1 we can replace Pareto-optimality by the property of being a refinement of $UC$. With this in mind, we inductively prove the statement.

The base case of $m = 5$ alternatives was verified using our computer-aided approach, i.e., we checked that there is no satisfying assignment for an encoding of $P^F$-strategyproofness and being a refinement of $UC$ with $|A| = 5$ alternatives, which the SAT solver confirmed within seconds.

For the induction step, we apply the contrapositive of Lemma 2 with $\mathcal{E} := F$, which yields the desired results. □

Note that this result does not form a contradiction to the fact that $TC$ is $P^F$-strategyproof, as, for $m \geq 4$ alternatives, $TC$ is strictly coarser than $UC$ and therefore not Pareto-optimal. Possibly, $TC$ is even the smallest majoritarian SCF that satisfies $P^F$-strategyproofness. We were able to verify this for up to 7 alternatives using our computer program, with the exception of 4 alternatives, where $UC$ is a strict refinement of $TC$ and (as our method shows) still $P^F$-strategyproof.[13]

# 5 Future Work

Based on the ease of adaptation of our proposed method, we anticipate many insights to spring from this approach in the future. Apart from simply applying our system to further investigate strategyproofness, we have identified three streams of future research that could arise from our contribution:

**Transfer to general SCFs** For reasons of reduced complexity, here we have studied *majoritarian* SCFs only. The framework, however, is applicable in the same way to general SCFs, which "operate" on full preference profiles (rather than majority graphs). The

[13] It is not obvious whether an inductive argument can extend these verified instances to larger numbers of alternatives (as, for instance, such an induction step would require at least 5 alternatives).

challenge then is to find a suitable representation of such preference profiles and potentially corresponding inductive arguments on the number of voters.

**Apply to other properties of SCFs** Some preliminary experiments suggest that our technique can easily be applied to a range of properties other than strategyproofness which deserve further investigation. In many cases it suffices to just formalize and implement the additional axioms. Of particular interest could be such properties that link the behavior of SCFs for different domain sizes, e.g., *composition-consistency*.

**Generalize inductive argument** It appears reasonable to investigate whether the inductive argument of Lemma 2 can be generalized to a whole class of properties/axioms, ideally based on their logical form. As in the work of Geist and Endriss [14], this could then (together with the previous item) enable an automated search for further theorems about SCFs.

## Acknowledgments

This material is based upon work supported by Deutsche Forschungsgemeinschaft under grants BR 2312/7-2 and BR 2312/9-1. The authors thank Jasmin Christian Blanchette, Markus Brill, and Hans Georg Seedig for helpful discussions and their support.

## References

1. S. Barberà. Manipulation of social decision functions. *Journal of Economic Theory*, 15 (2):266–278, 1977.

2. S. Barberà, W. Bossert, and P. K. Pattanaik. Ranking sets of objects. In S. Barberà, P. J. Hammond, and C. Seidl, editors, *Handbook of Utility Theory*, volume II, chapter 17, pages 893–977. Kluwer Academic Publishers, 2004.

3. A. Biere, M. Heule, H. van Maaren, and T. Walsh, editors. *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2009.

4. J. C. Blanchette and T. Nipkow. Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In *Proceedings of the First International Conference on Interactive Theorem Proving*, pages 131–146. Springer, 2010.

5. F. Brandt. *Tournament Solutions – Extensions of Maximality and Their Applications to Decision-Making*. Habilitation Thesis, Faculty for Mathematics, Computer Science, and Statistics, University of Munich, 2009.

6. F. Brandt. Group-strategyproof irresolute social choice functions. In T. Walsh, editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 79–84. AAAI Press, 2011.

7. F. Brandt and M. Brill. Necessary and sufficient conditions for the strategyproofness of irresolute social choice functions. In K. Apt, editor, *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 136–142. ACM Press, 2011.

8. S. Ching and L. Zhou. Multi-valued strategy-proof social choice rules. *Social Choice and Welfare*, 19:569–580, 2002.

9. A. Feldman. Manipulation and the Pareto rule. *Journal of Economic Theory*, 21: 473–482, 1979.

10. P. C. Fishburn. Even-chance lotteries in social choice theory. *Theory and Decision*, 3: 18–40, 1972.

11. P. C. Fishburn. Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977.

12. P. Gärdenfors. Manipulation of social choice functions. *Journal of Economic Theory*, 13(2):217–228, 1976.

13. P. Gärdenfors. On definitions of manipulation of social choice functions. In J. J. Laffont, editor, *Aggregation and Revelation of Preferences*. North-Holland, 1979.

14. 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.

15. J. S. Kelly. Strategy-proofness and social choice functions without single-valuedness. *Econometrica*, 45(2):439–446, 1977.

16. J.-F. Laslier. *Tournament Solutions and Majority Voting*. Springer-Verlag, 1997.

17. M. Le Breton. On the uniqueness of equilibrium in symmetric two-player zero-sum games with integer payoffs. *Économie publique*, 17(2):187–195, 2005.

18. B. D. McKay and A. Piperno. Practical graph isomorphism, II. *Journal of Symbolic Computation*, 2013.

19. K. Nehring. Monotonicity implies generalized strategy-proofness for correspondences. *Social Choice and Welfare*, 17(2):367–375, 2000.

20. M. R. Sanver and W. S. Zwicker. Monotonicity properties and their adaption to irresolute social choice rules. *Social Choice and Welfare*, 39(2–3):371–398, 2012.

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

22. A. D. Taylor. *Social Choice and the Mathematics of Manipulation*. Cambridge University Press, 2005.

Felix Brandt
Institut für Informatik
Technische Universität München
Munich, Germany
Email: `brandtf@in.tum.de`

Christian Geist
Institut für Informatik
Technische Universität München
Munich, Germany
Email: `geist@in.tum.de`

# Appendix

# A Omitted Proofs

## A.1 Equivalence of neutrality and the conjunction of Conditions (2) and (3)

The equivalence of neutrality and the conjunction of Conditions (2) and (3) is not obvious, why we give the proofs here. Let's first remind ourselves of the two conditions.

**Condition (2)** A choice function $f$ satisfies *canonical isomorphism equality* if

$$f(T) = \pi_T(f(T_{\mathfrak{C}})) \text{ for all tournaments } T.$$

**Condition (3)** A choice function $f$ satisfies the *orbit condition* if

$$O \subseteq f(T_{\mathfrak{C}}) \text{ or } O \cap f(T_{\mathfrak{C}}) = \emptyset$$

for all canonical tournaments $T_{\mathfrak{C}}$ and their orbits $O \in \mathcal{O}_{T_{\mathfrak{C}}}$.

We start by showing that the orbit condition is equivalent to a statement about automorphisms:

**Lemma 3.** *Let $f$ be choice function. Then the following statement is equivalent to the orbit condition:*

$$\alpha(f(T_{\mathfrak{C}})) = f(T_{\mathfrak{C}}) \text{ for all canonical tournaments } T_{\mathfrak{C}} \text{ and their automorphisms } \alpha. \quad (4)$$

*Proof.* Let $f$ be a choice function and $T_{\mathfrak{C}}$ a canonical tournament. For the direction from left to right, let furthermore $O \in \mathcal{O}_{\mathcal{T}_{\mathfrak{C}}}$ an orbit on $T_{\mathfrak{C}}$. Now pick two alternatives $a, b \in O$. We show that either both alternatives are chosen by $f$ or none is. Since $a$ and $b$ are in the same orbit there must be an automorphism $\alpha$ on $T_{\mathfrak{C}}$ for which $\alpha(a) = b$. Observe that $a \in f(T_{\mathfrak{C}})$ iff $b \in \alpha(f(T_{\mathfrak{C}}))$ iff $b \in f(T_{\mathfrak{C}})$, where the last step is an application of Condition (4).

For the converse, let $\alpha$ be an automorphism on $T_{\mathfrak{C}}$, pick an arbitrary alternative $a \in A$ and consider its inverse-image $\alpha^{-1}(a) =: b$. Since $a$ and $b$ are in the same orbit it holds by the orbit condition that $a \in f(T_{\mathfrak{C}})$ iff $b \in f(T_{\mathfrak{C}})$. Furthermore, as $\alpha(b) = a$ we get that $a \in f(T_{\mathfrak{C}})$ iff $a \in \alpha(f(T_{\mathfrak{C}}))$. Thus, $f(T_{\mathfrak{C}}) = \alpha(f(T_{\mathfrak{C}}))$, which is what we wanted to prove. □

Next one can prove a general statement about how to split any isomorphism into a canonical one and an automorphism.

**Lemma 4.** *Any isomorphism $\pi : T_{\mathfrak{C}} \to T$ can be decomposed into the canonical isomorphism $\pi_T$ and an automorphism $\alpha : T_{\mathfrak{C}} \to T_{\mathfrak{C}}$. I.e., for any isomorphism $\pi : T_{\mathfrak{C}} \to T$ there is an automorphism $\alpha : T_{\mathfrak{C}} \to T_{\mathfrak{C}}$ such that $\pi = \pi_T \circ \alpha$.*

*Proof.* Define $\alpha : T_{\mathfrak{C}} \to T_{\mathfrak{C}}$ by setting $\alpha := \pi_T^{-1} \circ \pi$. Since inverse functions and compositions of isomorphisms are themselves isomorphisms, it follows directly that $\alpha$ is an automorphism. Furthermore, $\pi_T \circ \alpha = \pi_T \circ \left(\pi_T^{-1} \circ \pi\right) = \left(\pi_T \circ \pi_T^{-1}\right) \circ \pi = \pi$. □

These two lemmata together can finally be used to prove the desired equivalence:

**Lemma 5.** *For any choice function, neutrality is equivalent to the conjunction of the orbit condition and canonical isomorphism equality.*

*Proof.* Let $f$ be a choice function and first note that by Lemma 3 we might use Condition 4 rather than the orbit condition. Therefore, the direction from left to right is trivially true.

For the direction from right to left, we first only show that (2) and (4) imply neutrality *for canonical tournaments*: So let $T_{\mathfrak{C}}$ be a canonical tournament, $\pi$ a permutation and define $T' := \pi(T_{\mathfrak{C}})$. By Lemma 4, we can decompose the isomorphism $\pi : T_{\mathfrak{C}} \to T'$ such that $\pi = \pi'_T \circ \alpha$ for some automorphism $\alpha$ on $T_{\mathfrak{C}}$. Then the following chain of equations holds, which proves the claim for canonical tournaments:

$$f(\pi(T_{\mathfrak{C}})) = f(T') \overset{(2)}{=} \pi_{T'}(f(T'_{\mathfrak{C}})) = \pi_{T'}(f(T_{\mathfrak{C}})) \overset{(4)}{=} \pi_{T'}(\alpha(f(T_{\mathfrak{C}}))) = \pi(f(T_{\mathfrak{C}})).$$

For arbitrary tournaments $T$ and permutations $\pi$ we write

$$\begin{aligned} f(\pi(T)) \quad &= \quad f(\pi(\pi_T(T_{\mathfrak{C}}))) = f((\pi \circ \pi_T)(T_{\mathfrak{C}})) \\ &\overset{\text{(canonical)}}{=} \quad (\pi \circ \pi_T)(f(T_{\mathfrak{C}}))) = \pi(\pi_T(f(T_{\mathfrak{C}}))) \overset{(2)}{=} \pi(f(T)), \end{aligned}$$

which proves the claim. □