Computer-aided proofs of Arrow's and other impossibility theorems ☆

Pingzhong Tang *, Fangzhen Lin

Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

ABSTRACT

Arrow's impossibility theorem is one of the landmark results in social choice theory. Over the years since the theorem was proved in 1950, quite a few alternative proofs have been put forward. In this paper, we propose yet another alternative proof of the theorem. The basic idea is to use induction to reduce the theorem to the base case with 3 alternatives and 2 agents and then use computers to verify the base case. This turns out to be an effective approach for proving other impossibility theorems such as Muller–Satterthwaite and Sen's theorems as well. Motivated by the insights of the proof, we discover a new theorem with the help of computer programs. We believe this new proof opens an exciting prospect of using computers to discover similar impossibility or even possibility results.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Recently, there has been much interest and work in applying economics models such as those from social choice theory to computer science problems as well as computational techniques to solving problems in social choice theory. In this paper, we consider a different application of AI to economics: using computers to help prove and discover theorems in social choice theory.

The particular theorems that we are interested in are the impossibility theorems such as those by Arrow [3], Sen [20], and Muller and Satterthwaite [15] in social choice theory [2], an area concerning about how individual preferences can be aggregated to form a collective preference in a society. Social choice theory has sometimes been called "a science of the impossible" because of the many famous impossibility theorems that have been proved in it. Among them, Arrow's theorem [3] on the non-existence of rational social welfare function is without doubt the most famous one. It shows the non-existence of the collective social preference (called social welfare function) even when some minimal standards such as Pareto efficiency and non-dictatorship are imposed. Arrow's original proof of this result is relatively complex, and over the years, quite a few alternative proofs have been advanced (see e.g. [4,7,8,21]).

In this paper, we propose yet another alternative proof of this result, with the help of computers. Briefly, Arrow's theorem says that in a society with at least three possible outcomes (alternatives) for each agent, it is impossible to have a social welfare function that satisfies the following three conditions: unanimity (Pareto efficiency), independent of irrelevant alternatives (IIA), and non-dictatorship. We shall show by induction that this result holds if and only if it holds for the base case when there are exactly two agents and three alternatives (the single agent case is trivial). For the base case, we verify it using computers in two ways. One views the problem as a constraint satisfaction problem (CSP), and uses a depth-first search algorithm to generate all social welfare functions that satisfy the first two conditions, and then verifies that all of

---

them are dictatorial. The other translates these conditions to a logical theory and uses a SAT solver to verify that the resulting logical theory is not satisfiable. Either way, it took less than one second on an AMD Opteron-based server (with 4 1.8 GHz CPUs and 8 GB RAM) for the base case to be verified.

As it turns out, this strategy works not just for proving Arrow's theorem. The same inductive proof can be adapted almost directly for proving other impossibility results such as Sen's and Muller–Satterthwaite theorems. We have used it to prove Gibbard–Satterthwaite theorem [9,19] as well, but we leave its proof to another paper.

As a byproduct of our proof of Arrow's theorem, the social welfare functions that satisfies IIA only in the base case can all be generated by our program. To our surprise, the number of such functions is so small that we are able to look at them one by one. By doing so, we form an interesting conjecture and then prove it using the same techniques as in the previous proofs. We then demonstrate the powerfulness of the newly proved theorem by showing that it subsumes both Arrow's and Wilson's theorems.

These proofs suggest that many of the impossibility results in social choice theory are all rooted in some small base cases. Thus an interesting thing to do is to use computers to explore these small base cases to try to come up with new conjectures automatically, and to understand the boundary between impossibility and possibility results. This is what we think the long term implication of our new proofs of Arrow's and other impossibility theorems lies, and the main reason why we want to formulate the conditions in these theorems in a logical language and use a SAT solver to check their consistency.

The rest of the paper is organized as follows. In Sections 2 and 3, we review Arrow's theorem and then describe our new inductive proof of this result. We then describe in Sections 4 and 5 how this proof can be adapted to prove Muller–Satterthwaite theorem and Sen's theorem. In Section 6, We describe the idea of making use of computer programs to discover new theorems in social choice theory and a theorem discovered this way. We then propose in Section 7 a logical language for social choice theory and describe how it can be used to axiomatize Arrow's theorem and how the base case in our inductive proof of Arrow's theorem can be checked using a SAT solver. Finally, we conclude this paper with a summary of our results and an outlook on future research.

2. Arrow's theorem

A voting model is a tuple (N, O), where N is a finite set of individuals (agents) and O a finite set of outcomes (alternatives). An agent's preference ordering is a linear ordering of O, and a preference profile $>$ of (N, O) is a tuple $(>_1, \ldots, >_n)$, where $>_i$ is agent i's preference ordering, and n the size of N. In the following, when N is clear from the context, we also call $>$ a preference profile of O. Similarly, when O is clear from the context, we also call it a preference profile of N.

Definition 1. Given a voting model (N, O), a *social welfare function* is a function $W : L^n \to L$, where L is the set of linear ordering of O, and n the size of N.

A social welfare function defines a social ordering for each preference profile. If we consider the social ordering given by a social welfare function as the aggregates of the preference orderings of the individuals in the society, it is natural to impose some conditions on it. For instance, it should not be dictatorial in that the aggregated societal preference ordering always is the same as a particular individual's preference. Arrow showed that a seemingly minimal set of such conditions turns out to be inconsistent.

In the following, given a preference profile $> = (>_1, \ldots, >_n)$, we sometimes write $>_W$ for $W(>)$. Thus both $a >_W b$ and $a\ W(>)\ b$ mean the same thing: the alternative a is preferred over the alternative b according to the societal preference ordering $W(>)$.

Definition 2. A social welfare function W is *unanimous* (Pareto efficient) if for all alternatives a_1 and a_2, we have that if $a_1 >_i a_2$ for every agent i, then $a_1 >_W a_2$

In words, if everyone ranks alternative a_1 above a_2, then a_1 must be ranked above a_2 socially.

Definition 3. A social welfare function W is *independent of irrelevant alternatives* (IIA) if for all alternatives a_1 and a_2, and all preference profiles $>'$ and $>''$, we have that $\forall i\ a_1 >'_i a_2$ iff $a_1 >''_i a_2$ implies that $a_1 >'_W a_2$ iff $a_1 >''_W a_2$.

Literally, IIA means that the relative social ordering of two alternatives depends only on their relative orderings given by each agent and has nothing to do with other alternatives.

Definition 4. An agent i is a dictator in a social welfare function W if for all alternatives a_1 and a_2, $a_1 >_W a_2$ iff $a_1 >_i a_2$. If there is a dictator in W, then it is said to be dictatorial. Otherwise, W is said to be non-dictatorial.

It is easy to see that if there are at least two alternatives, then there can be at most one dictator in any social welfare function.

Theorem 1 *(Arrow's theorem [3]). For any voting model (N, O), if $|O| \geqslant 3$, then any social welfare function that is unanimous and IIA is also dictatorial.*

Arrow's original proof of this result is somewhat complicated, and there are several alternative proofs by others, e.g. [4,7,8]. We now give yet another one using induction.

3. An inductive proof of Arrow's theorem

For ease of presentation, we assume the following notations.

- For any set S, we use S_{-a} to denote $S \setminus \{a\}$, i.e. the result of deleting a in S.
- We extend the above notation to tuples as well: if $t = (t_1, \ldots, t_n)$, then we use t_{-i} to denote the tuple $(t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_n)$. Furthermore, we use (t_{-i}, s) to denote the result of replacing ith item in t by s: $(t_{-i}, s) = (t_1, \ldots, t_{i-1}, s, t_{i+1}, \ldots, t_n)$. We use $t_{-\{i,j\}}$ to denote $(t_{-i})_{-j}$.
- If $>$ is a linear ordering of O, and $a \in O$, then we let $>_{-a}$ be the restriction of $>$ on O_{-a}: for any $a', a'' \in O_{-a}$, $a' >_{-a} a''$ iff $a' > a''$. On the other hand, if $>$ is a linear ordering of O_{-a} for some $a \in O$, then we let $>^{+a}$ be the extension of $>$ to O such that for any $a' \in O_{-a}$, $a' >^{+a} a$. Similarly, we let $>^{a+}$ to be the extension of $>$ to O such that for any $a' \in O_{-a}$, $a >^{a+} a'$. Thus if $>$ is a linear ordering of O, and $a \in O$, then $>_{-a}^{+a}$ is $(>_{-a})^{+a}$, i.e. the result of moving a to the bottom of the ordering. These notations extend to tuples of orderings. Thus if $>$ is a preference profile of (N, O_{-a}), then

$$>^{+a} = (>_1, \ldots, >_n)^{+a} = (>_1^{+a}, \ldots, >_n^{+a}),$$

which will be a preference profile of (N, O). Similarly for $>^{a+}$.

Like any inductive proof, there are two cases for our proof, the inductive case and the base case.

3.1. The inductive case

Lemma 1. *If there is a social welfare function for n individuals and $m + 1$ alternatives that is unanimous, IIA and non-dictatorial, then there is a social welfare function for n individuals and m alternatives that satisfies these three conditions as well, for all $n \geqslant 2, m \geqslant 3$.*

Proof. Let $N = \{1, \ldots, n\}$ be a set of n agents, O a set of $m + 1$ alternatives, and W a social welfare function for (N, O) that satisfies the three conditions in the lemma. We show that there is an $a \in O$ such that the "restriction" of W on O_{-a} also satisfies these three conditions.

For any $a \in O$, we define the restriction of W on O_{-a}, written W_a, to be the following function: for any preference profile $> = (>_1, \ldots, >_n)$ of O_{-a}, $W_a(>) = W(>^{+a})_{-a}$. In other words, $W_a(>)$ is the result of applying W to the preference profile $>^{+a}$ of O, and then projecting it on O_{-a}. The key property of this welfare function is that for any a' and a'' in O_{-a}, and any preference profile $>$ of O_{-a}, $a' W_a(>) a''$ iff $a' W(>^{+a}) a''$.

We show that W_a is unanimous and IIA:

- Suppose $a', a'' \in O_{-a}$ and $a' >_i a''$ for all i. By our definition $a' >_i^{+a} a''$ for all i as well. Since W is unanimous, $a' W(>^{+a}) a''$. Thus $a' W_a(>) a''$. This shows that W_a is unanimous.
- Let $a', a'' \in O_{-a}$ and $>', >''$ be two preference profiles of O_{-a} such that $\forall i$ $a' >_i' a''$ iff $a' >_i'' a''$. Thus $\forall i$ $a' >_i'^{+a} a''$ iff $a' >_i''^{+a} a''$ as well. Since W is IIA, $a' W(>'^{+a}) a''$ iff $a' W(>''^{+a}) a''$. Hence $a' W_a(>') a''$ iff $a' W_a(>'') a''$. This shows that W_a is also IIA.

We now show that there is an $a \in O$ such that W_a is not dictatorial. First for any $a \in O$ and any $a', a'' \in O_{-a}$, and any profile $>$ of O, we have

$$a' >_W a'' \quad \text{iff} \quad a' W(>_{-a}^{+a}) a''. \tag{1}$$

This follows because W is IIA and $a', a'' \in O_{-a}$.

Now let b be any alternative in O. Suppose W_b has a dictator, say agent 1 in it. Since W is not dictatorial, there must be a preference profile $>$ of O and some $c, d \in O$ such that $c >_1 d$ but $d >_W c$. Since $|O| = m + 1 > 3$, we can find an alternative $e \in O_{-b} \setminus \{c, d\}$. We now show that W_e is not dictatorial. Suppose otherwise. There are two cases:

- Agent 1 is again the dictator in W_e. Then $W_e(>_{-e})$ and $>_1$ agree on c and d. Thus $c W_e(>_{-e}) d$. By our definition of W_e, this means that $c [W(>_{-e}^{+e})]_{-e} d$. Since $c, d \in O_{-e}$, this means that $c W(>_{-e}^{+e}) d$. By (1), we have $c >_W d$, a contradiction with our assumption that $d >_W c$.

- Another agent, say agent 2 is the dictator in W_e. Let $a_1 \neq a_2$ be any two alternatives in $O \setminus \{b, e\}$. This is possible since $|O| > 3$. Let $>'$ be a preference profile of O such that $a_1 >'_1 a_2$ but $a_2 >'_2 a_1$. From $a_1 >'_1 a_2$, $\{a_1, a_2\} \subseteq O_{-b}$, and that agent 1 is the dictator in W_b, we can conclude $a_1 >'_W a_2$ as we have done in the previous case. Similarly, from $a_2 >'_2 a_1$, $\{a, a_2\} \subseteq O_{-e}$, and that agent 2 is the dictator in W_e, we can conclude $a_2 >'_W a_1$, a contradiction.

Thus we have shown that W_e cannot have a dictator. $\quad\square$

Note that it is essential for our proof that $m \geqslant 3$. Notice also that we only use the assumptions that W is IIA and non-dictatorial in our proof that W_a is not dictatorial for some $a \in O$. The assumption that W is unanimous is used only in showing that W_a is also unanimous.

Lemma 2. *If there is a social welfare function for $n + 1$ individuals and m alternatives that is unanimous, IIA and non-dictatorial, there will also be a social welfare function for n individual and m alternatives that satisfies these three conditions as well, for all $n \geqslant 2$, $m \geqslant 3$.*

Proof. Let $N = \{1, \ldots, n, n + 1\}$ be a set of agents, and O a set of m alternatives, and W a social welfare function for (N, O) that satisfies the three conditions in the lemma. For any $i \neq j \in N$, we define $W_{i,j}$ to be the following social welfare function for (N_{-i}, O): for any preference profile $>$ of (N, O), $W_{i,j}(>_{-i}) = W(>_{-i}, >_j)$, where $(>_{-i}, >_j)$, as we defined earlier, is the result of replacing $>_i$ in $>$ by $>_j$. Thus the social welfare function $W_{i,j}$ is defined through W by making agent i and agent j always agreeing with each other. Clearly, for any i, j, $W_{i,j}$ is unanimous and IIA because W satisfies these two conditions. We now show that we can find two distinct agents i and j such that $W_{i,j}$ is not dictatorial. Suppose otherwise, for every pair $i > j \in N$, $W_{i,j}$ is dictatorial. Now consider three distinct agents $i_1 < i_2 < i_3$ in N. This is possible because $|N| = n + 1 \geqslant 3$. Suppose i is the dictator in W_{i_1, i_2}, j the dictator in W_{i_1, i_3}, and k the dictator in W_{i_2, i_3}. There are two cases:

- Case 1: $i = j = k$. Since W is not dictatorial, there is a profile $>$ of (N, O) and two alternatives a_1 and a_2 such that $>_W$ and $>_i$ disagree on a_1 and a_2, say $a_1 >_i a_2$ but $a_2 >_W a_1$. Now at least two players from $\{i_1, i_2, i_3\}$ must agree on a_1, a_2. Let these two players be j_1 and j_2, and without loss of generality, suppose $j_1 < j_2$. Now consider the profile $(>_{-j_1}, >_{j_2})$. Since W is IIA, and because $>_{j_1}$ and $>_{j_2}$ agree on a_1 and a_2, $>_W$ and $W(>_{-j_1}, >_{j_2})$ must agree on a_1 and a_2. So $a_2 \; W(>_{-j_1}, >_{j_2}) \; a_1$. But i is the dictator in W_{j_1, j_2}, $W_{j_1, j_2}(>_{-j_1})$ must agree with $>_i$. Since $W_{j_1, j_2}(>_{-j_1})$ is defined to be $W(>_{-j_1}, >_{j_2})$, thus $W(>_{-j_1}, >_{j_2})$ agrees with $>_i$, so $a_1 \; W(>_{-j_1}, >_{j_2}) \; a_2$, a contradiction.
- Case 2: $i \neq j$ or $i \neq k$ or $j \neq k$. First, by our definition of $W_{x,y}$, and our assumption that agents i, j, and k are dictators in W_{i_1, i_2}, W_{i_1, i_3}, and W_{i_2, i_3}, respectively, for any preference profile $>$ of (N, O), if $>_{i_1} = >_{i_2} = >_{i_3}$, then $>_W = >_i$, $>_W = >_j$, and $>_W = >_k$. Since two of $\{i, j, k\}$ must be distinct, this means that $\{i, j, k\} \subseteq \{i_1, i_2, i_3\}$. Since i must be in N_{-i_1}, so $i \neq i_1$, thus $i \in \{i_2, i_3\}$. Similarly, $j \in \{i_2, i_3\}$ and $k \in \{i_1, i_3\}$. This leads to eight possible combinations for i, j, and k. Each of them will lead to a contradiction, using the following table:

(i, j, k)	$>_{i_1}$	$>_{i_2}$	$>_{i_3}$
(i_2, i_2, i_1)	$c > a > b$	$a > b > c$	$a > c > b$
(i_2, i_2, i_3)	case 1		
(i_2, i_3, i_1)	case 1		
(i_2, i_3, i_3)	$c > a > b$	$b > c > a$	$a > b > c$
(i_3, i_2, i_1)	$c > a > b$	$a > b > c$	$b > c > a$
(i_3, i_2, i_3)	$b > a > c$	$b > c > a$	$a > b > c$
(i_3, i_3, i_1)	$b > c > a$	$b > a > c$	$a > b > c$
(i_3, i_3, i_3)	case 1		

Each row in the above table either gives a preference profile that will lead to a contradiction or point to "case 1", meaning a contradiction can be derived similar to case 1. For instance, consider the row $(i, j, k) = (i_2, i_3, i_1)$, which says "case 1". This case can be reduced to "case 1" as follows. Since $i = i_2$ is the dictator in W_{i_1, i_2}, i_1 is the dictator in W_{i_2, i_1}. Similarly, i_1 is the dictator in W_{i_3, i_1} because $i_3 = k$ is the dictator in W_{i_1, i_3}. Thus i_1 is the dictator in W_{i_2, i_1}, W_{i_3, i_1}, and W_{i_2, i_3}, and the same reasoning in case 1 will lead to a contradiction here.

Now consider the first row $(i, j, k) = (i_2, i_2, i_1)$, and the preference profile $>$ given in the row:

$$c >_{i_1} a >_{i_1} b, \qquad a >_{i_2} b >_{i_2} c, \qquad a >_{i_3} c >_{i_3} b.$$

Because $i_2 = j$ is the dictator in W_{i_1, i_3}, $W(>_{-i_1}, >_{i_3}) = >_{i_2}$. But $>_{i_1}$ and $>_{i_3}$ agree on b and c, thus by IIA:

$$b >_W c \quad \text{iff} \quad b \; W(>_{-i_1}, >_{i_3}) \; c \quad \text{iff} \quad b >_{i_2} c.$$

So

$$b >_W c. \tag{2}$$

Similarly, $>_{i_1}$ and $>_{i_2}$ agree on a and b, and i_2 is the dictator in W_{i_1,i_2}, thus $a >_W b$ iff $a >_{i_2} b$. So

$$a >_W b. \tag{3}$$

Now $>_{i_2}$ and $>_{i_3}$ agree on a and c, and i_1 is the dictator in W_{i_2,i_3}, thus $a >_W c$ iff $a >_{i_1} c$. So $c >_W a$, which contradicts with (2) and (3). The other cases are similar.

This means that there must be some $i \neq j \in N$ such that $W_{i,j}$ is not dictatorial. $\quad\square$

Again notice that it is essential for our proof that $|N| = n + 1 \geqslant 3$, and that the existence of a non-dictatorial $W_{i,j}$ depends only on the assumptions that W is IIA and non-dictatorial.

By these two lemmas, we see that Arrow's theorem holds iff it holds for the case when there are exactly two agents and three possible outcomes.[1]

3.2. The base case

We now turn to the proof of the base case, and as we mentioned earlier, we use computer programs to do that.

The base case says that when $|N| = 2$ and $|O| = 3$, there is no social welfare function on (N, O) that is unanimous, IIA, and non-dictatorial. A straightforward way of verifying this is to generate all possible social welfare functions in (N, O) and check all of them one by one for these three conditions. However, there are too many such functions for this to be feasible on current computers: there are $3! = 6$ number of linear orderings of O, resulting in $6 \times 6 = 36$ total number of preference profiles of (N, O), and 6^{36} possible social welfare functions.

Thus one should not attempt to explicitly generate all possible social welfare functions. What we did instead is to generate explicitly all social welfare functions that satisfy the conditions of unanimity and IIA, and then check if any of them is non-dictatorial.

We treat the problem of generating all social welfare functions that satisfy the conditions of unanimity and IIA as a constraint satisfaction problem (CSP). A CSP is a triple (V, D, C), where V is a set of variables, and D a set of domains, one for each variable in V, and C a set of constraints on V (see, e.g. [18]). An assignment of the CSP is a function that maps each variable in V to a value in its domain. A solution to the CSP is an assignment that satisfies all constraints in C.

Now consider the voting model $(\{1, 2\}, \{a, b, c\})$ in our base case. We define a CSP for it by introducing 36 variables $x_1, \ldots, x_{36}$, one for each preference profile of the voting model. The domain of these variables is the set of 6 linear orderings of $\{a, b, c\}$, and the constraints are the instantiations of the unanimity and IIA conditions on the voting model. As can be easily seen, there is a one-to-one correspondence between the social welfare functions of the voting model and the assignments of the CSP. Furthermore, a solution to the CSP corresponds to a social welfare function that satisfies the unanimity and IIA conditions, and vice versa.

To solve this CSP, we use a depth-first search that backtracks whenever the current partial assignment violates the constraints, and implemented it in SWI-Prolog. As we mentioned earlier, when run on our AMD server machine, our Prolog program returned in less than one second two solutions, one corresponds to the social welfare function where agent 1 is the dictator, and the other agent 2 the dictator.

This verifies the base case of our inductive proof of Arrow's theorem, thus completes our proof. As mentioned in the introduction, we also verified the base case using a SAT solver. This requires a logical language to encode postulates in social choice theory, and will be described in a separate section below.

At last, it is worth noting that Suzumura [21] also provided, in his presidential address to the Japanese Economic Association, a specific backwards induction proof that reduces Arrow's theorem to two agents (but n alternatives) base case and then proves the base case with almost the same amount of efforts as the inductive case. In contrast, our further reduction to three alternatives case makes the computational verification possible and as we will show, our reduction to two agents case is more general and can be used to prove other impossibility theorems.

4. Muller–Satterthwaite theorem

As mentioned before, the same strategy that we used for proving Arrow's theorem can be used to prove other impossibility theorems. In fact, we have modified the above proof for proving Sen's and Muller–Satterthwaite theorems. We prove in the following Muller–Satterthwaite theorem (cf. e.g. [12]).

Arrow's theorem is about the social welfare function which maps a preference profile to a preference ordering. In comparison, Muller–Satterthwaite theorem concerns about *social choice function* which maps a preference profile to an outcome which is supposed to be the "winner" of the election (as represented by the preference profile).

Definition 5. Given a voting model (N, O), a social choice function is a function $C : L^n \to O$, where L is the set of linear orders on O, and n the number of agents in N.

---

[1] Technically speaking, we also need to consider the case when $|N| = 1$, but this is a trivial case.

Instead of the conditions of unanimity, IIA, and non-dictatorship in Arrow's theorem, Muller and Satterthwaite considered the following three corresponding conditions.

Definition 6. A social choice function C is *weakly unanimous* if for every preference profile $>$, if there is a pair of alternatives a_1, a_2 such that $a_1 >_i a_2$ for every agent i, then $C(>) \neq a_2$.

Thus according to this condition, an alternative that is dominated by another should never be selected.

Definition 7. A social choice function C is *monotonic* if, for every preference profile $>$ such that $C(>) = a$, if $>'$ is another profile such that $a >'_i a'$ whenever $a >_i a'$ for every agent i and every alternative a', then $C(>') = a$ as well.

In words, monotonicity means that if a choice function selects an outcome for a preference profile, then it will also select this outcome for any other preference profile that does not decrease the ranking of this outcome.

Definition 8. An agent i is a *dictator* in a social choice function C if C always selects i's top choice: for every preference profile $>$, $C(>) = a$ iff for all $a' \in O$ that is different from a, $a >_i a'$. C is *non-dictatorial* if it has no dictator.

Theorem 2 (*Muller–Satterthwaite theorem* [15]). *For any voting model* (N, O) *such that* $|O| \geqslant 3$, *any social choice function that is weakly unanimous and monotonic is also dictatorial.*

Like our proof of Arrow's theorem, we prove this theorem by induction. The inductive step is again by two lemmas similar to the ones for Arrow's theorem.

Lemma 3. *If there is a social choice function for n individuals and $m + 1$ alternatives that is weakly unanimous, monotonic and non-dictatorial, then there is also a social choice function for n individuals and m alternatives that satisfies these three conditions, for all $n \geqslant 2, m \geqslant 3$.*

Proof. Let (N, O) be a voting model such that $|N| = n$ and $|O| = m + 1$, and C a social choice function that satisfies the three conditions in the lemma. Just like our proof of the corresponding Lemma 1, for any $a \in O$, we define C_a to be a social choice function that is the "restriction" of C on O_{-a}: for any preference profile $>$ of O_{-a}, $C_a(>) = C(>^{+a})$. Again it can be easily seen that for any $a \in O$, C_a is weakly unanimous and monotonic. Now we show that there is one such a such that C_a is non-dictatorial.

Suppose otherwise: for any a, C_a is dictatorial. We start by assuming C_b has a dictator i. Since C is non-dictatorial, we can find a profile $> \in O$ such that $C(>) = c \neq d$, where d is top ranked outcome according to $>_i$. Since there are $|m + 1| \geqslant 4$ outcomes, we can find another outcome e that is distinct from b, c, d. Now we consider C_e, there are two cases:

- C_e still has agent i as its dictator. We have $d = C_e((>)_{-e}) = C((>)^{+e}_{-e})$, but according to monotonicity, we have $C((>)^{+e}_{-e}) = C(>) = c$, which leads to a contradiction since $c \neq d$.
- C_e has a dictator $j \neq i$. For any preference profile $>' \in O$ such that f is ranked top according to $>'_i$, g is ranked top according to $>'_j$, $f \neq g$ and f, g are distinct from b, e, we consider the following two preference profiles $>'' = ((>)^{+b}_{-b})^{+e}_{-e}$ and $>''' = ((>)^{+e}_{-e})^{+b}_{-b}$. Clearly, we have $C(>'') = g$ and $C(>''') = f$. However, according to monotonicity, we have $C(>'') = C(>''')$. This leads to a contradiction.

Therefore, C_e cannot have a dictator. So we have prove that there is always a outcome a so that C_a is non-dictatorial. □

Lemma 4. *If there is a social choice function for $n + 1$ individuals and m alternatives that is weakly unanimous, monotonic and non-dictatorial, then there is also a social choice function for n individuals and m alternatives that satisfies these three conditions, for all $n \geqslant 2, m \geqslant 3$.*

Proof. Let (N, O) be a voting model such that $|N| = n + 1$ and $|O| = m$, and C a social choice function that satisfies the three conditions in the lemma. Just like our proof of Lemma 2, for any pair of agents $i \neq j \in N$, we define $C_{i,j}$ to be the following social welfare function for (N_{-i}, O): for any preference profile $>$ of (N, O), $C_{i,j}(>_{-i}) = C(>_{-i}, >_j)$. Again it can be easily seen that for any pair of agents $i \neq j$, $C_{i,j}$ is weakly unanimous and monotonic.

We prove in the following that we can find two distinct agents i, j such that $C_{i,j}$ is non-dictatorial. Suppose not, then for every pair of agents i, j, there is an agent $d_{i,j}$ that is a dictator of $C_{i,j}$. We first show that $d_{i,j} = j$ for any i, j. Suppose otherwise, $d_{i,j} = k \neq j$. Since C is non-dictatorial, we can find a profile $>$ such that $a = C(>) \neq b$ where b is on top of $>_k$. We then still have $C(>_{-\{i,j\}}, (>_j)^{a+}_{-a}, (>_j)^{a+}_{-a}) = a$ according to monotonicity of C. But according to the dictatorship of $C_{i,j}$, we have $C(>_{-\{i,j\}}, (>_j)^{a+}_{-a}, (>_j)^{a+}_{-a}) = b$, a contradiction. Therefore, we have $d_{i,j} = j$ for any i, j.

Now consider a profile any $>$ on $(N + 1, O)$, any triple of agents i, j, k and any triple of alternatives (this is possible since $|N + 1| \geqslant 3$, $|O| \geqslant 3$) where

- $a >_i c >_i b >_i \ldots$ for $>_i$
- $c >_j b >_j a >_j \ldots$ for $>_j$
- $b >_k a >_k c >_k \ldots$ for $>_k$

Notice that $>_i, >_j, >_k$ only differ in $\{a, b, c\}$. There are the following cases:

1. $C(>) = a$, then we change $>_j$ to $>_k$ and denote the new profile $>'$. By monotonicity, we still have $C(>') = a$. This leads to the contradiction that $d_{j,k} \neq k$.
2. Other cases where $C(>) = b, c$ or other alternatives are similar to the case above.

Therefore, we conclude that there are two distinct agents i, j such that $C_{i,j}$ is non-dictatorial. $\quad\square$

For the base case again notice that the case for $N = 1$ is trivial, thus we need only to consider the case when there are two agents and three alternatives. Again the number of all possible social choice functions is too large to enumerate explicitly, but both our methods for verifying the base case in Arrow's theorem can be adapted here. For the depth-first search method, our program similarly reported that there are exactly two social choice functions that are weakly unanimous and monotonic, and both of them are dictatorial.

One additional interesting thing to note is that it is also extremely fast to generate all the social choice functions that satisfy monotonicity only. There are 17 functions returned in total: 2 are dictatorships, 3 are constant and the remaining 12 are all functions whose ranges contain 2 elements. Since a generalization of Muller–Satterthwaite theorem [16] says that the condition weak unanimity can be weakened by only requiring that the range contains at least 3 elements, these 12 functions are the only interesting ones to look at when one wants to completely generalize the monotonicity condition.

Notice that our proof outlined above parallels our earlier proof of Arrow's theorem but does not make use of Arrow's theorem. In contrast, the existing proofs such as those in [12,15,16] are more complicated and [12,15] rely on Arrow's theorem.

5. Sen's theorem

We show in the following that our proof can also be copied to prove the impossibility theorem by Sen [20].

Definition 9. A *collective choice rule* is a functional relationship $F : L^n \to R$ that specifies one and only one social preference relation r for any preference profile.

The set R of preference relations includes all the possible binary relations. Particularly, the members of R are not necessarily transitive or complete. However, Sen focused only on *social decision functions*, a subset of collective choice rules with certain restriction on R.

Definition 10. A *social decision function* is a collective choice rule $C : L^n \to R$ such that for each $r \in R$, r should generate a choice function.

A preference relation r should generate a "choice function" if according to r, there exists a best alternative in every subset of alternatives. In other words, there exists an alternative that is at least as preferred as any other alternative in that subset.

Sen then suggested three conditions which should be satisfied by any rational social decision function, namely unrestricted domain (condition U), unanimity (condition P, named after Pareto principle) and liberalism (condition L).

The first two conditions are mentioned explicitly or implicitly in Arrow's framework: unrestricted domain says that all the possible preference profiles should be included in the domain of a social decision function while unanimity is exactly the same one as in Arrow's theorem.

The third condition, liberalism, is somewhat debatable. The intuitive justification behind is that each individual has the freedom to determine at least one social choice. For example, I should feel free to have my own garden planted lily rather than rose.

Definition 11 (*Liberalism*). For each individual i, there is at least one pair of alternatives, say (a_1, a_2), such that this individual is decisive for (a_1, a_2).[2]

---

[2] i is decisive for (a_1, a_2) if i prefers a_1 to a_2 implies that a_1 is preferred to a_2 according to the social preference relation returned by the decision function.

Theorem 3 *(Sen's theorem [20]). There is no social decision function that can simultaneously satisfy U, P and L.*

Sen further weakened the condition L to be the following form L^*,

Definition 12 *(Liberalism*). There are at least two individuals such that for each of them there is at least one pair of alternatives over which he is decisive.

In other words, condition L^* only guarantees the freedom for two individuals instead of everyone in the society, as required by condition L. The following theorem subsumes Theorem 3.

Theorem 4 *(Sen's theorem [20]). There is no social decision function that can simultaneously satisfy conditions U, P, and L*, for any voting model with $|N| \geqslant 2$ and $|O| \geqslant 3$.*

We prove in the following Theorem 4. The inductive step consists of the following two lemmas

Lemma 5. *If there is a social decision function for $m + 1$ alternatives and n outcomes that satisfies U, P and L*, then there is a social decision function for m outcomes and n individuals that satisfies these three conditions as well, for all $m \geqslant 4$.*

Proof. Let (N, O) be a voting model such that $|N| = n$ and $|O| = m + 1$, and C a social decision function that satisfies the three conditions in the lemma. For any $a \in O$, we define C_a to be a function that is the "restriction" of C on O_{-a}: for any preference profile $>$ of O_{-a}, $C_a(>) = C(>^{+a})_{-a}$.

- C_a is still a social decision function. Since C is a social decision function, so the range of C is the set of preferences that can generate a choice function. That is, for any subset of outcomes, there is a best outcome. This outcome will still be the best after we restrict on C_a since a is less preferred than any other outcome by unanimity.
- The property of U and P of C_a follows directly from that of C.
- Since C satisfies L^*, we can always find two individuals and their decisive pairs (a_1, a_2) and (a_3, a_4) respectively. Since $|m + 1| \geqslant 5$, we can find an element a_5 that is not in $\{a_1, a_2, a_3, a_4\}$. Now we can see that C_{a_5} still satisfies L^* because the two decisive individuals are still decisive for their pairs of alternatives (a_1, a_2) and (a_3, a_4). $\quad\square$

Lemma 6. *If there is a social decision function for m alternatives and $n + 1$ outcomes that satisfies U, P and L*, then there is a social decision function for m outcomes and n individuals that satisfies these three conditions as well, for all $n \geqslant 2$.*

Proof. By the property L^* of C, we have two individuals j, k that are decisive for their own pair of outcomes. We can also find another distinct agent i, since there are at least $2 + 1$ three individuals for C. We now define $C_{i,j}$ to be the following social welfare function for (N_{-i}, O): for any preference profile $>$ of (N, O), $C_{i,j}(>_{-i}) = C(>_{-i}, >_j)$. Then $C_{i,j}$ is still a social decision function and all the three properties follows directly from that of C. $\quad\square$

Notice that we have $m \geqslant 4$ in Lemma 5, so the base case for Sen's theorem is $|N| = 2$ and $|O| = 3, 4$. We can still check it by our depth-first search algorithm, which we do not want to repeat here.

6. Discovering new theorems

We have been advocating a methodology of theorem discovering using computers [10,11]. The basic idea is to look for conjectures that are true in small domains using computers. Once we find such a conjecture, we then hope it to be true in general. In the following, we present a new theorem discovered this way.

6.1. An observation in small domain

Recall that in our CSP formulation of the base case of Arrow's theorem, constraints are the instantiations of both IIA and Unanimity conditions. Using the same algorithm, we can generate all the functions that satisfies IIA by restricting the constraints to be the instantiations of IIA only.

To our surprise (not so surprised if one is familiar with Wilson's theorem [22], which will be introduced later in this section), among the total 6^{36} social welfare functions, they are only 94 of them satisfying IIA. This seems to suggest that the impossibility in Arrow's theorem is actually not caused by the conflict between unanimity and IIA but mostly by IIA, which is too strong for a social welfare function to satisfy.

Among these 94 functions, 2 of them are dictatorial, 2 of them are inversely dictatorial which means the social order of the function is always opposite to someone's individual order, and each of the remaining 90 functions has at most two values in the range. Of course among the 90 functions, there are 6 constant functions, each of which has exactly one value. Moreover, for any of the remaining 84 functions which have two different values, the distance between these values is at

most one pair of outcomes. For example, if one value is $a_1 >_W a_2 >_W a_3$, then the other value can only be $a_2 >_W a_1 >_W a_3$ or $a_1 >_W a_3 >_W a_2$.

Definition 13. An agent i is a *inverse dictator* in W if for all alternatives a_1 and a_2, $a_1 >_W a_2$ iff $a_2 >_i a_1$. If there is a inverse dictator in W, then it is said to be *inversely dictatorial*.

Definition 14. The *(Kendall tau) distance* of two orderings on O is the number of pairs of outcomes where two orderings disagree.

When IIA holds for a function W, we can define from it a social welfare function $W_Y : L_Y^n \to L_Y$, the restriction of W on an arbitrary non-empty subset Y of O, where L_Y is the restriction of L on Y and for any profile $>' \in L_Y^n$, $W_Y(>') = W(>)_Y$, for any $> \in L^n$ such that $>_Y = >'$.

We then generalize the above observation in small domain into the following theorem.

Theorem 5. *If a social welfare function W on (N, O) satisfies IIA, then for every subset Y of O such that $|Y| = 3$,*

1. *W_Y is dictatorial, or*
2. *W_Y is inversely dictatorial, or*
3. *The range of W_Y has at most 2 elements, whose the distance is at most 1.*

Notice that 1–3 are pairwise disjoint. Fortunately, by observation we have already proved the base case for Theorem 5.

Lemma 7. *If a social welfare function W on (N, O) where $|N| = 2$, $|O| = 3$ satisfies IIA,*

1. *W is dictatorial, or*
2. *W is inversely dictatorial, or*
3. *The range of W has at most 2 elements, whose distance is at most 1.*

We show in the following the inductive step hold for this theorem too.

6.2. The inductive step

We first prove the following lemma that translates dictatorship to unanimity and translates inverse dictatorship to inverse unanimity under IIA.

Definition 15. A social welfare function W is *inversely unanimous* (inversely Pareto efficient) if for all alternatives a_1 and a_2, we have that if $a_1 >_i a_2$ for all agent i, then $a_2 >_W a_1$

Lemma 8. *If a social welfare function W on (N, O) where $|O| \geqslant 3$ satisfies IIA, then*

1. *W is dictatorial iff W is unanimous;*
2. *W is inversely dictatorial iff W is inversely unanimous.*

Proof. Assuming IIA,

1. if W is unanimous, by Arrow's theorem, it is dictatorial; if W is dictatorial, by the definition of unanimity, it is unanimous.
2. Now if W is inversely dictatorial, but W is not inversely unanimous, we can construct a new function W' such that $a >_W b$ iff $b >_{W'} a$ for any (a, b) and any preference profile $>$. We can see that W' satisfies IIA and dictatorial, but not unanimity. This contradicts to what we have proved above; similarly, if W is inversely unanimous but not inversely dictatorial, we can construct the same W' that satisfies IIA and unanimity but not dictatorial, violating Arrow's theorem. $\square$

By the following lemma, together with Lemma 8, we can extend Lemma 7 to voting models with any number of agents.

Lemma 9. *If there is a social welfare function for $n + 1$ individuals and 3 outcomes that is IIA, but not unanimous or inversely unanimous and its range has two elements whose distance is at least 2, then there is a social welfare function for n individuals and 3 outcomes that is IIA, but not unanimous or inversely unanimous and its range has two elements whose distance is at least 2 as well.*

Proof. Let $N = \{1, \ldots, n, n+1\}$ be a set of agents, and $O = \{a, b, c\}$ a set of 3 alternatives, and W a social welfare function for (N, O) that satisfies the four conditions in the lemma. The same as before, for any $i \neq j \in N$, we define $W_{i,j}$ to be the following social welfare function for (N_{-i}, O): for any preference profile $>$ of (N, O), $W_{i,j}(>_{-i}) = W(>_{-i}, >_j)$, where $(>_{-i}, >_j)$ is the result of replacing $>_i$ in $>$ by $>_j$. Clearly, for any i, j, $W_{i,j}$ is IIA, not unanimous or inversely unanimous because W satisfies these three conditions. We now show that we can find two distinct agent i, j such that the range of $W_{i,j}$ has two elements whose distance is at least 2.

Since there exist two preference profiles $>$ and $>'$ such that $W(>)$ differs from $W(>')$ in at least two pair of outcomes, say (a, b) and (a, c). Since W is IIA, its restrictions $W_{\{a,b\}}$ and $W_{\{a,c\}}$ are well defined. Now we consider $W_{\{a,b\}}(a > b, \ldots, a > b)$, $W_{\{a,b\}}(b > a, \ldots, b > a)$, $W_{\{a,c\}}(a > c, \ldots, a > c)$ and $W_{\{a,c\}}(c > a, \ldots, c > a)$. There are four cases as follows:

1. $W_{\{a,b\}}(a > b, \ldots, a > b) \neq W_{\{a,b\}}(b > a, \ldots, b > a)$ and $W_{\{a,c\}}(a > c, \ldots, a > c) \neq W_{\{a,c\}}(c > a, \ldots, c > a)$. Without loss of generality, we suppose $W(>)$ agrees with $W_{\{a,b\}}(a > b, \ldots, a > b)$ in (a, b) and agrees with $W_{\{a,c\}}(a > c, \ldots, a > c)$ in (a, c), therefore $W(>')$ agrees with $W_{\{a,b\}}(b > a, \ldots, b > a)$ in (a, b) and agrees with $W_{\{a,c\}}(c > a, \ldots, c > a)$ in (a, c). Now we consider a profile $>'' = (a > b > c, \ldots, a > b > c)$, clearly $W(>)$ agrees with $W(>'')$ in $(a, b), (a, c)$; similarly, for $>''' = (c > b > a, \ldots, c > b > a)$, $W(>')$ agrees with $W(>''')$ in $(a, b), (a, c)$. So $W(>'')$ and $W(>''')$ differ in $(a, b), (a, c)$. For two profiles $>''_{-i}, >'''_{-i}$, their values of $W_{i,j}$ differ in $(a, b), (a, c)$ for any j.

2. $W_{\{a,b\}}(a > b, \ldots, a > b) = W_{\{a,b\}}(b > a, \ldots, b > a)$ and $W_{\{a,c\}}(a > c, \ldots, a > c) \neq W_{\{a,c\}}(c > a, \ldots, c > a)$. Without loss of generality, we suppose $W(>)$ agrees with $W_{\{a,b\}}(a > b, \ldots, a > b)$ in (a, b) and agrees with $W_{\{a,c\}}(a > c, \ldots, a > c)$ in (a, c), therefore $W(>')$ agrees with $W(c > a, \ldots, c > a)$ in (a, c). Now we consider a profile $>'' = (a > b > c, \ldots, a > b > c)$, clearly $W(>)$ agrees with $W(>'')$ in $(a, b), (a, c)$; we can also construct another profile $>'''$ such that $>'''$ has $c > a$ for each agent and $>'''$ with $>'$ on (a, b) for each agent. So $W(>'')$ and $W(>''')$ differ in $(a, b), (a, c)$. Now we look at the relation of (a, b) in $>'''$, since there are at least 3 agents, we can always find two agents, say i, j that agree on (a, b). For profiles $>''_{-i}, >'''_{-i}$, their values of $W_{i,j}$ differ in $(a, b), (a, c)$.

3. $W_{\{a,b\}}(a > b, \ldots, a > b) \neq W_{\{a,b\}}(b > a, \ldots, b > a)$ and $W_{\{a,c\}}(a > c, \ldots, a > c) = W_{\{a,c\}}(c > a, \ldots, c > a)$. This case is similar to case 2 above.

4. $W_{\{a,b\}}(a > b, \ldots, a > b) = W_{\{a,b\}}(b > a, \ldots, b > a)$ and $W_{\{a,c\}}(a > c, \ldots, a > c) = W_{\{a,c\}}(c > a, \ldots, c > a)$. In this case, we first show that there exist two profiles $>^1$ and $>^2$ such that $W(>^1)$ and $W(>^2)$ differ in (b, c). We construct $>^1$ in such a way that $>^1$ agrees with either $>$ or $>'$ in (a, b) for each player so that $b >^1_W a$ and $>^1$ agrees with either $>$ or $>'$ in (a, c) for each player so that $a >^1_W c$. Therefore, we have $b >^1_W a >^1_W c$. Similarly, we can construct $>^2$ so that $c >^2_W a >^2_W b$. In this way, $W(>^1)$ and $W(>^2)$ differ in $(a, b), (b, c), (a, c)$. Now we consider further $W_{\{b,c\}}(c > b, \ldots, c > b)$ and $W_{\{b,c\}}(b > c, \ldots, b > c)$. There are two cases:
   - $W_{\{b,c\}}(c > b, \ldots, c > b) \neq W_{\{b,c\}}(b > c, \ldots, b > c)$, then this case will still be case 2 by considering $(a, b), (b, c)$ instead.
   - $W_{\{b,c\}}(c > b, \ldots, c > b) = W_{\{b,c\}}(b > c, \ldots, b > c)$. We prove in the following that this case is impossible. Suppose $W(a > b > c, \ldots, a > b > c) = o_1 > o_2 > o_3$ where (o_1, o_2, o_3) is a permutation of (a, b, c). Now we construct a new profile $>^*$ where o_2 is always on top of each agent's preference and $>^*$ agrees with either $>^1$ or $>^2$ in (o_1, o_3) for each player so that $o_3 W(>^*) o_1$. But since o_2 is always on top such that we have $o_1 W(>^*) o_2$ and $o_2 W(>^*) o_3$ because $W_{\{a,b\}}(a > b, \ldots, a > b) = W_{\{a,b\}}(b > a, \ldots, b > a)$, $W_{\{a,c\}}(a > c, \ldots, a > c) = W_{\{a,c\}}(c > a, \ldots, c > a)$ and $W_{\{b,c\}}(c > b, \ldots, c > b) = W_{\{b,c\}}(b > c, \ldots, b > c)$. By transitivity, we have $o_1 W(>^*) o_3$, which is a contradiction. $\square$

Since W is IIA, its restriction on any non-empty subset Y of $|O|$ is still IIA. Therefore our Theorem 5 follows from Lemmas 7, 8 and 9. Note that Theorem 5 is closely related to several existing results (see e.g. [22] Theorem 5, aka Wilson's partition lemma).

6.3. The implication of the new theorem

We show in the following how to use Theorem 5 to prove two existing theorems.

6.3.1. A brief proof of Arrow's theorem
One immediate implication of Theorem 5 is Arrow's theorem. Given that a social welfare function W on (O, N) is IIA, by applying Theorem 5, we know W_Y is either of the three cases when $Y \subseteq |O|, |Y| = 3$. If W is further unanimous assumed by Arrow's theorem, so is W_Y. Clearly W_Y can only be case 1 for any Y. In other words, The restriction of W on any three-element subset is dictatorial. Now we arbitrarily choose such a $Y = \{a_1, a_2, a_3\}$, suppose the dictator in W_Y is i. Then i will still be a dictator in W_{Y^1}, where $Y^1 = \{a_1, a_2, a_4\}$ for any $a_4 \in O \setminus Y$, since there can only be one agent that is decisive for the pair (a_1, a_2). Similarly, i is still the dictator for W_{Y^2}, where $Y^2 = \{a_1, a_3, a_4\}$ or $\{a_2, a_3, a_4\}$, $\{a_1, a_4, a_5\}$ or $\{a_4, a_5, a_6\}$ for any distinct a_5, a_6. Therefore, we prove that all the restrictions of W on three-elements subset have a common dictator i. Since i is decisive for any pair in O^2, i is a dictator in W.

6.3.2. A brief proof of Wilson's theorem
There have been fruitful researches on relaxing the unanimity condition in Arrow's framework. In other words, these researches also aim at finding the implication of IIA condition. One of the most famous one is Wilson's theorem [22]. It states

that even with a condition called nonimposition that is much weaker than unanimity, IIA can already imply dictatorship or inverse dictatorship.

Definition 16. A social welfare function W is *nonimposition* if for all distinct alternatives a_1 and a_2, there exists a preference profile $>$ such that $a_1 >_W a_2$

Theorem 6 (*Wilson's theorem [22]*). *For any voting model* (N, O), *if* $|O| \geqslant 3$, *then any social welfare function that satisfies nonimposition and IIA is either dictatorial or inversely dictatorial.*

Theorem 5 also implies Wilson' theorem as well. Given that a social welfare function W on (O, N) is IIA, by applying Theorem 5, we know W_Y is either of the three cases when $Y \subseteq |O|$, $|Y| = 3$. If W is further nonimposition assumed by Wilson's theorem, so is W_Y. Therefore W_Y can only be dictatorial or inversely dictatorial since case 3 in Theorem 5 obviously violates nonimposition. Dictatorship or inverse dictatorship then follows from similar arguments to those of Arrow's theorem above.

7. A logical language for social choice theory

As we mentioned earlier, we are not only interested in alternative proofs of existing theorems or even the manual discovery of new theorem like what we did in Section 5. Our long term goal is to automate the discovery of theorems in social choice theory, game theory, and others [10,11]. One insight of our new proofs is that these known impossibility results are all rooted in some small base cases. Thus by experimenting with other conditions in small cases, we could discover some new results. To fully automate the enumeration and verification process of these conditions, we propose a logical language for social choice theory.[3]

This language is a variant of the situation calculus [13,17], one of the best known languages in AI. For representing Arrow's theorem, we use two predicates: $p(x, a, b, s)$ (in the situation s, agent x prefers a over b) and $w(a, b, s)$ (in the situation s, a is preferred over b according to the social welfare function). The intuition is that in each situation, there is a preference ordering for each player (represented by predicate p), and a social welfare function for the society (predicate w). The requirement that the preferences be linear corresponds to the following axioms:

$$p(x, a, b, s) \lor p(x, b, a, s) \lor a = b, \tag{4}$$

$$\neg p(x, a, a, s) \land \neg w(a, a, s), \tag{5}$$

$$p(x, a, b, s) \land p(x, b, c, s) \supset p(x, a, c, s), \tag{6}$$

$$w(a, b, s) \lor w(b, a, s) \lor a = b, \tag{7}$$

$$w(a, b, s) \land w(b, c, s) \supset w(a, c, s), \tag{8}$$

where "$\supset$" is the logical implication operator. We have used the convention that all free variables in a formula are implicitly universally quantified from outside unless stated otherwise. So the full sentence for the first axiom above is:

$$\forall x, a, b, s. p(x, a, b, s) \lor p(x, b, a, s) \lor a = b.$$

We also need an axiom which says that the predicate w indeed represents a function that aggregates individual preferences:

$$[\forall x, a, b. p(x, a, b, s_1) \equiv p(x, a, b, s_2)] \supset [\forall a, b. w(a, b, s_1) \equiv w(a, b, s_2)]. \tag{9}$$

The unanimity condition corresponds to the following axiom:

$$\forall a, b, s. [\forall x\ p(x, a, b, s)] \supset w(a, b, s), \tag{10}$$

the non-dictatorship condition the following axiom:

$$\neg \exists x \forall s, a, b. p(x, a, b, s) \equiv w(a, b, s), \tag{11}$$

and the IIA condition the following one:

$$\forall a, b, s_1, s_2. [\forall x. p(x, a, b, s_1) \equiv p(x, a, b, s_2)] \supset [w(a, b, s_1) \equiv w(a, b, s_2)]. \tag{12}$$

---

[3] In fact, there has been some work on encoding Arrow's axiom system in temporal logic. For example, [1] presented a logic that included Arrow's theorem as a theorem in the logic. However, we propose the new logic language mainly because of its simplicity in syntax and semantics as well as its immediate translation to propositional logic for fast implementation by SAT.

Furthermore, we need to say that each preference profile is represented by some situation (the assumption of unrestricted domain). One way to do it is to introduce an action $swap(x, a, b)$ which when performed will swap the positions of a and b in agent x's preference ordering.

$$p(x, a, b, do(swap(x, a, b), s)) \equiv p(x, b, a, s),$$

where in general, $do(A, s)$ denotes the situation resulting from doing action A in s. We also need other axioms to say that in the new situation, agent x prefers a' over b iff she prefers a' over a before, she prefers a' over a iff she prefers a' over b before, that this action has no effects on the orderings of other pairs of alternatives, and no effect on the preference orderings of other agents. All these can be conveniently specified using Reiter's successor state axioms [17]:

$$p(x, a, b, do(swap(y, a_1, b_1), s)) \equiv p(x, a, b, s) \wedge [x \neq y \vee (a \neq a_1 \wedge a \neq b_1 \wedge b \neq a_1 \wedge b \neq b_1)] \vee$$

$$x = y \wedge a = a_1 \wedge b = b_1 \wedge p(x, b, a, s) \vee$$

$$x = y \wedge a = a_1 \wedge b \neq b_1 \wedge b \neq a \wedge p(x, b_1, b, s) \vee$$

$$x = y \wedge b = b_1 \wedge a \neq a_1 \wedge b \neq a \wedge p(x, a, a_1, s).$$

This way, given an initial situation S_0 that encodes any preference profile, we can get any other preference profile by performing a sequence of swapping actions in S_0.

However, if we are given a specific voting model, we can name each preference profile explicitly by a situation constant. For instance, for the voting model $(\{1, 2\}, \{a, b, c\})$ corresponding to the base case in our proof of Arrow's theorem, there are 36 different profiles, so we introduce 36 situation constants $S_1, \ldots, S_{36}$, and add axioms like the following ones to define them:

$$p(1, a, b, S_1) \wedge p(1, a, c, S_1) \wedge p(1, b, c, S_1),$$

$$p(2, a, b, S_1) \wedge p(2, a, c, S_1) \wedge p(2, b, c, S_1).$$

In fact, this is what we did for using a SAT solver to verify the base case in our inductive proof of Arrow's theorem. We instantiated the axioms (10)–(12) as well as the general axioms about p and w on $(\{1, 2\}, \{a, b, c\})$, and converted them as well as the axioms like the above ones for the 36 situation constants to clauses. The resulting set of clauses has 35973 variables and 106354 clauses, and we were surprised that the SAT solver Chaff2 [14] returned in less than 1 second when run on our AMD server machine and confirmed that the set of clauses has no models.

8. Conclusion and future work

We have given a new proof of Arrow's theorem. The basic idea is extremely simple: use induction to reduce it to the base case which is then verified using computers. One remarkable thing about it is that it appears to be a very general approach for proving other theorems in the area. In fact, we have adapted it almost straightforwardly to proving two other well-known theorems of the same nature, one by Muller and Satterthwaite and the other by Sen.

One insight we have obtained from the proof is that theorems that are verified to be true in the small base cases are extremely likely to be true in general. That is how we have discovered and proved our new theorem in Section 6.

If all these axioms in social choice theory can be checked in base case as fast as those in Arrow's theorem, an interesting future work is to verify all the possible combinations of these candidate axioms using a computer program and then try to extend the survivors to general case using the "two-lemma trick" introduced in the inductive step. To facilitate the above systematical generation and verification process, it becomes nature to describe these axioms in a logical language that is easy in syntax and semantics as well as allows for fast implementation. That is why we have proposed a new logical formalism for social choice theory despite the rich literature. In fact, we did discover this way two theorems, as described in [6]. It is pity that both theorems can be implied immediately by existing theorems. We are still exploring this territory to see if we could come up with something new.

In a recent note [5] in celebration of John Nash's 80th birthday, Binmore remarked,

Nash was not shy of taking his ideas to the big names in the academic world. He famously proposed a scheme for reinterpreting quantum theory to Albert Einstein, who responded by suggesting that he first learn some physics. It is unfortunate that Nash got similar treatment from Von Neumann, when he showed him his existence theorem…

We are not shy of taking our ideas to those big names either, and we get computers to help us.

Acknowledgements

References

[1] T. Agotnes, W. Van der Hoek, M. Wooldridge, Towards a logic of social welfare, in: Proceedings of the Seventh International Conference on Logic and the Foundations of Game and Decision Theory, 2006.

[2] K.J. Arrow, A.K. Sen, K. Suzumura (Eds.), Handbook of Social Choice and Welfare, vol. 1, Elsevier, 2002.

[3] K. Arrow, A difficulty in the concept of social welfare, Journal of Political Economy (1950) 328–346.

[4] S. Barberà, Pivotal voters: A new proof of Arrow's theorem, Economics Letters 6 (1) (1980) 13–16.

[5] Binmore, Nash's work in economics, in: Games and Economic Behavior, 2009.

[6] F. Lin, P. Tang, Computer-aided proofs of Arrow's and other impossibility theorems, in: Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, 2008.

[7] P.C. Fishburn, Arrow's impossibility theorem: Concise proof and infinite voters, Journal of Economic Theory 2 (1) (1970) 103–106.

[8] J. Geanakoplos, Three brief proofs of Arrow's impossibility theorem, Economic Theory 26 (1) (2005) 211–215.

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

[10] F. Lin, P. Tang, Discovering theorems in game theory: Two-person games with unique Nash equilibria, http://www.cs.ust.hk/faculty/flin/papers/zerosum.pdf, 2007.

[11] F. Lin, Finitely-verifiable classes of sentences, in: Proc. of 2007 AAAI Spring Symposium on Logical Formalization of Commonsense Reasoning, http://www.ucl.ac.uk/commonsense07/, 2007.

[12] A. Mas-Colell, M.D. Whinston, J.R. Green, Microeconomic Theory, Oxford University Press, 1995.

[13] J. McCarthy, Situations, actions and causal laws, in: M. Minsky (Ed.), Semantic Information Processing, MIT Press, Cambridge, MA, 1968, pp. 410–417.

[14] M.W. Moskewicz, C.F. Madigan, Y. Zhao, L. Zhang, S. Malik, Chaff: Engineering an efficient SAT solver, in: Proceedings of the 38th Design Automation Conference (DAC'01), 2001.

[15] E. Muller, M.A. Satterthwaite, The equivalence of strong positive association and strategy-proofness, Journal of Economic Theory 14 (2) (1977) 412–418.

[16] R. Myerson, Fundamentals of social choice theory, Northwestern University, Center for Mathematical Studies in Economics and Management Science, Discussion Paper 1162, 1996.

[17] R. Reiter, Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems, The MIT Press, 2001.

[18] S.J. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, second ed., Prentice Hall, Englewood Cliffs, NJ, 2003.

[19] 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) (1975) 187–217.

[20] A. Sen, The impossibility of a Paretian liberal, Journal of Political Economy 78 (1) (1970) 152–157.

[21] K. Suzumura, Welfare economics beyond welfarist-consequentialism, The Japanese Economic Review 51 (1) (2000) 1–32.

[22] R. Wilson, Social choice theory without the Pareto principle, Journal of Economic Theory 5 (3) (1972) 478–486.