Computer-Aided Proofs of Arrow's and Other Impossibility Theorems*

Fangzhen Lin and **Pingzhong Tang**
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 Sen's and Muller-Satterthwaite's theorems as well. Furthermore, we believe this new proof opens an exciting prospect of using computers to discover similar impossibility or even possibility results.

Introduction

Recently, there has been much interest and work in applying economics models such as those from game theory to AI as well as AI techniques to solving problems in economics. 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 in this paper are the impossibility theorems such as those by Arrow (1950), Sen (1970), and Muller and Satterthwaite (1977) in social choice theory (Arrow, Sen, & Suzumura 2002), 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 (Arrow 1950) on the non-existence of rational social welfare function is no 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 (Fishburn 1970; Barbera 1980; Geanakoplos 2005)).

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 1.8GHz CPUs and 8GB 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 for proving other impossibility results such as Sen's and Muller-Satterthwaite's theorems. We shall outline our proof for Muller-Satterthwaite's theorem along this line but leave Sen's theorem to the full version of the paper.

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 other impossibility or possibility results, and to understand the boundary between these two type of results. This is what we think the long term implication of our new proofs of Arrow's and other impossibility theorems will lie, 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. Our work in this direction is still preliminary. However, we do have two results to report, one tries to relax the unanimity condition and the other the IIA condition in Arrow's theorem.

The rest of the paper is organized as follows. We next review Arrow's theorem and then describe our new inductive proof of this result. We then briefly describe how this proof can be adapted to prove Muller-Satterthwaite's theorem. We

*Authors' names are listed according to the alphabetical order.
This work was supported in part by HK RGC CERG 616707.
Copyright © 2008, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

then propose 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 briefly describe our idea on using computers to discover new theorems in social choice theory, and then conclude this paper.

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 $(>_1, ..., >_n)$, 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 $> = (>_1, ..., >_n)$, 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 (1950))*** *For any voting model (N, O), if $|O| \geq 3$, 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 (Fishburn 1970; Barbera 1980; Geanakoplos 2005). We now give yet another one using induction.

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 $t = (t_1, ..., t_n)$, then t_{-i} denotes the tuple $(t_1, ..., t_{i-1}, t_{i+1}, ..., t_n)$. Furthermore, we use (t_{-i}, s) to denote the result of replacing ith item in t by s: $(t_{-i}, s) = (t_1, ..., t_{i-1}, s, t_{i+1}, ..., t_n)$.
- 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 $>^{+a}$ to be the extension of $>$ to O such that for any $a' \in O_{-a}$, $a' >^{+a} a$ (we could insert a anywhere in the order, but for concreteness we put it in the last position). 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

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

  which will be a preference profile of (N, O).

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

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 $n \geq 2, m \geq 3$.*

Proof: Let $N = \{1, ..., n\}$ 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 $> = (>_1, ..., >_n)$ 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 $\forall i \;\; a' >_i^{\prime+a} a''$ iff $a' >_i^{\prime\prime+a} a''$ as well. Since W is IIA, $a' \; W(>^{\prime+a}) \; a''$ iff $a' \; W(>^{\prime\prime+a}) \; a''$. 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

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

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. ■

Note that it is essential for our proof that $m \geq 3$. 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 $n \geq 2, m \geq 3$.*

Proof: Let $N = \{1, ..., n, n+1\}$ 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 $|N| = n + 1 \geq 3$. 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 that $>_{j_1}$ 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:

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

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:

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

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

$$a >_W b. \quad (3)$$

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. ■

Again notice that it is essential for our proof that $|N| = n + 1 \geq 3$, 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]

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. (Russell & Norvig 2003)). 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 $x_1, ..., x_{36}$, 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.

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

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.

Some other impossibility theorems

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's impossibility theorems. We briefly describe how this is done for Muller-Satterthwaite's theorem, and leave Sen's theorem to the full version of the paper.

Arrow's theorem is about social welfare functions which map a preference profile to a preference ordering. In comparison, Muller-Satterthwaite's theorem concerns about so-called *social choice functions* which map a preference profile to an outcome which is supposed to be the "winner" of the voting (as represented by the preference profile).

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

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 (1977))*** *For any voting model (N, O) such that $|O| \geq 3$, 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 $n \geq 2, m \geq 3$.*

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 shown that for any $a \in O$, C_a is weakly unanimous and monotonic, and there is one such a such 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 $n \geq 2, m \geq 3$.*

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 shown that for any pair of agents $i \neq j$, $C_{i,j}$ is weakly unanimous and monotonic, and that there exists one such pair such that $C_{i,j}$ is non-dictatorial. (Actually the proof here is much simpler than the one in the proof of Lemma 2.) ■

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.

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 of Muller-Satterthwaite's theorem that we know of (Muller & Satterthwaite 1977; Mas-Colell, Whinston, & Green 1995) are much more complicated and rely on Arrow's theorem.

A Logical Language for Social Choice Theory

As we mentioned earlier, we are not just interested in alternative proofs of Arrow's and other known theorems. Our long term goal is to use computers to discover theorems in social choice theory, game theory, and others (Lin 2007; Lin & Tang 2007). One implication 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. Towards this end, we propose a logical language for social choice theory.

This language is a variant of the situation calculus (McCarthy 1968; Reiter 2001), 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 unanimity condition corresponds to the following axiom:

$$\forall a, b, s.[\forall x\ p(x, a, b, s)] \supset w(a, b, s), \tag{4}$$

the non-dictatorship condition the following axiom:

$$\neg\exists x \forall s, a, b.p(x, a, b, s) \equiv w(a, b, s), \tag{5}$$

and the IIA condition the following one:

$$\forall a, b, s_1, s_2.[\forall x.p(x, a, b, s_1) \equiv p(x, a, b, s_2)] \supset [w(a, b, s_1) \equiv w(a, b, s_2)] \tag{6}$$

We also need some axioms to say that both p and w represent linear orderings, and that w is a function of p. 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. 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 $S_1, ..., S_{36}$, 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 (4) – (6) 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 (Moskewicz *et al.* 2001) returned in less than 1 second when run on our AMD server machine and confirmed that the set of clauses has no models.

Discovering New Theorems

We have been advocating a methodology of theorem discovering through exhaustive search in small domains using

computers (Lin 2007; Lin & Tang 2007). Along this line, given the structure of our inductive proofs for the impossibility theorems that we have here, an interesting thing to do is to look for similar theorems by systematically exploring a so-called finitely-verifiable class of conditions (Lin 2007), meaning whether a condition in this class is a theorem can be verified by checking if it holds in some small domains.

Our work in this direction is still preliminary. But we do have two results to report, one tries to relax the unanimity condition and the other the IIA condition.

Recall that the unanimity condition says that if every agent ranks one option over another, so must the resulting societal preference ordering. We can try to relax this, and require that consensus work only when every one has the same preference ordering.

Definition 9 *We say that a social welfare function W is* globally unanimous *if $W(>, ..., >) \; = >$ for every preference ordering $>$ of the set of alternatives.*

As it turned out, this weakening of the unanimity condition does not affect Arrow's theorem. For the base case of 2 individuals and 3 alternatives, our depth-first search program shows that every social welfare function that is globally unanimous and IIA is also dictatorial. Thus the almost same inductive proof of Arrow's theorem above shows that this is true in general as well.

Theorem 3 *Provided that there are more than 2 alternatives, every social welfare function that is globally unanimous and IIA is also dictatorial.*

Now let's consider a possible relaxing of the IIA condition. Recall that the condition says that if two preference profiles agree on the ordering between two alternatives, then their corresponding societal preference orderings should also agree on it. We could relax this, and instead of pairs of alternatives, require this only for triples of alternatives:

Definition 10 ***(IIA')*** *A social welfare function W is IIA' if for all alternatives a_1, a_2, a_3 and all preference profiles $>'$ and $>''$, $a_m >'_i a_n$ iff $a_m >''_i a_n$ for each agent i and each m and n in $\{1, 2, 3\}$ implies that $a_m >'_W a_n$ iff $a_m >''_W a_n$ for each m and n in $\{1, 2, 3\}$.*

Using our depth-first search algorithm we can verify that any social welfare function on $(\{1, 2\}, \{a_1, a_2, a_3, a_4\})$ that is unanimous and IIA' is also dictatorial. Thus by suitably modifying the inductive step in our proof of Arrow's theorem, we get the following theorem:

Theorem 4 *Any social welfare function that is unanimous and IIA' is also dictatorial, provided the number of alternatives is not less than 4.*

In fact, this theorem is closely related to Arrow's theorem through the so-called *intersection principle* (cf. Chapter 1 in (Arrow, Sen, & Suzumura 2002)).

Concluding Remarks

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 Sen and the other by Muller and Satterthwaite.

If all these impossibility theorems can be reduced to some small base cases, then an interesting thing to do is to use computers to conduct search in these small domains to try to discover new impossibility or possibility results and to discover the boundaries between these two type of results. Our preliminary results described above indicate that this is a promising line of research, and we are currently pursuing it using our situation calculus formalization of social choice theory given earlier in the paper.

References

Arrow, K. J.; Sen, A. K.; and Suzumura, K., eds. 2002. *Handbook of Social Choice and Welfare*, volume 1. Elsevier.

Arrow, K. 1950. A difficulty in the concept of social welfare. *Journal of Political Economy* 328–246.

Barbera, S. 1980. Pivotal voters : A new proof of arrow's theorem. *Economics Letters* 6(1):13–16.

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

Geanakoplos, J. 2005. Three brief proofs of arrow's impossibility theorem. *Economic Theory* (26(1)):211–215.

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

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

Mas-Colell, A.; Whinston, M. D.; and Green, J. R. 1995. *Microeconomic Theory*. Oxford University Press.

McCarthy, J. 1968. Situations, actions and causal laws. In Minsky, M., ed., *Semantic Information Processing*. MIT Press, Cambridge, Mass. 410–417.

Moskewicz, M. W.; Madigan, C. F.; Zhao, Y.; Zhang, L.; and Malik, S. 2001. Chaff: Engineering an Efficient SAT Solver. In *Proceedings of the 38th Design Automation Conference (DAC'01)*.

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

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

Russell, S. J., and Norvig, P. 2003. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, N.J. : Prentice Hall, 2nd edition.

Sen, A. 1970. The impossibility of a paretian liberal. *Journal of Political Economy* 78(1):152–57.