ARROW'S IMPOSSIBILITY THEOREM: INTRODUCTION AND PROOF

Name	WZQ
Student No.	112019
Class	07161901
Department	CS
Email	
Date	December 1, 2021

Abstract

Arrow's impossibility theorem, named after economist Kenneth J. Arrow, is known as the general impossibility theorem. It states that a clear order of preferences cannot be determined while adhering to mandatory principles of fair voting procedures. Here, I give a brief introduction to this problem and then give a proof method. Hopefully you can enjoy it!

Keywords: Arrow's impossibility theorem, Economy

Contents

1. Introduction

1.1 Why I chose this?

I believe that many people, just like me, think the theorem is very magical when they first see it. It proves that there is no "ideal" voting mechanism in the world. And aggregating individual preferences is hard. Besides, well, this topic can also make me look sincere enough. So, because of interest and score I choose this topic as my term project assignment.

1.2 A vivid example

Here, an example illustrates the type of problems highlighted by Arrow's impossibility theorem. Consider the following example, where voters are asked to rank their preference of three projects that the country's annual tax dollars could be used for: A; B; and C. This country has 99 voters who are each asked to rank the order, from best to worst, for which of the three projects should receive the annual funding.

- 33 votes A > B > C (1/3 prefer A over B and prefer B over C)
- 33 votes B > C > A (1/3 prefer B over C and prefer C over A)
- 33 votes C > A > B (1/3 prefer C over A and prefer A over B)

Therefore,

- 66 voters prefer A over B
- 66 voters prefer B over C
- 66 voters prefer C over A

So a two-thirds majority of voters prefer A over B and B over C and C over A—a paradoxical result based on the requirement to rank order the preferences of the three alternatives.

2. Problem Statement

2.1 Environment: where we are?

- Finite set $\mathbf{A} = \{A, B, C, ...\}$ of at least three different policy options
- Finite number N of different individuals $i = 1, 2, ..., N$
- Each person i has preferences over the policy options $\succeq_i$, which are complete and transitive(strict)
  - so for any individual i and any two policies a and b, either $a \succ_i b$ or $b \succ_i a$
  - and for any three policies a, b, and c, if $a \succ_i b$ and $b \succ_i c$, then $a \succ_i c$
- And that's all.

2.2 Some basic concepts

We're looking for a way to aggregate preferences – we want a way to turn each possible set of individual preferences $\{\succ_1, \succ_2, ..., \succ_N\}$ into a preference relation $\succeq^*$for "society"

- **Social preferences need to be defined for any set of individual preferences**
  - The unrestricted domain of F is all possible combinations of personal preferences. Of course, all personal preferences satisfy completeness and transitivity.
  - We can also call it "universe domain".
- **Social preferences should be complete and transitive**
  - just like the describe in the previous definition.
- **There are at least 3 different policy options**
  - If there are only two candidates, it means that each voter has only two choices, and the paradox has no conditions to produce. In other words, one of the conditions of arrow's impossibility theory "can produce collective decision-making" is automatically satisfied, and you do not need to impose additional restrictions on the voting mechanism.
- **Social preferences should respect Pareto Efficiency (Unanimity)**
  - If everyone in society agrees that policy a is strictly better than b, then the social preferences defined by our social welfare function (SWF[1]) should also strictly prefer a to b.

[1] I will use SWF refer to social welfare function in the following sections.

- In short, we get unanimous vote.

- **The SWF should satisfy independence of irrelevant alternatives(IIA[2])**
  - Basically, this says that if we're trying to figure out whether society prefers a to b, what people think of c should not matter.
  - Suppose we start with some set of individual preferences $\{\succ_1, \succ_2, ..., \succ_N\}$, and the SWF picks a social preference function under which $a \succeq^* b$. If we modify one guy's preferences $\succ_i$ such that his preference between a and b stays the same, the SWF should still pick a social preference function that prefers a to b.

- **Nondictatorship, ideally**
  - a SWF is a dictatorship if the social preference always just reflects the same one guy's preferences, that is, if there's some individual k such that regardless of anyone else's preferences, $a \succ^* b$ if and only if $a \succ_k b$.

- **Ordinal, not cardinal**
  - All we know is whether an individual prefers one policy to another – we have no way to talk about how much he prefers one to another.
  - We don't have money, to allow people to barter.

OK, that's all the basic concepts.

[2] I will use IIA refer to independence of irrelevant alternatives in the following sections.

3. The Theorem And The Proof

Theorem (Arrow). Any SWF which respects transitivity, unanimity, and independence of irrelevant alternatives is a dictatorship.

- The proof has several steps. We assume we have a SWF which satisfies transitivity, unanimity, and IIA, and then show that this means there must be some voter whose preferences always match the social preferences, who is therefore a dictator.

proofs are as following:

3.1 Step 1: The Extremal Lemma

Lemma 1. For any policy b, if **every** individual i ranks b **either** strictly best **or** strictly worst, then $\succeq^*$ must rank b either strictly best or strictly worst as well.

proof:

1. Suppose the lemma were false. There must be two other policy a c which satisfy that $a \succeq^* b \succeq^* c$, although each individual i ranks b either best or worst.
2. Then, we can execute this regulation:
   - For each individual i who has b at the top of their list: move c up to second on the list.
   - And for each individual i who has b at the bottom of their list, move c to the top of the list.
3. Obviously, this measurement doesn't change any individual's rank between a and b. So by IIA, society still prefers a to b.
4. Also, it doesn't change any individual's rank between b and c. So by IIA, b is still preferred to c.
5. And by transitivity, $a \succeq^* b$ and $b \succeq^* c \Rightarrow a \succeq^* c$.
6. However, everyone has c ranked above a, thus unanimity should require $c \succeq^* a$ which gives a contradiction.

3.2 Step 2: Identify individual R_{i^*}

Here, we have already know (by unanimity) that if everyone in society puts b last, then $\succeq^*$ puts b last too; and if everyone puts b first, $\succeq^*$ puts b first.

Also, we have just proofed that if some people put b first and the rest of people put b last, then $\succeq^*$ puts b either first or last.

So, consider this situation: a set of individual preferences where everyone in society puts b last.[1]

[1] it is the same when b is always the best.

$$\begin{array}{ccc} \underline{R_1} & \ldots & \underline{R_n} \\ a & \ldots & a \\ \vdots & \ldots & \vdots \\ b & \ldots & b \end{array}$$

where the dotted ranges represent the other alternatives in fixed but arbitrary locations in the two profiles. Now following this method:

- change R_1 preferences by moving b from last to first. Since everyone in society either likes b the most or the least, b must be either first or last in $\succeq^*$.
- change R_2 preferences by moving b from last to first. The last one become x[2].
- Keep going like this until society ranks b first.

That is, the first time that the social preference $\succeq^*$ switched from having b at the bottom to b at the top. And that's the individual R_{i^*}.

For the profile I

$$\begin{array}{ccc} \underline{R_1 \ldots R_{i^*-1}} & \underline{R_{i^*}} & \underline{R_{i^*+1} \ldots R_n} \\ b & a & a \\ \vdots & \vdots & \vdots \\ x & b & b \end{array}$$

the social preference is $a \succeq^* b$ and for the profile II

$$\begin{array}{ccc} \underline{R_1 \ldots R_{i^*-1}} & \underline{R_{i^*}} & \underline{R_{i^*+1} \ldots R_n} \\ b & b & a \\ \vdots & \vdots & \vdots \\ x & x & b \end{array}$$

the social preference is $b \succeq^* a$.

3.3 Step 3: R_{i^*} is a dictator over policies that aren't b

Start with arbitrary preferences where $a \succeq^* c$, and call these preferences Profile IV.

Make the following changes to preferences:

- Move a to the top of R_{i^*}'s preference list, and b to second on R_{i^*}'s preference list.
- For individuals 1 up to R_{i^*}, move b to the top of their list.
- For individuals R_{i^*+1} up to N, move b to the bottom of their list.
- Call this new set of preferences Profile III

In this progress, we didn't change anyone's ranking of a versus c. So by IIA, the societal preference between a and c has to be the same at profile IV as at profile III. Thus, at profile III, society prefer a to c. And we can proof that:

[2]x means any option that is not equal to a and b.

1. At profile I, society put b last, which means $a \succeq^* b$ at profile I. And everyone's ranking of a versus b is the same at profile I as at profile III. So by IIA, $a \succeq^* b$ at profile III.

2. At profile II, on the other hand, society put b first, which means $b \succeq^* c$ at profile II And everyone's ranking of b versus c is the same at II as at III. So by IIA, $b \succeq^* c$ at profile III.

3. Since preferences at profile III must be transitive, $a \succeq^* c$.

4. So at at profile IV, $a \succeq^* c$. $\Rightarrow$ whenever $a \succ_{R_{i^*}} c$, $a \succeq^* c$.

Thus, R_{i^*} is a dictator over policies that aren't b.

3.4 Step 4: R_{i^*} is also a dictator when it comes to b

This is the easy part. We need to proof that if $b \succ_{R_{i^*}} a$, $b \succeq^* a$; and if $a \succ_{R_{i^*}} b$, $a \succeq^* b$. Now, we pick a policy that's different from a and b, like c. And repeat things we have done before:

1. We move preferences from c worst to c best, find the R'_{i^*} and proof that he is a dictator for any two policies that aren't c.
   - So far, we do not know this new dictator is R_{i^*}.
2. We already know that R_{i^*}'s preferences over b sometimes matter.
   - When we moved from profile I to II, R_{i^*}'s preferences for b were all that changed. And that shifted society from putting b last to putting b first
3. So if the new dictator isn't R_{i^*}, we have a contradiction; so the new dictator must also be R_{i^*}.
   - Because there should only be one dictatorship in all individuals.

3.5 Step 5: It's over

From step 3 and step 4. we know that R_{i^*} is dictator over any pair of policies that excludes b, and over any pair of policies that includes b. Thus, R_{i^*} is a dictator.

4. Acknowledgements

The completion of this assignment is unbelievable for me honestly. Since if I read any paper, I will skip the math part directly. Not to speak of trying to understand and described enigmatic mathematical theorem.

Besides, This assignment was edited on overleaf and I used the template released at beijing-institute-of-technology-report-template.

Finally, I would like to express my sincere gratitude to our teacher, Zhengyang Liu, for his conscientious and warmhearted teaching. I will never forget the man proposed model which tells me to always be proactive.

5. References

[1] https://www.bilibili.com/read/cv7554159/

[2] https://scholar.princeton.edu/sites/default/files/ashvin/files/arrow-theorem.pdf

[3] https://www.rochester.edu/college/faculty/markfey/papers/ArrowProof3.pdf

[4] https://plato.stanford.edu/entries/arrows-theorem/

[5] https://www.cs.cmu.edu/~arielpro/15780/readings/svensson.pdf

[6] https://scholar.harvard.edu/files/maskin/files/the_arrow_impossibility_theorem.pdf?m=1360041257