Usher, Dan

Working Paper

Interpreting Arrow's Impossibility Theorem

Queen's Economics Department Working Paper, No. 1384

Provided in Cooperation with:
Queen's University, Department of Economics (QED)

Suggested Citation: Usher, Dan (2017) : Interpreting Arrow's Impossibility Theorem, Queen's Economics Department Working Paper, No. 1384, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:
https://hdl.handle.net/10419/188896

Queen's Economics Department Working Paper No. 1384

Interpreting Arrow's Impossibility Theorem

Dan Usher
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

6-2017

Interpreting Arrow's Impossibility Theorem

Dan Usher June 25, 2017

Abstract: Arrow's Impossibility Theorem is commonly understood to invoke a dictatorship that is somehow lurking within our voting arrangements. A well-recognized statement of the theorem is that "any constitution that respects transitivity, independence of irrelevant alternatives and unanimity is a dictatorship". The theorem is really not about dictatorship at all. It is more appropriately interpreted as about the spoiler problem, about the possibility that the presence of a candidate who cannot win the election himself may, nevertheless, violate the "independence of irrelevant alternatives" by switching the outcome of the election between two other candidates. The theorem becomes that no electoral system is guaranteed to avoid the spoiled problem altogether, regardless of the options and regardless of voter preferences.

Key Words: Impossibility Theorem, Spoilers, Dictatorship

JEL Code: D60, D72

As stated by Geanakopolis (2005,page 212), the Arrow's Impossibility Theorem is that

"Any constitution that respects transitivity, independence of irrelevant alternatives and unanimity is a dictatorship"

As stated by Sen (2014,page 34), the theorem is

"If there are at least three distinct social states, and a finite number of individuals, then no social welfare function can satisfy ***U, I, D*** and **P**", where

U is unrestricted domain, meaning that any constellation of preferences is admissible,

P is the Pareto principle that social choice always ranks one option over another whenever everybody agrees that the former is preferable,

I is independence of irrelevant alternatives, meaning that *social choice between two options is never dependent on whether or not some third option is available*, and

D, commonly referred to as non-dictatorship, is that social choice is not restricted to any given person's preferences.

Sen goes on to say that "One common way of putting this result is that a social welfare function that satisfies unrestricted domain, independence and the Pareto principle has to be dictatorial. This is a repugnant conclusion – emanating from a collection of reasonable-looking axioms" (2014, 35)

Focussing on the evils of dictatorship Sen adds that "We cannot begin to understand the intellectual challenge involved in Arrow's impossibility theorem without coming to grips with the focus on the informational inclusiveness that goes with a democratic commitment which is deeply offended by a dictatorial procedure. This is so even when the dictatorial result is entailed by axiomatic requirements that seem reasonable, taking each axiom on its own." (2014, 31-32)

The argument is this note is three-fold: i) the implicit meaning of "dictatorship " within the theorem is very different from and very much less threatening than dictatorship as the word is commonly understood, ii) the emphasis in describing the theorem should be not on how the axioms **U, P** and **I** are inconsistent with **D**, but on how the axioms **U, P** and **D** are inconsistent with **I**, and iii) the principal take-away from the theorem should be to design voting systems that minimize the harm from spoilers, from candidates who affect outcomes of elections they cannot win. None of this is criticism of the theorem itself, which, so far as I can tell, is completely correct.

Consider the simplest possible example. Ten people have lunch together every day, and, each day, they must choose one among three kinds of sandwiches, cheese (C) , turkey (T) and ham (H). Assume, no matter why, that they cannot choose sandwiches individually; they must all have the same type of sandwich each day, though they can change types from one day to the next. Suppose four of the ten

people have orders of preference C-H-T (meaning that cheese is preferred to ham which, in turn, is preferred to turkey), three people have orders of preference H-- C and the remaining three have orders of preference T-H-C.

One of the ten people is a dictator, in Arrow's sense of the term if the group's choice among sandwiches reflects that person's preferences exclusively, regardless of what anybody else prefers. If that person prefers ham, then ham is what the ten people get to eat. If that person's order of preference is H-T-C, then T is chosen if only T and C are available. Independence of irrelevant alternatives, **I ,** would automatically hold in that case, but the non-dictatorship axiom, **D**, would be violated. Axioms **U, P** and **I** would be preserved by violating **D**, exactly as he theorem requires. Nothing in the theorem specifies which of the ten people the dictator would be. Appointment of any of the ten people would meet the requirements of the theorem.

Or consider a rule of social choice according to which each person gets to choose the sandwich once every ten days. The ten people take turns as dictator, violating axiom **D** every day, but such dictatorship (as implicitly defined within the theorem) is completely innocuous with none of the evils we normally associate with dictatorship.

Reference to dictatorship in the context of the impossibility theorem invokes images of Hitler and Stalin, images that have nothing to do with what the theorem is really about. There is no Nazi Party or Communist Party. The dictator as conceived in axiom **D** is not an evil fellow who murders people or puts them in concentration camps. He is not the *destroyer* of democratic government or even the *candidate* who wins by rigging the election. He is merely a *voter* who gets his way regardless of anybody else's preference. Typically, a real dictator holds no elections or rigs elections so that he comes out the winner; how the dictator himself votes is almost irrelevant. Putin does not cease to be a dictator if he abstains or if he chooses to vote for the opposition, as would be the case for a dictator in the world of Arrow's theorem. Nothing in the context of the impossibility theorem indicates how the dictator is chosen, requires the dictator to be the same person from one election to the next or even guarantees that dictatorship is harmful. Dictatorship within the context of the impossibility theorem is not what we ordinarily mean by the term. In the "impossibility" world, dictators get what they vote for. Actual dictators get what they want regardless of elections.

A case can be made for abandoning the term "dictator" altogether, and using the term "diversity" instead. Axiom **D** would require that social choice respect society's diversity of preference, at least to the extent that no one person's preference is destined to prevail regardless of the preferences of the rest of the community. Strictly speaking, the meaning of the axiom would be unchanged, but the repugnant connotations of the word dictator would be removed.

Axiom **I**, the independence of irrelevant alternatives, is a projection from individual preference to social choice. It is that the option rejected in a choice between two options is never accepted when a

third option becomes available as well. If a rational person prefers ham to cheese when just ham and cheese are available, his only possible orders of preference among ham, cheese and turkey are ham-cheese-turkey, ham-turkey-cheese and turkey-cheese-ham. With all three options available, this person might select turkey or he might continue to select ham, but he would be irrational to select cheese which has already been rejected when only ham and cheese were available.

In social choice, violation of the independence of irrelevant alternatives is commonly called the spoiler problem. In our simple ten-person example, four people have preferences C-H-T, three people have preferences H-T-C, three people have preferences T-H-C and public choice is by majority rule first-past-the-post voting. With these preferences, the independence of irrelevant alternative may be violated, when H is the spoiler in what was originally an electoral contest between C and T, or when T is the spoiler in what was originally an electoral contest between C and H.

In a vote between C and T alone, the winner is T with 6 votes to 4. Introducing an additional option H creates a three-way election among options C, T and H. To avoid H becoming the spoiler, the outcome of this three-way election must be either a win by the new option H or by the original winner T. Neither of these happens. Instead, the introduction of H switches the outcome of the election from a win by T to a win by C with a plurality of 4 votes against 3 for each of the other options. Spoilers cannot be avoided except by a choice mechanism in which some particular order is imposed (for example H-T-C) regardless of what anybody prefers, in which case people with preferences H-T-C are co-dictators in Arrow's sense of the term.

Axiom **I** is justified by an analogy between individual rationality and social decision-making. A person who violates **I** is crazy. A rational person has a consistent order of preference. A person who chooses ham over cheese when nothing else is available but switches to cheese when turkey becomes available is irrational. Should we think of society as irrational when axiom I is violated in elections? Perhaps we should, but, if so, we must think of society as irrational whenever a spoiler affects the outcome of an election, and we must recognize that such irrationality is ubiquitous in democratic decision-making.

Axiom **I** is very strong. The axiom is that for any and every method of aggregating people's preferences into social preference – by voting or by some other means – there is no constellation of individual preferences, however peculiar or unusual, for which the social choice between two given options is dependent on whether or not some third option is available. Axiom **I** is that there can be no spoiler, that social choice between two options is *never* dependent on whether or not some other option is available, regardless of the voting method and regardless of the constellation of preference. The word "never" is critical. A single incident, however unusual or unrealistic, is enough to violate the axiom.

The most cited violation is the contest among Bush, Gore and Nader in the 2001 US Presidential election where the difference in votes between Bush and Gore was much less that the number of votes for Nader and where, it is commonly assumed, Nader's votes would have gone to Gore rather than

Bush if Nader had dropped out of the race. There are more such incidences, but most elections are not like that. Dependence on irrelevant alternatives is a somewhat rare event. Independence of irrelevant alternatives means that such events can never happen, no matter what the structure of preferences or how individual preferences are combined in social decision-making.

In a recent note on how he came to discover the impossibility theorem (2014, page 144), Arrow had this to say: "The social ordering must satisfy two properties: it must reflect in some sense the preference ordering of individuals, and in making social choice from any given set of alternatives, it should use information about the preference ordering of individuals among those alternatives only. Further, it should be defined for any conceivable set of individual preference orderings, i.e., it is a functional, called the social welfare functional…..The second property above, called Independence of Irrelevant Alternatives, has the particular implication that the choice from any two-alternative set depends only on the preferences of individuals as between those alternatives." Arrow takes it as axiomatic that a process of public decision-making should inherit features of private decision-making, that what is rational for private decision-making should be inherent in public decision-making too. Alas, that is just not so. Though characteristic of any rational person's decision-making, independence of irrelevant alternatives is violated in public decision-making based upon individual preferences.

Following a suggestion by Ian Little, Arrow (1963, page 106) draws a distinction between "a social welfare function" and a "social decision process", where the former can be expected to preserve independence of irrelevant alternatives but the latter cannot. The social welfare function is somebody's assessment of the welfare of the nation as a whole, sometimes, though not necessarily, expressed as a "uniform income equivalent", an income which, if everybody had precisely that income, would create the same social welfare (as seen by the person whose function it is) as the actual distribution of income. The social welfare function is analogous to a person's utility function, with incomes of different people playing the role of amounts of goods, but with one mind combining incomes of different people into a social measure. In the one as in the other, rationality of individual choice – including the independence of irrelevant alternatives - must be preserved.[1]

[1] Imagine a person j whose perception of social welfare, W_j, is representable by average utility

$$W_i = (1/n)\sum_1^n u^j(y_i)$$

where n is total population, y_i is the income of person i and $u^j(y_i)$ is person j's perception of the utility of a person with income y_i. Person j's uniform income equivalent of the entire distribution of income is $\boldsymbol{Y}_j$ defined implicitly by the equation

$$W_j = u^j(\boldsymbol{Y}_j)$$

So defined, the uniform income equivalent is the income such that social welfare as seen by person j would be the same if everybody had that income as it is with the actual income distribution. Imagine

Social decision processes are different, in part because the social welfare function is seen differently by different people and in part because, in voting or other aspects of public decision-making, people act to promote their own interests as well as the interests, as they see them, of society as a whole. Social decision processes cobble together public decisions where there is no universally-recognized social welfare function and where the rationality to be expected from individual decision-making is more than one can reasonably expect.

Strictly speaking, the impossibility theorem is that the four axioms **U**, **P, I** and **D** are inconsistent and cannot all be true at once. In discussing the theory, it is customary to rephrase it as saying that **U**, **P** and **I** cannot all be true unless **D** is false, where D is non-dictatorship and violation of **D** means that there must be a dictator as defined within the theorem. With **D** interpreted as "diversity" (that nobody gets to determine the outcome all by himself), the theorem could equally-well be restated as showing that, together **U**, **P** and **D** require that **I** be false, meaning only that, regardless of how individual preferences are aggregated, there is always some constellation of preferences for which a spoiler may arise.

This alternative formulation seems preferable for two reasons: The first might be called logical. To say that **U**, **P** and **D** require that **I** be false is to infer the possibility of a common occurrence from three generally-recognized properties of public choice. The other interpretation is convoluted. To say that **U**, **P, I** violate **D** is to combine as postulates two generally-recognized characteristics of public choice with a characteristic of individual behaviour which we know does not extend to public choice at all, and to infer from this mix a counter-intuitive inference with little relevance to voting or to any other method of public decision-making. That one postulate we know to be false implies another that is equally false is of little help in actual decision-making.

The other reason is practical. The lesson in the impossibility theorem becomes not that dictatorship is inevitable (for it really has nothing to say about dictatorship), but that a major consideration in the design of voting arrangements and other methods of social choice is, as much as possible, to keep the spoiler away. Keeping the spoiler away is an important argument in favour of the alternative vote (sometimes called instant run-off elections) as compared with ordinary first-past-the-post voting and is central to Maskin and Sen's (2017) advocacy of what they call majority voting. The impossibility

---

an election in which voters' only concerns about the different were the distributions of income that their policies would provide. By assumption, person j wants people to be prosperous and happy, but he doesn't care whether the population is large or small. If all voters' utility of income functions were the same and people voted altruistically to attain the largest attainable social welfare, then the candidate supplying the largest uniform income equivalent would win the election unanimously and independence of irrelevant alternatives would be preserved. If candidate x beats candidate y when candidate z is not in the race, the entrance of candidate z could never swing the election to candidate y. Either candidate z would win or candidate x would remain the winner.

theorem shows that the spoiler cannot be banished completely. He may, nevertheless, be more likely to appear in some arrangements than in others.

How much does all this matter? Much depends on whether the wrong candidate – wrong in the sense that most voters prefer some other candidate - is elected through political machinations or by random quirks in the electoral procedure. As long as mistakes really are random, benefiting neither good guys nor bad guys on average, and not too frequent, they may make little difference in the long run. It does not matter a great deal if the candidate preferred by 49% of the electorate occasionally defeats the candidate preferred by 51% of the electorate when the error is inherent in a voting system that has been in force since time immemorial and is not subject to manipulation by politicians today. The spoiler problem is closer to a random mistake. There are worse electoral diseases: gerrymandering, disenfranchisement of some class of voters to increase incumbents' chances of re-election, huge disproportions in competing parties' access to campaign funds and the not-completely-avoidable risk of a real dictator being elected.

I do not know whether it is possible to reverse the order of the proof of the theorem, postulating **U, P** and **D** and showing that **I** cannot hold. Regardless, the theorem might be understood that way, with emphasis on the implication of **I** that social choice may not be completely independence of irrelevant alternatives. It is independent for some patterns of preferences, but not for every conceivable pattern.

References:

Arrow, Kenneth, *Social Choice and Individual Values*, second edition, Cowles Commission, 1963.

Arrow, Kenneth, "The Origins of the Impossibility Theorem" pages 29-42 in Maskin, E and Sen, A.K , *The Arrow Impossibility Theorem*, Columbia University Press, 2014.

Geanakopolos, J., "Three Brief Proofs of Arrow's Impossibility Theorem", *Economic Theory*, volume26, 2005, 211-215.

Maskin, Eric and Sen, Amartya, "The Rules of the Game: A New Electoral System", *New York Review of Books*, January 19, 2017.

Sen, Amartya, "Arrow's Impossibility Theorem", pages 143-148 in Maskin, E and Sen, A.K , *The Arrow Impossibility Theorem*, Columbia University Press, 2014