Economic Inquiry

Journal of the Western Economic Association

(Formerly WESTERN ECONOMIC JOURNAL)

Volume XII, Number 4 **DECEMBER 1974**

A VERY UNSUBTLE VERSION OF ARROW'S IMPOSSIBILITY THEOREM

ALLAN M. FELDMAN
Brown University

Kenneth Arrow's Impossibility Theorem (1963) is subtle and delicate. But subtlety can be confusing and mysterious, especially when compounded with unintuitive notions like "independence of irrelevant alternatives," "almost decisiveness," and "decisiveness." This paper offers a straightforward proof of the theorem for a special case. The purposes are first, to illustrate the assumptions of the theorem, second, to show plainly why these assumptions imply the non-existence of a rule for generating social preference relations and, third, to show why the Impossibility Theorem remains unaffected by a substantial weakening of the universality assumption.

For simplicity I will assume, as Arrow does in a preliminary argument (1963, pp. 48-51), that there are only two individuals in society, and three social alternatives $\{x, y, z\}$. I will also suppose that no individual is ever indifferent between any two alternatives. Henceforth, aP_ib will mean "i prefers a to b." Individual i's preference ordering is assumed to be complete (he can always choose between any pair of alternatives) and transitive (if he prefers a to b, and prefers b to c, then he prefers a to c). Since there are only three alternatives and indifference is disallowed, there are just six ways individual *1* can order the alternatives. (He can prefer x to y to z, or he can prefer x to z to y, or he can prefer y to x to z, and so on). The same is true of individual *2*. Therefore, if any preference ordering for *1* or *2* is allowed, there are exactly 6 x 6 = 36 different constellations of individual preferences possible in this small society. Figure 1 includes them all.

Each cell in this figure shows a possible pair of rankings of the three alternatives by individuals *1* and *2*. On the left side the alternatives are ordered, from top to bottom, according to trader *1*'s preferences, and on the right side they are ordered according to *2*'s preferences. For example, the 1st row, 4th column cell,

	1	2
1st	x	y
2nd	y	z
3rd	z	x

says that *1* prefers x to y and y to z, while *2* prefers y to z and z to x.

The problem which Arrow considered, and which will be considered here, is this: Does there exist a sensible "collective choice" rule, which can take any constellation of individual preferences (or any cell in Figure 1), and transform that constellation of individual preferences into a social preference ordering? The answer to the question hinges, of course, on what is meant by "sensible." Let us write down five plausible requirements that one might expect a collective choice rule to meet. These five requirements, taken together, will define "sensibleness," and will in fact y that *no sensible rule exists for generating social preference orderings.*

Choices \ Individuals	1	2	1	2	1	2	1	2	1	2	1	2
1st	x	x	x	x	x	y	x	y	x	z	x	z
2nd	y	y	y	z	y	x	y	z	y	x	y	y
3rd	z	z	z	y	z	z	z	x	z	y	z	x
1st	x	x	x	x	x	y	x	y	x	z	x	z
2nd	z	y	z	z	z	x	z	z	z	x	z	y
3rd	y	z	y	y	y	z	y	x	y	y	y	x
1st	y	x	y	x	y	y	y	y	y	z	y	z
2nd	x	y	x	z	x	x	x	z	x	x	x	y
3rd	z	z	z	y	z	z	z	x	z	y	z	x
1st	y	x	y	x	y	y	y	y	y	z	y	z
2nd	z	y	z	z	z	x	z	z	z	x	z	y
3rd	x	z	x	y	x	z	x	x	x	y	x	x
1st	z	x	z	x	z	y	z	y	z	z	z	z
2nd	x	y	x	z	x	x	x	z	x	x	x	y
3rd	y	z	y	y	y	z	y	x	y	y	y	x
1st	z	x	z	x	z	y	z	y	z	z	z	z
2nd	y	y	y	z	y	x	y	z	y	x	y	y
3rd	x	z	x	y	x	z	x	x	x	y	x	x

Figure 1

REQUIREMENTS ON THE COLLECTIVE CHOICE RULE

First a bit of notation. R denotes a social preference relation, so aRb means "a is socially at least as good as b." P is a strict social preference relation, so aPb means "a is socially preferred to b," i.e., aRb and *not* bRa. Finally, I is a social indifference relation, so aIb means "society is indifferent between a and b," i.e., aRb and bRa. Now to the requirements:

(1) *Completeness and Transitivity*. The social preference relations generated by a collective choice rule must be complete and transitive. If some constellation of individual preferences is transformed into a particular R, then for any pair of alternatives a and b, either aRb or bRa must hold, and for any triple a,b,c, aRb and bRc must imply aRc. The requirement says that a collective choice rule must always permit social choices between alternatives, and that social choices must be consistent, or not inherently self-contradictory.

Several well-known collective choice rules *don't* generate complete and transitive social preference relations. For example, unanimous voting produces incomplete social rankings: if *1* prefers x to y and *2* prefers y to x, neither alternative wins a unanimous vote over the other. Majority voting produces nontransitive social rankings. As an example in the two-person case (where the usual voting paradox of Condorcet cannot be generated), consider an R defined by: aRb if a gets at least as many votes as b. Then aIb means a and b tie, while aPb means a gets more votes than b. Suppose *1* prefers x to z to y, and *2* prefers y to x to z. If a vote is taken between x and y, there is a tie (*1* votes for x, and *2* votes for y). If a vote is taken between y and z, there is again a tie. According to the majority voting rule, therefore, xIy and yIz; transitivity for R then requires that xIz. But in a vote between x and z, x gets 2 votes and z none; so xPz, and R is *not* a transitive social preference relation.

(2) *Universality*. A collective choice rule should work no matter what individual preferences happen to be. This means that the rule should give us a social preference ordering for *every* cell in Figure 1, not just for the "easy" ones, like those where there is unanimous agreement (the diagonal cells).

Universality is a strong requirement. If certain restrictions are put on what individual preferences are allowable, majority voting becomes a suitable collective choice rule (i.e., it generates complete and transitive social preferences), providing the number of voters is odd (Black, 1948). (There are two voters in the example above.) If much more stringent restrictions are put on individual preferences, even unanimous voting would work. Why then should one require that a collective choice rule *always* work, no matter how divergent individual feelings happen to be?

First, it is difficult to see where to draw the line between permissible and impermissible individual preferences. Which cells in Figure 1 should be disallowed or ignored? How much diversity can be expected in society? When is there so much conflict that the very idea of social welfare becomes implausible? There are no easy answers to these questions, and so I will not draw arbitrary lines. Second, the Impossibility Theorem remains valid even when the universality requirement is substantially weakened, and I will indicate how much it can be weakened in a subsequent section.

(3) *Pareto Consistency*. A collective choice rule should be consistent with the Pareto criterion: If both individuals prefer a to b, a must be socially preferred to b.

Pareto consistency is a very mild requirement for a collective choice rule. One would not expect it to hold in societies that are ruled by external forces; in which, for example, everyone prefers lust and gambling, on the one hand, to chastity and frugality on the other, but where, according to a Holy Book, the social state of chastity and frugality is preferable to the social state of lust and gambling. Economists naturally would recommend lust and gambling.

(4) *Non-Dictatorship.* A collective choice rule must make no one a dictator. Individual i is defined to be a dictator if his wishes prevail, no matter how j feels; that is, if aP_ib implies aPb for all a and b, irrespective of P_j. Ruling out dictatorship does not mean that it is never possible to have aP_ib implying aPb for all a and b. Obviously, if both parties agree on the relative desirability of all alternatives (so that $P_1 = P_2$, as in the diagonal cells of Figure 1), then it is perfectly reasonable to have the social preference relation agreeing with *1*'s (and *2*'s) preference relation, and in fact, the Pareto consistency requirement makes such agreement necessary. Non-dictatorship simply says that *1* (or *2*) must not *always* prevail, no matter how *2* (or *1*) happens to feel.

Why is dictatorship undesirable? First, it's undesirable because your worst enemy might be dictator. Second, it is not a *collective* choice rule, except in a degenerate sense of the word collective.

(5) *Independence of Irrelevant Alternatives.* If people's feelings change about a whole set of "irrelevant" alternatives, but do not change about a and b, then a collective choice rule must preserve the social ordering of a and b. The social choice between a and b must be *independent* of individual orderings on other pairs of alternatives.

Independence is the most subtle of the five requirements, but it really does make sense. Suppose society chooses communism over democracy when fascism is a third alternative lurking in the wings. Next suppose everyone suddenly changes his mind about the desirability of fascism, but *no one* changes his mind about communism *vs.* democracy. The independence requirement says that, if society is faced with the choice between democracy and communism, and *only those two*, it must still choose communism over democracy.

The standard example of a collective choice rule that violates independence is the rule of weighted voting. Let society be made up of individuals *1* and *2*, as usual, and (continuing the communism, fascism, democracy story) suppose *1*'s initial preferences are for fascism (f) over communism (c) over democracy (d), while *2*'s initial preferences are for d over c over f. Suppose a person's first choice gets a weight of five points, a second choice gets four points, and a third choice gets one point. (The weights are obviously, and unavoidably, arbitrary.) Assume that the collective choice rule is: xRy if x gets as many points as y. When the social choice is between c and d, c gets $4 + 4 = 8$ points, and d gets $1 + 5 = 6$ points; so communism is socially preferred to democracy. Now let *1* become totally disillusioned with fascism; his ordering changes to c over d over f. If the *c vs. d* vote is repeated, *c* gets 5 + 4 = 9 points, and *d* gets 4 + 5 = 9 points. Society has become indifferent between communism and democracy, even though neither *1* nor *2* has changed his mind about these two alternatives!

APPLYING THE REQUIREMENTS

Requirements 1, 2, 3 and 5 can now be applied to Figure 1. The applications should clarify the meanings of the four requirements. They will also lay the groundwork for the proof of the Impossibility Theorem.

The completeness and transitivity and the universality requirements say that, when applied to a cell of Figure 1, a collective choice rule must generate a complete and transitive social ordering and that the rule must generate such an ordering in every cell.

The Pareto consistency requirement says a collective choice rule must respect unanimous opinion: if both *1* and *2* prefer one alternative to another, then society must also prefer the one to the other. For example, given the configuration of individual preferences of the 1st row, 3rd column cell of Figure 1, the Pareto requirement says x and y must be socially preferred to z. Application of Pareto consistency over the entirety of Figure 1 gives rise to Figure 2. Each cell of this figure is produced by applying Pareto consistency to the corresponding cell of Figure 1, and therefore, any rule for generating social preferences must be entirely consistent with Figure 2.

xPy xPz yPz	xPy xPz	xPz yPz	yPz	xPy	
xPy xPz	xPy xPz zPy	xPz		xPy zPy	zPy
xPz yPz	xPz	xPz yPx yPz	yPx yPz		yPx
yPz		yPx yPz	yPz yPx zPx	zPx	yPx zPx
xPy	xPy zPy		zPx	xPy zPx zPy	zPx zPy
	zPy	yPx	yPx zPx	zPx zPy	yPx zPx zPy

Figure 2

Next, consider the independence requirement. Suppose that, when applied to the constellation of preferences

	1	2
1st	x	y
2nd	y	x
3rd	z	z ,

a collective choice rule declares x is socially preferred to y. Then by Requirement 5, x must always be socially preferred to y providing that xP_1y and yP_2x, no matter how *1* and *2* change their feelings about the irrelevant alternative z. Similarly, if we have yPx (or xIy) when

	1	2	
1st	x	y	
2nd	y	x	
3rd	z	z	holds,

then we must have yPx (or xIy) whenever xP_1y and yP_2x. In short, the independence requirement forces a collective choice rule to give rise to social preferences which *agree* over certain constellations of individual preferences.

Let us detail those areas of agreement. Independence requires that all the cells in Figure 1 where xP_1y and yP_2x must yield identical social rankings of x and y. Similarly, all the cells where yP_1x and xP_2y must yield identical social rankings of x and y. There is no presumption, however, that the social x-y ranking on the xP_1y and yP_2x cells need be the same as the social x-y ranking on the yP_1x and xP_2y cells. Such a "neutrality" condition is unnecessary for the proof of the Impossibility Theorem, although it is intuitively appealing and useful in other contexts.

Independence also implies these areas of agreement: all the cells of Figure 1 where xP_1z and zP_2x must give rise to identical social rankings of x and z; all the cells where zP_1x and xP_2z must give rise to identical social rankings of x and z; all the cells where yP_1z and zP_2y must give rise to identical social rankings of y and z; and, finally, all the cells where zP_1y and yP_2z must give rise to identical social rankings of y and z.

All of this information can be incorporated in a third figure. Figure 3A indicates where the social rankings of x *vs.* y must agree because xP_1y and yP_2x in all the X-ed cells, and where the social rankings of x *vs.* y must agree because yP_1x and xP_2y in all the O-ed cells. Figure 3B and 3C show the areas of agreement which arise from applications of the independence requirement to the social choices between x and z, and y and z, respectively.

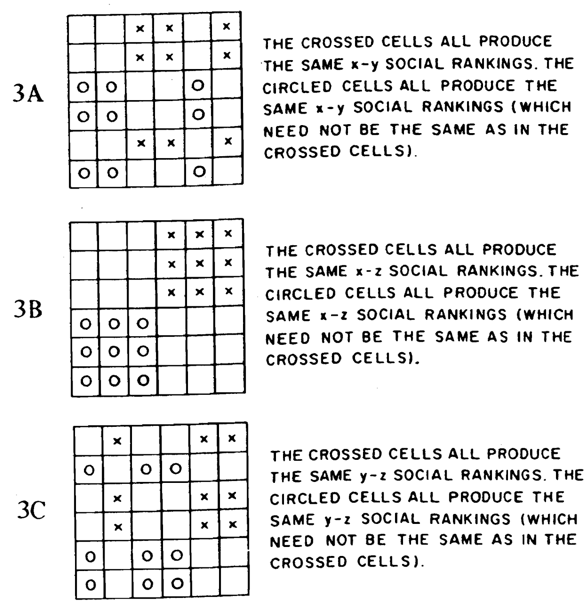

Figure 3

With these preliminaries out of the way, we can turn to the Impossibility Theorem.

THE IMPOSSIBILITY THEOREM

Arrow's Impossibility Theorem:

Any rule for generating social preference relations which is consistent with the requirements of (1) completeness and transitivity, (2) universality, (3) Pareto consistency, and (4) independence of irrelevant alternatives, makes one of the traders a dictator. Therefore, there is no rule which satisfies all five requirements.

Proof:

Start by looking at the constellation of individual preferences in the 1st row, 2nd column cell of Figure 1. For these preferences Pareto consistency requires xPy and xPz (Figure 2). There are three and only three complete and transitive social preference orderings which satisfy xPy and xPz. They are:

(1) xPy, xPz and yPz,
(2) xPy, xPz and zPy, and
(3) xPy, xPz and yIz.

Each of these three possibilities will be considered in turn.

Case 1: yPz. First a word about strategy. The Pareto consistency requirement tells a lot about what social preferences must be, but it leaves a lot unsaid; Figure 2 is full of blank and partially blank spaces. I will now fill in all the blanks by repeatedly applying the independence requirement (Figure 3). When I am done I will have exactly reproduced individual *1*'s preferences and thus made him a dictator.

If yPz holds in the 1st row, 2nd column cell, then independence of irrelevant alternatives (Figure 3C) requires that y be socially preferred to z whenever individual preferences about y and z are the same as they are in that cell. Therefore, yPz holds in all of these cells (those referred to in the argument are circled—they are the crucial ones for the purposes of this proof):

	yPz			yPz	yPz
	yPz			yPz	yPz
	yPz			yPz	yPz

Now consider the 1st row, 5th column cell. Pareto consistency (Figure 2) requires that xPy here, but xPy and yPz implies xPz, by transitivity (Requirement 1). So in this cell one must also have xPz. But if xPz holds in the 1st row, 5th column cell, then independence of irrelevant alternatives (Figure 3B) requires that x be socially preferred to z whenever individual preferences about x and z are the same as they are in that cell. Therefore, xPz holds in all of these cells (those referred to in the argument are circled):

			xPz	xPz	xPz
			xPz	xPz	xPz
			xPz	xPz	xPz

Now consider the 2nd row, 6th column cell. Pareto consistency (Figure 2) requires that *zPy* here, but *xPz* and *zPy* implies *xPy*, by transitivity (Requirement 1). So in this cell one must also have *xPy*. But if *xPy* holds in the 2nd row, 6th column cell, then independence of irrelevant alternatives (Figure 3A) requires that *x* be socially preferred to *y* whenever individual preferences about *x* and *y* are the same as they are in that cell. Therefore, *xPy* holds in all of these cells (those referred to in the argument are circled):

		xPy	xPy		xPy
		xPy	xPy		xPy
		xPy	xPy		xPy

Now consider the 5th row, 4th column cell. Pareto consistency (Figure 2) requires that *zPx* here, but *zPx* and *xPy* implies *zPy* by transitivity (Requirement 1). So in this cell one must also have *zPy*. But if *zPy* holds in the 5th row, 4th column cell, then independence of irrelevant alternatives (Figure 3C) requires that *z* be socially preferred to *y* whenever individual preferences about *z* and *y* are the same as they are in that cell. Therefore, *zPy* holds in all of these cells (those referred to in the argument are circled):

zPy		zPy	zPy		
zPy		zPy	zPy		
zPy		zPy	zPy		

Now consider the 6th row, 3rd column cell. Pareto consistency (Figure 2) requires that *yPx* here, but *zPy* and *yPx* implies *zPx*, by transitivity (Requirement 1). So in this cell one must also have *zPx*. But if *zPx* holds in the 6th row, 3rd column cell, then independence of irrelevant alternatives (Figure 3B) requires that *z* be socially preferred to *x* whenever individual preferences about z and x are the same as they are in that cell. Therefore, zPx holds in all of these cells (those referred to in the argument are circled):

zPx	zPx	zPx			
zPx	zPx	zPx			
zPx	zPx	zPx			

Finally consider the 4th row, 1st column cell. Pareto consistency (Figure 2) requires that yPz here, but yPz and zPx implies yPx by transitivity (Requirement 1). So in this cell one must also have yPx. But if yPx holds in the 6th row, 3rd column cell, then independence of irrelevant alternatives (Figure 3A) requires that y be socially preferred to x whenever individual preferences about y and x are the same as they are in that cell. Therefore, yPx holds in all of these cells (the one referred to in the argument is circled):

yPx	yPx			yPx	
yPx	yPx			yPx	
yPx	yPx			yPx	

This completes the repeated exploitation of the transitivity, Pareto consistency, and independence requirements. At the outset one social preference ranking for one pair of alternatives was assumed and six stages of inference followed that initial assumption. Each stage gave rise to new information about eight pairwise social preference rankings implied by the independence requirement. Five rankings of pairs were found through use of the transitivity requirement, and the Pareto requirement (Figure 2) provided 54 such rankings. The total number of social rankings of pairs of alternatives is 3 x 36 = 108.

Now let me add the information contained in the six diagrams above to the information of Figure 2. Writing *a, b* and *c* vertically over each other to indicate *aPb, bPc* and *aPc*, the accumulated pattern of social preferences looks like this:

x y z	x y z	x y z	x y z	x y z	x y z
x z y	x z y	x z y	x z y	x z y	x z y
y x z	y x z	y x z	y x z	y x z	y x z
y z x	y z x	y z x	y z x	y z x	y z x
z x y	z x y	z x y	z x y	z x y	z x y
z y x	z y x	z y x	z y x	z y x	z y x

But these are precisely trader *1*'s preferences. Therefore, in *Case 1, 1 is a dictator*. He gets his way no matter how *2* feels.

Case 2: *zPy*. If *zPy* holds in the 1st row, 2nd column cell, an argument precisely analogous to the one above establishes that *2 is a dictator*.

Case 3: *yIz*. Suppose *yIz* holds in the 1st row, 2nd column cell. Then by independence of irrelevant alternatives (Figure 3C) *yIz* must also hold in the 3rd row, 2nd column cell, as well as the 4th row, 5th column cell. By Pareto consistency (Figure 2), *z* must be socially preferred to *x* in the latter cell. Now by transitivity (Requirement 1), *yIz* and *zPx* implies *yPx* for the 4th row, 5th column cell. By independence again (Figure 3A), *yPx* in the 4th row, 5th column cell implies *yPx* in the 3rd row, 2nd column cell. Using transitivity (Requirement 1) again, *yIz* and *yPx* implies *zPx* in this cell. However, this contradicts Pareto consistency (Figure 2), that says *xPz* holds here. Therefore, *Case 3* is impossible.

	yIz				
	yIz yPx				
				yIz yPx	

The proof of the theorem is now complete, for it has been shown that Requirements 1, 2, 3, and 5 together imply that either

(i) *1* is a dictator, or
(ii) *2* is a dictator. *Q.E.D.*

Perhaps it would be useful to summarize briefly the proof. *Case 3*, in which *yIz* appears in the 1st row, 2nd column cell, is impossible; it quickly gives rise to a contradiction. *Case 1*, in which *yPz* holds in that starting cell, makes *1* a dictator. It does so because of repeated applications of the independence requirement that, stage by stage, fill in all the blanks left by applications of the Pareto consistency requirement. At each stage, individual *1*'s ranking of a pair is entered in a single box, transitivity is used, and then, through independence, one infers that *1*'s preferences must hold over a large area. When the process is complete nothing is left but *1*'s preferences; he is, therefore, a dictator. The case that makes *2* a dictator is precisely analogous to *Case 1*, and *Cases 1* through *3* exhaust the possibilities.

RELAXING THE UNIVERSALITY REQUIREMENT

It was said at the beginning of this paper that Requirement 2, which demands that a collective choice rule work for *any* constellation of individual preferences, is overly strong. This section will indicate why. In fact, the construction of a sensible collective choice rule remains impossible even if a large number of possible constellations of individual preferences are disallowed.

To show this it will suffice to indicate which constellations of preferences are actually used in the proof of the theorem. In Figure 4 below, the shaded cells correspond to constellations of individual preferences exploited in *Case 3*. The cells used in *Case 1* contain Arabic numerals ordered according to the order of their appearances in the proof. A sequence of cells that could be used to establish *Case 2* is indicated with Roman numerals. The shaded and numbered cells, therefore, represent a whole set of *crucial* instances, and the Impossibility Theorem holds *even if any or all of the remaining cells are discarded.* Incidentally, the indicated cells are not the only set which establish the theorem—there are other sequences of steps which prove it. But they are a "full set," in the sense that they will do the job, and as long as they are all retained, any or all of the other cells are disposable.

There are 13 shaded and numbered cells, out of a total of 36 different constellations of individual preferences. Therefore, the universality requirement could be weakened substantially without destroying the the-

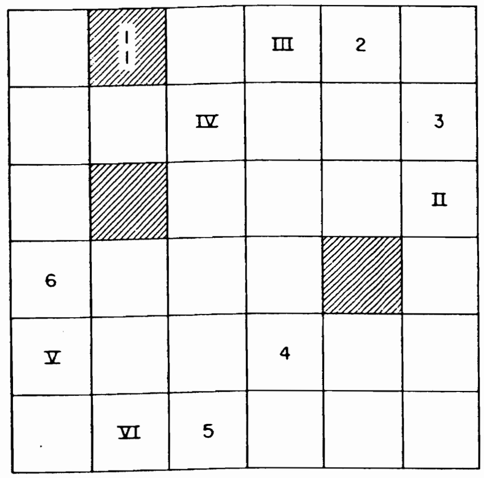

Figure 4

orem; in fact, 23 of the 36 constellations of preferences could be disallowed, including, for instance, those six constellations of individual preferences where there is a total disagreement between the two individuals.

CONCLUDING REMARKS

This paper has provided a proof of Arrow's Impossibility Theorem that depends on an exhaustive examination of all configurations of individual preferences. In the proof, it is clear that the crucial requirements of Pareto consistency and independence of irrelevant alternatives force the pattern of social preferences to mirror the pattern of *one* individual's preferences and, therefore, make one individual a dictator. By seeing exactly which requirement implies what about social preferences, we get a feeling for the nature of the requirements and their strengths. By tracing the sequence of implications of the independence requirement, we see how it forces social preferences to line up with the dictator's. And finally, by considering which configurations of preferences are critical for the proof of the Impossibility Theorem, we get some intuition about how the universality requirement can be weakened without weakening the profound conclusion of the theorem: there is no rule for constructing social preference relations which is not in some fundamental way objectionable.

REFERENCES

1. K. J. Arrow, *Social Choice and Individual Values*, 2nd edition, John Wiley & Sons, New York, 1963.
2. D. Black, "On the Rationale of Group Decision Making," *Jour. Pol. Econ.*, *56*, 1948.