GEORGES BORDES AND NICOLAUS TIDEMAN

INDEPENDENCE OF IRRELEVANT ALTERNATIVES IN THE THEORY OF VOTING

ABSTRACT. In social choice theory there has been, and for some authors there still is, a confusion between Arrow's *Independence of Irrelevant Alternatives* (IIA) and some *choice consistency* conditions. In this paper we analyze this confusion. It is often thought that Arrow himself was confused, but we show that this is not so. What happened was that Arrow had in mind a condition we call *regularity*, which implies IIA, but which he could not state formally in his model because his model was not rich enough to permit certain distinctions that would have been necessary. It is the combination of regularity and IIA that he discusses, and the origin of the confusion lies in the fact that if one uses a model that does not permit a distinction between regularity and IIA, regularity looks like a consistency condition, which it is not. We also show that the famous example that 'proves' that Arrow was confused does not prove this at all if it is correctly interpreted.

Keywords: Social choice, voting theory, choice consistency, Arrow, history of science.

1. INTRODUCTION

The somewhat quixotic purpose of this paper is to clarify the meaning of Arrow's (1951/1963) Condition of *Independence of Irrelevant Alternatives* (IIA) in the theory of voting, and, correlatively, to state clearly and unambiguously what the Arrow Theorem means.

This may seem useless, and in truth it should be. After nearly forty years, what the Condition and the Theorem mean should be common knowledge, not worth talking about except in textbooks. That it is not useless is indicated by the fact that in a recent book on voting procedures Michael Dummett (1984) misinterprets IIA and the Arrow Theorem, and he is far from being the only one to do so.

A kind of legend has grown up around IIA and 'what Arrow meant by IIA'. It focuses on the example Arrow gave of a supposed violation of IIA (1951/1963, p. 27), which has been 'proved' to be not a violation of IIA as stated as Condition 3 on the same page, but of a consistency condition. Also, part of Arrow's informal discussion of IIA is called upon either to 'prove' that Arrow confused IIA with a consistency condition (first thesis) or that Arrow meant IIA to contain

Theory and Decision **30**: 163–186, 1991.

a consistency condition (second thesis). The first thesis, that Arrow was confused, was the view of one of us (Bordes), while the second thesis, that Arrow meant IIA to contain a consistency condition (the apparent view of Dummett among others), was the view of the other of us (Tideman) prior to this collaborative effort.

In this paper we show that both theses are wrong. Arrow was not confused, though it is understandable that some of his readers were confused by what he wrote. Arrow meant (slightly) more by IIA than what is in Condition 3, but what he meant is not what Dummett and others have thought he meant. His model was simply insufficient for him to state formally what he meant.

We proceed in the following way. First, we build a formal model of voting rules which is more general than the 'Social Choice Functions' model currently in use. Then we introduce a new condition on voting rules, which cannot be formally stated in the usual model, and prove that this condition implies IIA, though it looks like a 'consistency condition', which it definitely is not. With this result in mind, we then re-read the crucial pages 26–27 of Arrow's book, and show that it was this new condition that Arrow meant by IIA (but could not state formally because of his model), and that it was because this condition looks like a consistency condition that Arrow has been misinterpreted. We show that Arrow's example can be interpreted both as illustrating a breach of consistency and as illustrating a breach of this new condition. We conclude by discussing how the Arrow Theorem is then to be interpreted. (One last word of warning before proceeding: In what follows, we are concerned only with voting theory; we are *not at all* concerned with the controversial issues raised regarding Arrow's theorem in the quite different field of welfare economics.)

2. THE MODELLING OF VOTING RULES

In many papers dealing with voting rules, the 'Notations and Definitions' section begins with something like this:

> X is the set of alternatives and K is the set of all non-empty finite subsets of X. $V = \{1, 2, \ldots, n\}$ is the set of voters. We denote by r_i the preferences of voter i on X. Preferences

> are assumed to be total, reflexive and transitive, that is, complete weak orderings of X. An n-tuple $(r_1, \ldots, r_n)$ of preferences is called a profile. D is the set of logically possible profiles. A *Social Choice Function* (SCF) F is a function from $K \times D$ into K such that for all $A \in K$ and all $(r_1, \ldots, r_n) \in D$: $F(A, (r_1, \ldots, r_n)) \subseteq A$.

Then the author goes on, considering some particular class of SCFs, proving theorems and commenting on them, and so on. He thinks, and his readers think too, that the SCFs (or the corresponding model of collective choice rules) are a model of *voting rules*, and that the results he gets for SCFs are results about voting rules. He is almost right: the SCF model is a good model for voting rules, *but it is not a complete model*. More precisely, the SCF model is a model for voting rules which is complete enough for almost, but not quite, everything. The fact that this was not recognized is at the origin of the confusion about Independence of Irrelevant Alternatives.

Let us put things this way. When one speaks about a 'voting rule', one does not have in mind any precise SCF. An SCF is a *function*, and a function cannot be defined independently of its domain and range. For example, let us consider the *plurality* voting rule. It is quickly and precisely defined as:

> The candidate that receives the greatest number of first-place votes is elected.

(We do not consider – it has no bearing on the problem – what happens in cases of ties.)

Suppose two countries use the plurality rule for the election of their presidents. At the time of the election, in country I the set of potential candidates is X, the set of non-empty finite subsets of X is K, the set of voters is V and the set of possible profiles is D. For such a situation, the plurality rule yields an SCF F. But in country II, the set of potential candidates is not X but Y (with $X \cap Y = \emptyset$), the set of non-empty finite subsets of Y is not K but H, the set of voters is not V but W, and so on. So for country II, the *same* voting rule yields a *different* SCF, G. F and G cannot be identical since they have neither the same domain nor the same range. A function is a function.

So what is really meant when it is said (or implied) that the SCF model is a model for voting rules is this: depending on what X, V and so on are, a given voting rule will yield different SCFs. But if one takes any of these SCFs yielded by the voting rule and proves theorems about it that do not depend on the fact that one has chosen this particular SCF rather than any other to represent the voting rule, then these theorems are actually theorems about the voting rule itself. This is quite true, and raises no problem. Most of the time . . .

But . . .

So let us consider a more general model for voting rules. A voting rule is 'something' that defines an SCF for each set X of alternatives (and hence its family k of non-empty finite subsets) and each set V of voters (and hence the set D of profiles). Let $\mathbf{X}$ be some 'universal' set of sets of alternatives. That is, $X \in \mathbf{X}$ means that X is some possible set of alternatives. Note that what is written is $X \in \mathbf{X}$ and *not* $X \subset \mathbf{X}$. Let $\mathbf{V}$ be some 'universal' set of sets of voters. $V \in \mathbf{V}$ means that V is some possible set of voters. Let Ω and P be the functions defined in the following way:

> For all $X \in \mathbf{X}$: $\Omega(X) = K$ = the set of non-empty finite subsets of X.
> For all $X \in \mathbf{X}$, all $V \in \mathbf{V}$: $P(X, V) = D$ = the set of all logically possible profiles when the set of alternatives is X and the set of voters is V.

We can now define:

> A *voting rule* is a function $\mathbf{F}$ defined on $\mathbf{X} \times \mathbf{V}$ such that for all X and V, $\mathbf{F}(X, V)$ is an SCF the domain of which is $K \times D = \Omega(X) \times P(X, V)$.

3. A THEOREM ABOUT VOTING RULES

Let us go back to the SCFs. In the definition of an SCF, X is some set of *potential candidates*. On the other hand, an $A \in K$ is a *possible set of actual candidates*. For example, in a French presidential election, X is

the set of all the individuals that have a constitutional right to run for the presidency (something like 20 or 30 million people), but any set of actual candidates will (happily) contain less than a dozen individuals. A set A of actual candidates being given, the elements of $X \backslash A$ are the *potential-but-not-actual* candidates (Georges Bordes, for example, in a French presidential election).

Now consider a 'usual' voting rule, for example the plurality rule. The set of voters $V = \{1, \ldots, n\}$ *being given* once and for all, for a given set X of potential candidates, the choice rule yields an SCF, F. The profile $(r_1, \ldots, r_n)$ being given, and also the set of actual candidates A, we get the choice $F(A, (r_1, \ldots, r_n))$. This choice is made by electing the actual candidate that gets the greatest number of first-place votes when compared to the other actual candidates. That is, A being given, each voter indicates which of the candidates in A he considers best, and the candidate in A who gets the greatest number of votes is elected. (We ignore the problem of ties.)

Obviously, with such a rule, what happens with respect to the potential-but-not-actual candidates does not matter as long as the voters' preferences over the actual candidates do not change. For example, if Mr. x who is a potential-but-not-actual candidate dies, it has no effect on the choice from A (except maybe, of course, if, as in a French presidential election, Mr. x is also a voter – but we suppose that we do not have this difficulty).

We will call such a voting rule *regular*. More precisely, first we denote:

> If R is any binary relation on a set S, and if T is a subset of S, $R|_T$ is the *restriction of R to T*.

Then:

DEFINITION 1. A voting rule $\mathbf{F}$ is *regular* iff: for all (X, V) and (Y, W) in its domain, with $F = \mathbf{F}(X, V)$ and $G = \mathbf{F}(Y, W)$, if $W = V$ and $Y \subseteq X$, then for all $B \in \Omega(Y)$ (and hence B belonging to $\Omega(X)$ since $\Omega(Y) \subseteq \Omega(X)$) and all $(r_1, \ldots, r_n) \in P(X, V)$,

$$F(B, (r_1, \ldots, r_n)) = G(B, (r_1|_Y, \ldots, r_n|_Y)) .$$

Regularity means that, given a voting rule, a set of actual candidates, a set of voters and a preference profile, if the set of potential-but-not-actual candidates shrinks but the voters' preferences over the remaining potential candidates (including the actual ones) do not change, then the choice from the set of actual candidates does not change.

Now, let us turn to an 'old friend':

DEFINITION 2. X and V being given and hence $K = \Omega(X)$ and $D = P(X, V)$, an SCF F satisfies *Independence of Irrelevant Alternatives* (IIA) iff: for all $B \in K$ and all $(r_1, \ldots, r_n), (r'_1, \ldots, r'_n) \in D$, if for all $i \in V$: $r_i|_B = r'_i|_B$, then

$$F(B, (r_1, \ldots, r_n)) = F(B, (r'_1, \ldots, r'_n)) .$$

IIA means that if the voters' preferences over the potential-but-not-actual candidates change while their preferences over the actual candidates stay the same, then the choice among the actual candidates stays the same.

Now we can state:

THEOREM. *If a voting rule is regular, then the SCFs it yields satisfy IIA.*

Proof. Let **F** be a regular voting rule, and let the antecedent of IIA hold. Consider the SCF: $G = \mathbf{F}(B, V)$. The conditions for the application of regularity hold, and so we have:

$$F(B, (r_1, \ldots, r_n)) = G(B, r_1|_B, \ldots, r_n|_B)) \quad \text{and}$$
$$F(B, (r'_1, \ldots, r'_n)) = G(B, (r'_1|_B, \ldots, r'_n|_B)) ,$$

and hence the conclusion of IIA. □

This result is intuitively obvious since regularity for the voting rule and IIA for the SCFs mean almost the same thing. In *real world* voting, the voters give *only* information about their preferences with respect to the *actual* candidates (and usually not even all of that information), and do

not give any information about their preferences over the potential-but-not-actual candidates. Obviously, no voting rule can make any use of information it does not have, and what transformation occurs in the information a rule does not have, whether that information shrinks (regularity) or changes (IIA), does not matter as long as the information the rule has and makes use of does not change.

So to ask a voting rule to be a regular or an SCF to satisfy IIA means exactly that we want it to represent a possible *real-world* voting rule and not some fairy-world voting rule. In fact, in the Arrow Theorem we can do without the conditions of Universal Domain, Choice Consistency, and Unanimity and still have a recognizable voting rule, *but we cannot do without IIA*. For it is IIA (along with Nondictatorship) that ensures that the Arrow Theorem is a theorem about real-world voting rules. Without IIA it would be a theorem about a fairy-world, and hence quite uninteresting.[1] Dummett (1970, p. 54) is therefore wrong when he writes that IIA ". . . lacks complete intuitive justification". But in fact, as can be seen from p. 118 and other places in his book as well, what Dummett thinks is IIA is not IIA. And Dummett is not the only one to make this mistake. So in the next section we attempt to clear up the confusion over IIA.

4. CLEARING UP THE CONFUSION OVER IIA

At least since the publication of P. Ray's paper (1973), it should have been clear that IIA has at times been confused with the following condition, which we call C (for consistency):

DEFINITION 3. An SCF F with domain (K, D) satisfies condition C iff: for all $(r_1, \ldots, r_n) \in D$ and all $A, B \in K$, if $A \subseteq B$ and $A \cap F(B, (r_1, \ldots, r_n) \neq \emptyset$, then

$$F(A, (r_1, \ldots, r_n)) = A \cap F(B, (r_1, \ldots, r_n)) .$$

Condition C (sometimes called WARP: Weak Axiom of Revealed Preference) is to be found as Condition C4 in Arrow (1959). In Arrow's *Social Choice and Individual Values* (1951/1963), the role of Condition C is played by the slightly stronger Axioms I and II:

AXIOM I: *For all x and y, either xRy or yRx.*

AXIOM II: *For all x, y, and z, xRy and yRz imply xRz.*

where 'xRy' is defined as 'x is preferred or indifferent to y' (Arrow, 1951/1963, pp. 12–13).

The following result is now classical:

> An SCF F satisfies C iff for all $(r_1, \ldots, r_n) \in D$ there exists a complete weak ordering of X, R, such that for all $B \in K$: $F(B, (r_1, \ldots, r_n)) = \{X \in B \mid \text{for all } y \in B: xRy\}$.

It follows that with an SCF F that satisfies C, one can associate a function F° from D into the set of complete weak orderings of X, and conversely that such a function F° determines uniquely an SCF F that satisfies C. The F° functions will be called *Arrow functions*.

Now it is obvious that these two conditions, IIA and C, are different:

- IIA concerns what happens to the choice when *the set of actual candidates being given*, *the profile changes* in a certain way;
- C concerns what happens to the choice when *the profile being given*, *the set of actual candidates changes* in a certain way.

Not only are they different, they can be proved to be logically independent. One can exhibit SCFs that satisfy none, one but not the other, or both. Though in the last case, it follows from a theorem by Wilson (1972) that the SCFs that satisfy both are rather 'unpalatable', which is in fact the gist of the Arrow Theorem, of which Wilson's theorem is a generalization. That is, IIA is an *interprofile* property while C (WARP) is an *intraprofile* property. It is to be remarked, by the way, that the condition of 'regularity' for voting rules cannot be definitely labeled as either 'interprofile' or 'intraprofile', since this distinction does not apply straightforwardly for conditions on voting rules: it applies to conditions on SCFs, which are representations of voting rules, a distinctly different species.

So we are confronted with the following question: why two conditions that are different, that look different, that are logically independent, have been confused one for the other or thought to be logically related in some way?

Some, like Radner and Marschak (1954), were accessory after the fact, when they compared a condition by Chernoff (1954), which is related to C, to Arrow's IIA. But the original culprit is Arrow, though he is not guilty of what he is usually thought to be guilty of.

One source of confusion is Arrow's use of two types of names and a hierarchical structure for his assumptions in *Social Choice and Individual Values* (1951/1963). After introducing Axioms I and II, Arrow defines preference and indifference (p. 14):

DEFINITION 1: *xPy is defined to mean not yRx.*

DEFINITION 2: *xIy means xRy and yRx.*

Arrow proves some lemmas about relationships among R, P and I, and then defines the concept of choice (p. 15):

If S is the set of alternatives available, which we will term the *environment*, let $C(S)$ be the alternative or alternatives chosen out of S . . . .

DEFINITION 3: *$C(S)$ is the set of alternatives x in S such that, for every y in S, xRy.*

At the end of the subsequent section, the Ordering of Social States, Arrow writes (p. 19):

Similarly, society as a whole will be considered provisionally to have a social ordering relation for alternative social states, which will be designated by R, sometimes with a prime or a second. Social preference and indifference will be denoted by P and I, respectively, primes or seconds being attached when they are attached to the relation R.

Throughout this analysis it will be assumed that individuals are rational, by which is meant that the ordering relations R_i satisfy Axioms I and II. The problem will be to construct an ordering relation for society as a whole that will also reflect rational choice-making, so that R may also be assumed to satisfy Axioms I and II.

The meaning of 'R' has thus been changed from 'preferred or indiffer-

ent' to 'socially preferred or indifferent according to a relation that satisfies Axioms I and II'.

The next element in the development of the formal structure (p. 23) is

DEFINITION 4: *By a* social welfare function *will be meant a process or rule which, for each set of individual orderings* $R_1, \ldots, R_n$ *for alternative social states (one for each individual), states a corresponding social ordering of alternative social states,* R.

What Arrow called a 'social welfare function' is what we call an 'Arrow function'. The use by Arrow of the term 'social welfare function' appears to be at the origin of other confusions that are outside the scope of this paper. For that reason we chose to use the neutral term 'Arrow function'.

Following Definition 4, Arrow proceeds to develop the other assumptions employed in his theorem (pp. 24–31). These, however, are called not 'Axioms', but rather 'Conditions'. In specifying the Condition, Arrow makes use of three additional definitions:

Let an *admissible* set of individual ordering relations be a set for which the social welfare function defines a corresponding social ordering, i.e., a relation satisfying Axioms I and II. (p. 24)

DEFINITION 5: *A social welfare function will be said to be* imposed *if, for some pair of alternatives* x *and* y, xRy *for any set of individual orderings* $R_1, \ldots, R_n$, *where* R *is the social ordering corresponding to* $R_1, \ldots, R_n$.

DEFINITION 6: *A social welfare function is said to be* dictatorial *if there exists an individual* i *such that, for all* x *and* y, xP_iy *implies* xPy *regardless of the orderings* $R_1, \ldots, R_n$ *of all individuals other than* i, *where* P *is the social preference relation corresponding to* $R_1, \ldots, R_n$.

The five Conditions are:

Condition 1: *Among all the alternatives there is a set of* S *of three alternatives such that, for any set of individual orderings* $T_1, \ldots, T_n$ *of the alternatives in* S, *there is an admissible set of individual orderings* $R_1, \ldots, R_n$ *of all alternatives such that, for each individual* i, xR_iy *if and only if* xT_iy *for* x *and* y *in* S.

Condition 2: *Let* $R_1, \ldots, R_n$ *and* $R'_1, \ldots, R'_n$ *be two sets of individual ordering relations,* R *and* R' *the corresponding social orderings, and* P *and* P' *the corresponding social preference relations. Suppose that for each* i *the two individual ordering relations are connected in the following ways: for* x' *and* y' *distinct from a given alternative* x, $x'R'_iy'$ *if*

and only if $x'R_iy'$; for all y', xR_iy' implies xR'_iy'; for all y', xP_iy' implies xP'_iy'. Then, if xPy, $xP'y$.

Condition 3: *Let $R_1, \ldots, R_n$ and $R'_1, \ldots, R'_n$ be two sets of individual orderings and let $C(S)$ and $C'(S)$ be the corresponding social choice functions. If, for all individuals i and all x and y in a given enviroment S, xR_iy if and only if xR'_iy, then $C(S)$ and $C'(S)$ are the same* (independence of irrelevant alternatives).

Condition 4: *The social welfare function is not to be imposed.*

Condition 5: *The social welfare function is not to be dictatorial* (nondictatorship).

A careful reading of the Conditions reveals that they are defined in a framework that presumes the Axioms: Condition 1 refers to an 'admissible set', which is defined in terms of the Axioms. Condition 2 refers to R and P; Condition 3 refers to $C(S)$ which is defined in terms of R; Conditions 4 and 5 refer to the social welfare function, which is defined in terms of R. And R, the quotation from page 19 implies, is to be assumed to satisfy the Axioms. However, it would be easy to overlook the fact that the Conditions presume the Axioms. Thus a reader, understanding the Arrow Theorem to be one of the inconsistencies of a set of conditions, might ask, 'Which one of Arrow's Conditions do actual voting rules violate'? and overlook the possibility that the answer could be 'None of them or all of them, depending on how you look at it, as they violate his Axioms I and II'.

In Arrow's original statement of his theorem (p. 59) he makes it clear that Axioms I and II as well as the conditions are involved:

THEOREM 2 (General Possibility Theorem): *If there are at least three alternatives which the members of the society are free to order in any way, then every social welfare function satisfying Conditions 2 and 3 and yielding a social ordering satisfying Axioms I and II must be either imposed or dictatorial.*

However, when Arrow revised and slightly weakened the theorem in the second edition (1963, Theorems 2 and 3 of Chapter VIII on pages 97 and 103 respectively), to take account of Blau's (1957) discovery of an error in the proof and a counterexample, he did not mention the Axioms:

THEOREM 2: *Conditions* 1′, 3, *P, and* 5 *are inconsistent.*

THEOREM 3: *Conditions* 1, 3, *P, and* 5′ *are inconsistent.*

The newly introduced conditions are:

Condition P: *If* xP_iy *for all* i, *then* xPy.

Condition 1′: *All logically possible orderings of alternative social states are admissible.*

Condition 5′: *Among the triples of alternatives satisfying Condition* 1, *there is at least one on which no individual is a dictator.*

The Axioms are definitely needed for Theorems 2 and 3 of Chapter VIII; they are needed in particular for the following step on page 100:

Hence, yRz, and, since xPy, society must prefer x to z.

In view of the fact that the Axioms were mentioned in the original statement of the Theorem, the absence of any reference to them in the revised statements is most easily interpreted as reflecting an implicit understanding that Theorems 2 and 3 of Chapter VIII, like Theorem 2 of the first edition, are theorems about "every social welfare function . . . yielding a social ordering satisfying Axioms I and II . . .". That is, the Conditions are defined in a framework where the Axioms hold. The possibility that readers would not understand that Axioms I and II are required for the framework in which Theorems 2 and 3 of Chapter VIII are defined is more unfortunate in view of the fact that Theorem 2 of Chapter VIII is now often regarded as the canonical version of the Arrow Theorem.

A failure to take note of the relationship between the Axioms and the Conditions may also have contributed to the current idea among some people who know of the confusion between IIA and C (or some form of C) and who have read Arrow's comments on IIA (1951/1963, pp. 26–28), that Arrow himself confused IIA and C. And in fact, when one reads these pages, it is easy to feel that Arrow is sometimes talking about IIA and sometimes about C, and that the example he gives of a violation of IIA is actually an example of a violation of C.

But we claim that if Arrow is guilty of anything in these pages, it is of having used an insufficient model, not of having confused IIA with C.

Before presenting our case, we identify a consequence of C, which we call C′:

$$\text{If } F(B, (r_1, \ldots, r_n)) \subseteq A \subseteq B, \text{ then}$$
$$F(A, (r_1, \ldots, r_n)) = F(B, (r_1, \ldots, r_n)) .$$

Now we are ready to proceed.

Consider regularity. While it seems extraordinary that anyone should confuse IIA with C, a confusion between C (or even more so C′) and regularity is plausible. In both cases, there is a *given profile*, and a *set of candidates that shrinks* (and for C′ *a choice that stays the same*). Of course, in the two cases, it is not the same set of candidates that shrinks: For regularity, it is the set of *potential* candidates, shrinking in such a way that the original set of actual candidates is still included in the set of potential candidates after the shrinking. For C′, on the other hand, it is the set of *actual* candidates, shrinking in such a way that the set of winning candidates is still included in the set of actual candidates. But if one uses a loose model, in which one speaks of 'candidates', making no formal distinction between actual candidates and potential candidates, then how can one make the distinction between regularity and C′? The answer is that one cannot. And Arrow's model is such a loose model.

So our claim is that while informally discussing IIA (pp. 26–28), Arrow did not confuse IIA with C or C′, but was simultaneously discussing regularity and IIA, which he considered (almost correctly) as equivalent. ('Almost' because regularity of a voting rule is in fact slightly stronger than IIA for the SCFs it yields.[2] But that need not concern us here, and we can go on acting as if regularity and IIA were equivalent.)

Everything points to this explanation. First, Arrow could not have confused IIA and C or C′. In his model, he is concerned mainly with Arrow functions (functions that satisfy his Axioms I and II), and at the time the book was written, he knew that if an SCF derived from an Arrow function as above, then it would necessarily satisfy C. C and the consequence of C, that the SCF under consideration derives from an Arrow function, were not published formally before 1959 (Arrow, 1959), but these results are found in an earlier hectographed note by

Arrow (1948). (See Arrow, 1959, footnote 2, p. 123.) So why would Arrow insist that the SCF should satisfy C or C′ since he already knew that, since the SCFs he implicitly considered derived from Arrow functions, they would necessarily satisfy C and C′?

Second, if one reads pages 26 and 27 with the idea in mind that Arrow meant regularity and IIA, one finds no contradiction. (The case of the unhappy example will be explained later.) The opening sentence of Section 3 of Chapter III of *Social Choice and Individual Values* (p. 26) is:

> If we consider $C(S)$, the choice derived from the social ordering R, to be the choice society would actually make if confronted with a set of alternatives S, then, just as for a single individual, the choice made from any fixed environment S should be independent of the very existence of alternatives outside of S.

What is meant by this sentence is clearly regularity, and not C or C′. It states that for 'any fixed environment', that is, any given set of actual candidates, the choice should be independent of the very existence of alternatives outside it, that is, independent of whether potential-but-not-actual candidates exist, and hence, if such potential-but-not-actual candidates do exist, independent of who they are and of what the voters' preferences over them are. This is regularity.

The next sentences have sometimes been presented as a 'proof' that Arrow meant C by IIA, or confused C and IIA. Let us take a closer look:

> Suppose that an election is held, with a certain number of candidates in the field, each individual filing his list of preferences, and then one of the candidates dies. Surely the social choice should be made by taking each of the individual's preference lists, blotting out completely the dead candidate's name, and considering only the orderings of the remaining names in going through the procedure of determining the winner. That is, the choice to be made among the set S of surviving candidates should be independent of the preferences of individuals for candidates not in S.

What Arrow is describing is an instance in which the set of what we call actual candidates shrinks, as in C or C′. But if Arrow had meant something like C or C′, he would have said something like: 'The death of the candidate does not change the choice, except if the dead candidate was himself the choice.' But to the contrary he says here that

whatever the choice would have been before the death of the candidate, it should be forgotten, the name of the candidate should be blotted out, and the choice determined entirely by the voters' preferences over the new set of actual candidates.

One could dispute whether Arrow is correct in asserting that the information contained in the voters' rankings of the candidate who died *ought to be* ignored. It might be possible to use this information to make inferences about the intensities of preferences over the remaining candidates, and thereby improve the choice from among those who remained. *That, however, is a separate issue.* What Arrow is discussing is in any case again some form of IIA or regularity: the choice from a set of actual candidates is to be a function only of the preferences of the voters over this set of actual candidates, as is stated in succeeding sentences:

> Therefore, we may require of our social welfare function that the choice made by society from a given environment depend only on the orderings of individuals among the alternatives in that environment.

So up to this point, what Arrow had in mind was clearly regularity. However, in his model, it was impossible to give a formal statement of regularity: to do so, he would have had to build, as we have seen, a much more elaborate model. He did not see its usefulness, and so, having in mind both regularity and the theorem we stated in Section 3, he goes on:

> Alternatively stated, if we consider two sets of individual orderings such that, for each individual, his ordering of those particular alternatives in a given environment is the same each time, then we require that the choice made by society from that environment be the same when the individual values are given by the first set of orderings as they are when given by the second.

This is clearly an informal statement of IIA, which he states immediately afterward as Condition 3, this Condition 3 being formally equivalent to our Definition 2.

Up to this point, everything is consistent with Arrow having regularity/IIA in mind, and nothing is consistent with his having some form of C in mind. But now we get to the example. We quote the paragraph in full:

The reasonableness of this condition can be seen by consideration of the possible results in a method of choice that does not satisfy Condition 3, the rank-order method of voting frequently used in clubs. With a finite number of candidates, let each individual rank all the candidates, i.e., designate his first-choice candidate, second choice candidate, etc. Let pre-assigned weights be given to the first, second, etc. choices, the higher weight to the higher choice, and then let the candidate with the highest weighted sum of votes be elected. In particular, suppose that there are three voters and four candidates, x, y, z and w. Let the weights for the first, second, third, and fourth choice be 4, 3, 2, and 1, respectively. Suppose that individuals 1 and 2 rank the candidates in the order x, y, z, and w, while individual 3 ranks them in the order z, w, x, and y. Under the given electoral system, x is chosen. Then, certainly, if y is deleted from the ranks of the candidates, the system applied to the remaining candidates should yield the same result, especially since, in this case, y is inferior to x according to the tastes of every individual; but, if y is in fact deleted, the indicated electoral system would yield a tie between x and z.

Here, people who think that Arrow confused IIA with C, or that Arrow meant some form of C by IIA (which is not the same thing) will say, 'Aha! We've got you here. This example is clearly one of violation of C, and not of IIA or regularity.' Our answer is, 'Don't be so sure.'

The difficulty arises from the fact that there are two ways of understanding what Arrow meant by 'the rank-order method'. Let X, the set of potential candidates be finite. For simplicity, we will suppose that the voters cannot be indifferent between two candidates, that is, that for all i and all x and y $(x \neq y)$ we cannot have simultaneously xR_iy and yR_ix. (We could do away with this simplification, but then we would have to define more general rank-order methods that permit indifference, which would lead to useless complications.)

The first understanding of the rank-order method, which we will call the *global* rank-order method, consists of calculating the candidates' scores from the preferences of the voters over the whole of X. Then a subset A of actual candidates being given, the choice from A is the candidate(s) with the highest score.

The second understanding of the rank-order method, which we will call the *local* rank-order method, consists of calculating the candidates' scores from the restrictions of the voters' preferences to a subset A of actual candidates, the choice from A being the candidate(s) with the highest of these scores.

Since for each profile the global rank-order method generates a complete weak ordering over the whole of X which will be used to determine the choice over the set of actual candidates, the correspond-

ing SCF will satisfy C. On the other hand, this SCF does not satisfy IIA and the voting rule does not satisfy regularity.

The local rank-order method, on the other hand, does not satisfy C. Since to determine the choice over a set of actual candidates A, only the voters' preferences restricted to A are taken into account (no information about their preferences over the potential-but-not-actual candidates is used), it satisfies regularity and the SCF satisfies IIA.

So the question is: which of the two methods did Arrow have in mind, the global or the local? Everything points to the answer that it was the global method, which satisfies C but violates IIA and regularity.

First, Arrow's framework was not SCFs, but what he called 'Social Welfare Functions', that is, what we call, to avoid confusion with other concepts, Arrow functions. From all that precedes his discussion of IIA in his book, it is clear that for him the Arrow function, which yields a complete weak social ordering of X for each profile, comes *first*, the choices over 'environments', that is subsets of X, sets of actual candidates, coming *afterward*. The succession is: first we order the whole set of potential candidates, then we choose the set of actual candidates, and then we choose from among the actual candidates using the ordering of the set of all potential candidates that had been defined earlier. Now the global rank-order method, *not* the local one, corresponds to this succession. So to suppose that Arrow had in mind the local rank-order method in his example would be to suppose that this example is in contradiction with the rest of the book.

Second, at the time when Arrow was writing his book, it was not realized that there could be two different rank-order methods and that it was important to make a distinction between them. To the best of our knowledge, the first precise distinction between the global and local methods was in a paper presented by Sen at the Third World meeting of the Econometric Society in Toronto, in 1975 (published as Sen, 1977). So Arrow thought that he was using the global method, that violates IIA, and did not think that it could raise any problem.

Third and more important: it is true that one can interpret Arrow's example in such a way that it illustrates a violation of C. But that does not mean that it cannot be interpreted in such a way that it illustrates a violation of regularity/IIA. And it can be!

The key is this: suppose that X is finite and it happens that all the potential candidates are actual candidates, that is, $A = X$. The profile being given, a choice is made using 'the rank-order method'. Question: how can we know whether it was the global or local method that was used? The answer is that we cannot. In this case, both methods obviously give the same result, through the same reckoning.

So consider Arrow's example this way: in a first time, the set of potential candidates is $X = \{x, y, z, w\}$ and the set of actual candidates is $A = X$. Given the profile, the choice from A is x. In a second time, y has died and the set of potential candidates is now $Y = \{x, z, w\}$. Again all the potential candidates are actual candidates, and $B = Y$. With the 'same' profile, the choice from B is $\{x, z\}$.

First interpretation: the *local* rank-order method is used. X being the set of potential candidates, the set of actual candidates shrinks from A to B. The choice changes from x to $\{x, z\}$ in violation of C. Then the set of potential candidates shrinks from X to Y. Because of regularity, the choice from B does not change.

Second interpretation: the *global* rank-order method is used. X being the set of potential candidates, the set of actual candidates shrinks from A to B. Because of C, the choice from B is still x. Then the set of potential candidates shrinks from X to Y. The choice from B changes from x to $\{x, z\}$ in violation of regularity/IIA.

So while Arrow may be guilty of not having stated his example carefully enough (but with his model it was not obvious) he is certainly not guilty of having confused IIA with C, or of having meant IIA to contain some form of C.

In fact, by adding *five letters* (not one more!) to his example, Arrow would have saved everybody a lot of trouble. In the penultimate sentence, let us add 'over z' so that now it reads: 'Under the given electoral system, x is chosen *over* z'. This would identify implicitly the set of actual candidates as $\{x, z\}$ and the example becomes clearly an example of the violation of regularity/IIA by the global rank-order method. To convince yourself that this is so, just re-read the paragraph with the five letters added.

In case you are still not convinced, let us read a little further:

> The condition of independence of irrelevant alternatives implies that in a generalized sense all methods of social choice are of the type of voting.

And much later, on p. 110, in the chapter added to the 1963 edition, Arrow says about IIA:

> After all, every known electoral system satisfies this condition.

As said earlier, all 'real-world' voting rules satisfy regularity/IIA, but, *and this is precisely a consequence of Arrow's Theorem*, none satisfies C, or C′, or even weaker 'consistency' conditions.

So the case is closed. The verdict is: not guilty. There is not the shadow of a proof that Arrow confused IIA with C or something like C, or that he meant IIA to be something more than what is stated in his Condition 3, p. 27. To the contrary, there is substantial evidence that he did not.

5. THE MEANING OF THE ARROW THEOREM

Now, with the confusion over IIA cleared up, we can look at what the Arrow Theorem really means. First, we have to consider condition C, or more generally the 'consistency conditions'.

Consider the following example. The set $X = \{x, y, z\}$ is the set of potential candidates, and we consider two sets $A = X$ and $B = \{x, y\}$ of actual candidates. The voters' preferences are as follows:

for 45% of the voters: xP_iy, yP_iz

for 35% of the voters: yP_ix, xP_iz

for 20% of the voters: zP_iy, yP_ix

The vote takes place under the plurality rule. With this rule, if the set of actual candidates is A, then x is elected, but if the set of actual candidates is B, then y is elected. So the SCFs generated by the plurality rule violate C.

Let us take a closer look at this violation of C. z stands no chance of being elected, unless he is the only actual candidate. So we can consider that the two 'serious' candidates are x and y. But here, *who is it that chooses* whether it will be x or y that will be elected: the voters . . . *or* z? One can very well consider it to be z. Or suppose that

x, y and z are not 'flesh and blood' candidates, but alternative public works projects, the voters being the members of the legislative branch of the government of some country. x and y are two 'serious' competing projects, and according to the voters' preferences, y will pass with 55% of the votes. However, the executive branch is in favor of x. Since it has the possibility of proposing projects itself, it proposes z (for which it does not care) and insures that x will pass. The question is: who decided on the project? The legislative branch or the executive branch?

Let us consider another problem. One way to arrive at a choice is *sequential voting*. Instead of taking a vote on the whole set of actual candidates A, one compares a pair of elements of A, x_1 and x_2, using majority rule, and then compares the winner of this paired comparison with the third element of A, x_3, again using majority rule, and so on, the winner of the last paired comparison being the choice.

A voting rule satisfies the intraprofile condition called *Path-Independence* (PI) by Plott (1973) if:

- whatever the order of presentation of the candidates, the final choice is always the same; and
- this final choice is the same as the one that would have resulted from a vote on the whole of A.

PI is implied by C. That is, a violation of PI is a violation of C. If a voting rule does not satisfy PI it means that there are circumstances in which the entity that decides the order in which the candidates will be presented can manipulate this order to its advantage, and change the decision. In fact, the desirability of PI was alluded to by Arrow on the last page of the second edition of *Social Choice and Individual Values* (1963), as a justification for Axioms I and II.

So violation of C, or of PI and hence of C, implies that at least some of the power to make collective decisions lies not with the voters, where it is supposed to be, but rather with some agenda-setting entity or entities (which may be chance if, for example, the order of presentation of the voters is decided by drawing lots). So C is a condition (the strongest of such conditions – see Sen, 1977) to the effect that the choice cannot be 'manipulated' through 'strategic

candidacies', strategic construction of a sequence of paired comparisons, or similar maneuvers. This shows, by the way, that to ask a voting rule to yield SCFs satisfying C is not as unwarranted as some have claimed.

Now, with our terminology, the conditions that are employed in the 1963 classical version of the Arrow Theorem are the following:

- Universal Domain: the SCF is defined for every logically possible profile.
- IIA.
- Unanimity (or Pareto): if there are only two actual candidates x and y and if all the voters vote that they strictly prefer x to y, then x is chosen.
- Nondictatorship: there does not exist a voter (a 'dictator') such that whatever x and y, whatever the preferences of the other voters, if x and y are the only actual candidates and he strictly prefers x to y, then x is chosen.
- Condition C, that is, in Arrow, that the SCF derives from an Arrow function.

The Arrow Theorem states that these five conditions are logically inconsistent, that is, there cannot exist an SCF that satisfies all of them. Considered for voting rules, IIA and Universal Domain are conditions to the effect that the voting rule should be a 'real-world' voting rule. IIA we already discussed at length. As for Universal Domain, since one does not know, when defining a voting rule, the conditions and purposes for which it might in the future be used, it is desirable that it be capable of coping with any situation. The consequences of not satisfying the Universal Domain have been explored theoretically at great length, but virtually all real-world voting rules satisfy this condition. Unanimity and Nondictatorship are conditions to the effect that the voting rule should be 'reasonable'. We do not think it useful to comment at length on these. And last, we have just discussed condition C.

So the meaning of the Arrow Theorem is that all reasonable, real-world voting rules violate C. That is:

Every reasonable real-world voting rule is manipulable through strategic candidacies and similar maneuvers.

At least, this is the general meaning of the Arrow Theorem for the positive theory of voting.

One last word. Since the publication of Arrow's theorem, numerous 'Arrovian' results have been obtained. The beautiful theorem by Wilson (1972) characterizes the SCFs that simultaneously satisfy IIA (and hence represent possible real-world voting rules) and C (are 'non-manipulable' in the meaning we defined). Wilson's theorem shows that they are definitely unpalatable and unreasonable.

Other works explore the weakening of C, which is the strongest 'consistency' (non-manipulability) condition. For example, C implies PI, but the converse is not true, and Sen (1969) in fact showed that if C is replaced by PI in the five conditions above, the conditions are no longer incompatible: The 'unanamity' voting rules satisfy all of them. However, further works by Mas-Colell and Sonnenshein (1972), Blair, Bordes, Kelly and Suzumura (1976), and Bordes (1981) lead to the conclusion that the *only* 'reasonable' voting rules that satisfy the Arrovian conditions with C replaced by PI are the unanimity rules. That is, such rules are quite undiscriminating, since it is sufficient for two voters to have opposite preferences for a tie to occur. Other weakenings of C have been considered, but they usually result in an Arrovian impossibility (Sen, 1977). For a general overview see Kelly (1978) and Suzumura (1983).

NOTES

[1] There are some slight exceptions to this. It is possible for the set of candidates who have *actually been ranked* by voters to be *slightly* larger than the set of *actual candidates*. This would happen if a candidate were to die after the voters had ranked the candidates. It could also happen that individuals who were not actual candidates were included in a set to be ranked by voters, for the sake of having more information. Thus there are some voting rules that are affected by changes in voters' rankings of non-actual candidates (and hence violate IIA) that are potential real world, not fairy-world, voting rules. But it remains true that no real-world voting rule is or could be based on rankings of *all* candidates since this set is often not really defined. (Think for example of cases where the 'candidates' are public works projects: the set of potential candidates is the set of all

conceivable alternative public works projects!) This is in fact Fishburn's (1973, p. 8) main argument in favor of IIA.

[2] A 'voting rule' that selected the plurality winner among the actual candidates when the number of potential candidates was odd, and selected the plurality loser among the actual candidates when the number of potential candidates was even, would not be regular, although it would yield SCFs satisfying IIA.

REFERENCES

Arrow, K. J.: 1948, 'The Possibility of a Universal Social Welfare Function', Project RAND, RADL-289, 26 Oct. 1948, Santa Monica, Cal., hectographed.

Arrow, K. J.: 1951, *Social Choice and Individual Values*, Wiley, New York (first edition).

Arrow, K. J.: 1959, 'Rational Choice Functions and Orderings', *Economica* (New Series) **26**(102), 121–127.

Arrow, K. J.: 1963, *Social Choice and Individual Values*, Yale University Press, New Haven (second edition).

Blair, D. H., Bordes, G., Kelly, J. S., and Suzumura, K.: 1976, 'Impossibility Theorems without Collective Rationality', *Journal of Economic Theory* **13**(3), 361–379.

Blau, J. H.: 1957, 'The Existence of Social Choice Functions', *Econometrica* **25**(2), 302–313.

Bordes, G.: 1981, 'Procédures d'aggrégation et Fonctions de Choix', pp. 45–74 in Batteau, Jacquet-Lagreze and Monjardet (eds.), *Analyse et Aggrégation des Préférences*, Economica, Paris.

Chernoff, H.: 1954, 'Rational Selection of Decision Functions', *Econometrica* **22**(4), 422–433.

Dummett, M.: 1984, *Voting Procedures*, Clarendon Press, Oxford.

Fishburn, P.: 1973, *The Theory of Social Choice*, Princeton University Press, Princeton.

Kelly, J. S.: 1978, *Arrow Impossibility Theorems*, Academic Press, New York.

Mas-Colell, A. and Sonnenshein, H.: 1972, 'General Possibility Theorems for Group Decision', *Review of Economic Studies* **39**(2), 185–192.

Plott, C.: 1973, 'Path Independence, Rationality and Social Choice', *Econometrica* **41**(6), 1075–1091.

Radner, R. and Marschak, J.: 1954, 'Note on Some Proposed Decision Criteria', pp. 61–68 in Thrall, Coombs, and Davis (eds.), *Decision Processes*, Wiley, New York.

Ray, P.: 1973, 'Independence of Irrelevant Alternatives', *Econometrica* **41**(5), 987–991.

Sean, A. K.: 1969, 'Quasi-transitivity, Rational Choice and Collective Decision', *Review of Economic Studies* **36**(3), 381–393.

Sen, A. K.: 1977, 'Social Choice Theory: A Re-examination', *Econometrica* **45**(1), 53–89.

Suzumura, K.: 1983, *Rational Choice, Collective Decision, and Social Welfare*, Cambridge University Press, Cambridge.

Wilson, R.: 1972, 'Social Choice Theory Without the Pareto Principle', *Journal of Economic Theory* **5**(3), 478–486.

Georges Bordes
Lare (UA-CNRS N° 944),
Université de Bordeaux I,
Faculté des sciences économiques,
Avenue Léon Duguit-33604 Pessac,
France.

Nicolaus Tideman
Economics Department,
Virginia Polytechnic Institute and State University,
Blacksburg, VA 24061,
U.S.A.