JEAN-MARIE BLIN

HOW RELEVANT ARE 'IRRELEVANT' ALTERNATIVES?

ABSTRACT. Arrow's Independence of Irrelevant Alternatives Condition is examined. It is shown why the standard rationale for (or against) the condition tends to be inconclusive as it fails to consider the basic 'game' issue in social choice. Specifically it is explained how some recent results (Gibbard-Satterthwaite) on the general non-existence of strategy-proof voting procedures provide the strongest rationale for the independence condition. Also, it is shown that this rationale was exactly the one used by Condorcet in his work on decision rules for juries and elections.

I. INTRODUCTION

Choice between conflicting alternatives derives from preference over them. This proposition forms the basis of all theories of rational decision making. One may argue about what is the proper notion of 'rationality' one should use; but choice without preference appears as a logical inconsistency. Which choosing unit this principle applies to is very much an open question. In particular, if one interprets preference as a complete reflexive and transitive binary relation, i.e., a complete pre-ordering over the set of alternatives, it has been known since Condorcet [5], and formally proven by Arrow [2], that such a concept may not be applicable to rationalize group choices. This is so, of course, provided that one adopts the individualistic value judgment, for which preference can only be based upon some aggregation of individual preferences. By and large, western democratic ideals include some recognition of this value. It is hardly surprising, then, to find that the discovery of (what has come to be known as) 'Arrow's paradox' has generated so much debate and controversy in the social sciences. As Friedland and S. J. Cimbala [6] have pointed out, the paradox shares with all paradoxes a strange mixture of elusive, yet inescapable, truth capable of undermining the most basic foundations of a field. The challenge it presents has led to numerous further discoveries and refinements in social choice theory. Without reviewing some of the highlights of this line of research, it seems worthwhile to take a further look at a crucial condition in Arrow's theorem: the so-called 'independence of irrelevant alternatives' condition. Actually,

Theory and Decision 7 (1976) 95–105.

as we shall later argue, in one sense, a more accurate name for it would be 'independence of preferences for irrelevant alternatives'. The fact that Arrow's theorem crucially depends on that condition has made it a prime candidate for criticisms from those seeking to circumvent the paradox. Arrow, himself, recognized its key role in his rebuttal of some of these criticisms (see, in particular, pp. 109–120 in the second edition of *Social Choice and Individual Values*[1]). However, part of the problem stems from a simple misunderstanding, which Arrow himself helped create when he put forth as an alleged illustration an example which is actually unrelated to his condition (see pp. 27–28, op. cit.). Plott [17] and Hansson [13] both noticed the irrelevance of this example in arguing for (or against) Arrow's independence condition. The latter author, in particular, has conducted a very thorough investigation of the actual meaning of this condition and its role in Arrow's theorem and other impossibility theorems. Simultaneously, Ray [20] has clarified the relation between Arrow's independence condition and other similarly named conditions in the economics and psychology literature. (Radner and Marschak [19]; Luce [14].) The object of this paper is to extend the clarifications provided by these authors and cast a different light on the rationale for this condition by relating it to some recent results in social choice theory (Satterthwaite [21]). Additionally, this discussion places the independence condition in its historical context and, in particular, in the work of Condorcet. Here it will be seen that it represents a longstanding tradition, and that one of the strongest reasons for requiring it was specifically proposed by Condorcet.

II. THE INDEPENDENCE CONDITION

Arrow states his independence condition as follows:

> 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.[2] If, for all individuals i and all x and y in a given environment S, xR_iy if and only if xR'_iy, then $C(S)$ and $C'(S)$ are the same.

The following notation is used through our analysis: R_i are the individual preference pre-orderings $(i = 1, \ldots, n)$; R is a social preference pre-ordering; A is the set of all conceivable alternatives, (m of them); a generic element S of the power set of A, denoted $\mathscr{P}(A)$, is called an 'environment'. $\mathscr{R}$ is the set of all conceivable preference pre-orderings on

an m-element set (i.e. the set of all complete, reflexive and transitive binary relations on A). Then a social welfare function φ (or aggregation mapping) is:

$$(1) \qquad \varphi: \prod_{i=1}^{n} \mathscr{R}_i \to \mathscr{R}.$$

A social choice function[3] C is:

$$(2) \qquad C: \prod_{i=1}^{n} \mathscr{R}_i \times \{\mathscr{P}(A) \setminus \emptyset\} \to \{\mathscr{P}(A) \setminus \emptyset\}.$$

In words, for any given 'profile' $(R_1 \ldots, R_n) \in \prod_{i=1}^{n} \mathscr{R}_i$ and any (non-vacuous) environment $S \in \{\mathscr{P}(A) \setminus \phi\}$, C chooses a set of elements in $\{\mathscr{P}(A) \setminus \phi\}$ such that $C(S) \subset S$ and $C(S) \neq \phi$. The relationship between choice functions and social welfare functions (or more generally orderings) has been thoroughly studied by K. Arrow [1], [2] and B. Hansson [10]. It is clear that from a social welfare function, a social choice function can be derived by simply picking as $C(S)$ the R-maximal element(s) in S (R being the social pre-ordering determined by φ). One may also note that the social choice function concept is more general inasmuch as a choice function is not necessarily rationalizable by a complete transitive binary relation, even though the existence of such a relation is certainly sufficient to derive a social function (see C. Plott [18] for a thorough discussion of this point). In the sequel, however to keep the argument simple, we shall only consider social welfare functions.[4] In terms of such functions, Arrow's independence condition reads:

Let

$$(3) \qquad [R] = (R_1, \ldots, R_n) \quad \text{and} \quad [R'] = (R'_1, \ldots, R'_n)$$
$$xR_iy \leftrightarrow xR'_iy \to x\varphi([R])y \leftrightarrow x\varphi([R'])y \; \forall (x, y) \in A \times A.$$

In words, if the two profiles $[R]$ and $[R']$ agree on some pair (x, y), then the social outcome on (x, y) under φ should also be unchanged.[5] A change in the profiles of individual preferences on some other pair(s), say (w, t) – with w and/or $t \neq x$ and/or y – should not affect the social outcome on (x, y). Here, the issue is not the absence or presence of certain alternatives, but rather, what are the potential changes in individual preferences under which the social outcome is *invariant*. Of course, the issue of

whether a candidate is or is not in the race – e.g. if he dies after the votes have been taken, but before they are counted as in Arrow's example – does not matter. It is simply not the premise of Arrow's independence condition – unless one makes the individual preferences a function of the set of alternatives A. In such a case, of course, the presence or absence of certain alternatives would induce changes in the profiles of preference, and one would be back to the premise of the independence condition. Is it reasonable, though, to resort to such a roundabout argument to explain the apparent irrelevance of Arrow's example as an illustration of the independence condition? Is it not simpler to conclude that the example is just that, irrelevant? Until recently, such a conclusion appeared justified and the case could be dismissed. But, in light of both the context of Condorcet's discovery of the voting paradox and some very recent work by Gibbard [7] and Satterthwaite [21], a strong argument can be made for reopening the case. We first proceed to show an important implication of the independence condition.

III. INDEPENDENCE AND GENERALIZED MAJORITY VOTING

Let us consider, for a moment, the mathematical formalization of an Arrow-type social welfare function. If φ is such a mapping, then

$$(4) \qquad \varphi: \prod_{i=1}^{n} \mathscr{R}_i \to \mathscr{R} \quad \text{or} \quad \varphi(R_1, \ldots, R_n) = R,$$

where R is a typical image of an n-preference profile, i.e., a complete preordering on A. Now $R \subset A \times A$ and R is complete, reflexive and transitive. What if we look at the restriction of the mapping φ to a pair (x, y): $\varphi(R_1^{xy}, \ldots, R_n^{xy}) \equiv \varphi_{xy}$. There are $\binom{n}{2} = n(n-1)/2$ such restricted mappings φ_{xy} for $x, y = 1, 2, \ldots, n$. A natural question to raise is whether the image under φ of such an (x, y)-sub-profile is always that pair (x, y) itself, or some other pair. Clearly, Arrow's independence condition is simply a requirement that this always hold true for *any* pair. But, if this is so, then the overall image of any n-profile under φ can always be built up as the union of the sub-images $[\varphi_{xy}(\ldots), \ldots, \varphi_{wz}(\ldots)]$. This also means that, once the distribution of individual preferences on any pair, say (x, y), is given, the social outcome for that pair is also determined. This holds true

regardless of the distribution of preferences on other pairs – even if they involve x or y (but not both!). Arrow labels this property 'generalized majority voting' (pairwise). It should be clear that simple pairwise majority voting is a special case. A more general case of majority voting is that of 'qualitative' majorities; i.e., when no system of numerical weights on the individual votes can represent these majorities. Certain 'harmonious' groups are declared the winning groups ('decisive' in Arrow's terminology), whereas others are the losing groups.[6] What consistency requirements should be imposed on such families of winning groups is a separate issue, which is addressed by Arrow's other conditions (positive association, citizen's sovereignty and non-dictatorship). Actually, the link between the independence condition and generalized majority voting is clearly stated by Arrow:

> The condition of the independence of irrelevant alternative implies that in a generalized sense all methods of social choice are of the type of voting. If S is the set (x, y), this condition tells us that the choice between x and y is determined solely by the preferences of the members of the community as between x and y. That is, if we know which members of the community prefer x to y, which are indifferent and which prefer y to x, then we know what choice the community makes.

It is helpful to stress this fact more formally.

LEMMA: *A social welfare function φ is of the generalized majority voting type*[7] *if and only if it verifies the independence condition.*

Proof. (i) That the independence condition necessarily holds for a function of the generalized majority voting-type is clear.

(ii) The sufficiency part follows by an a contrario argument: if φ is not of the generalized majority voting type, there exists some $(x, y) \in A \times A$ such that $\varphi_{xy}^{-1}(\ldots) \neq (R_1^{xy}, \ldots, R_n^{xy})$. That is, for at least one individual i, R_i^S – where S is non-empty and $S \neq (x, y)$ – enters the inverse image of φ_{xy}; formally stated $\varphi_{xy}^{-1} = (R_1^{xy}, \ldots, R_i^S, \ldots, R_n^{xy})$. Then if we change the ith preference ordering on S, while leaving the (x, y) sub-profile unchanged for all i, the social outcome on (x, y) is affected contrary to the independence requirement. Q.E.D.

But, one might ask, this fact still does not affect our original finding, that Arrow's example bears no relation to his independence notion; whether or not some element x is in A is irrelevant. Or is it really? Only if we can safely assume that the composition of the set A does *not* affect the individual preferences R_i. In Arrow's work, of course, these prefer-

ences are normally taken as *fixed.* Whether, in his terminology, these preferences are a representation of an individual's 'tastes' or his 'values' (i.e. allowing for broader evaluation than the self-centered one dictated by his tastes) makes apparently little difference. But the point, here, is precisely that it may be unrealistic, if not impossible, to conceive of a situation where the R_i's are not themselves a function of A and, indirectly, of each other. This is, of course, nothing else but the gaming issue which Arrow himself recognized but decided not to address. In his words (p. 7 op. cit.):

> ... there still remains the problem of devising rules of the game so that individuals will actually express their true tastes even when they are acting rationally (...). In addition to ignoring game aspects of the problem of social choice we will also assume in the present study that individual values are taken as data and are not capable of being altered by the nature of the decision process itself.[8]

In recent contributions, this game aspect has been independently and successfully tackled by Gibbard [7] and Satterthwaite [21]. They call strategy-proof a voting procedure (a social choice function in our terminology) which is such that there never exists a profile of individual preferences where at least one individual would benefit from misrepresenting his true preferences. Here, benefit means that, judged by his true preferences, he can secure, through lying, a more favorable social outcome!

THEOREM (Gibbard-Satterthwaite): *If there are three or more alternatives, the only strategy-proof voting procedures are the dictatorial ones (in Arrow's sense).*

Satterthwaite also shows very clearly the link between this result and Arrow's theorem. Specifically, he shows that there exists a one-to-one correspondence between strategy-proof voting procedures and social welfare functions verifying the Pareto and Independence condition.

This fact sheds a rather different light on the independence condition and the rationale for requiring it, since one can hardly argue against the desirability of strategy-proofness as a property of a social choice mechanism. It would be interesting to explore the implicit fairness notion which underlies such a judgment.[9] Such a line of research will not be pursued here. But this strong rationale for the independence condition remains. Thus it is interesting to see how Arrow's example can be interpreted in

light of these findings. This is the case of the election where the presence or absence of a candidate affects the ranking of two other candidates (pp. 26–27, op. cit).

Now if this were to happen, it would be a very clear example of lack of strategy-proofness of the process and an unmistakable cue to astute voters. As a matter of fact, in non-majority voting election systems, sophisticated strategies are often laid out by political parties to take advantage of such facts. Negotiations to avoid third-party candidates in two-party systems, platform adjustments and various log-rolling processes all stem from a recognition of this fact. If, on the other hand, the outcome on any pair only depends on what individual preferences are for *that pair*, regardless of any other alternative, then one can intuitively see how immune to strategies such a system would be. The Gibbard-Satterthwaite result formally demonstrates this fact. Along the same lines of reasoning, it is worthwhile to give an example adapted from Luce and Raiffa [15], which argues against the independence condition. In a strict sense, as in Arrow's example, it seems unrelated to the condition. Briefly stated it considers the plausibility of the following individual choice behavior: on a dinner menu featuring steak and spaghetti, and not knowing anything about the restaurant, a safe strategy might be to state a preference for steak over spaghetti; but if one is then told that lobster is also available, it is perfectly reasonable to interpret this as a sign of overall quality of the restaurant and decide to state a preference for spaghetti over steak (assuming that this happens to be the actual preference of this individual). This example illustrates very clearly the relationship between the composition of the alternative set and the preferences over that set. For informational (as in this latter case) or strategic considerations (as in the case of elections), 'irrelevant' alternatives may actually be of utmost relevance! They are relevant whenever the decision process is susceptible of strategic manipulation by the individuals who state their preferences. If it is, then their preferences will be a function of the available alternatives and change as these alternatives vary.

In conclusion and to emphasize these remarks further, we now take a brief look at the origin of the independence condition in Condorcet's work. It will then be seen how the strategy-proofness issue was Condorcet's prime motivation for requiring independence. This will also clarify a rather fundamental point in the history of economic thought.

IV. CONDORCET VS BORDA: HONEST VOTERS OR INDEPENDENCE OF IRRELEVANT ALTERNATIVES?

The natural starting point here is the work of Borda [4] on election systems. Dissatisfied with the loss of information about individual preferences when plurality voting is used, he proposed to assign scores $(n-1)$ through 0 from the top-ranked to the bottom-ranked alternative, to sum these individual scores and use the order of these collective scores as the group order. It is well-known that the Borda count method is only one of a multitude of methods sometimes referred to as 'positionalist voting functions' (Gärdenfors [8]), or 'scoring functions' (Young [23]). Without reviewing these procedures, it suffices to say that they are extremely sensitive to changes in the individual scores (witness Arrow's example) and thus make the individual play of strategies (i.e. preference misrevelation) very likely. After Borda presented his memoir to the members of the French Academy of Science, his method was adopted for elections and other group decisions at the Academy. Condorcet, as an Academy member, became aware of this procedure and soon pointed out how easily manipulable the Borda count was. Borda conceded this defect but retorted 'My system is only for honest men'!

Rather than blindly trust voters and rely on some sense of civic duty, Condorcet set himself the task of finding a system which would be both informationally efficient, i.e. use all the information contained in the individual preference orderings (as the Borda count does) *and* non-manipulable. Condorcet expressed this notion of non-manipulability by seeking a system that would discover the 'true' social preferences. All his lengthy discussion which deals with individual probabilities of discerning the 'truth' and aggregating them through a process which is most likely to reflect this 'truth', can be readily interpreted from this viewpoint.[10] He also connected it with his work on decision rules for juries. Far from there being some sort of discontinuity in his work, when he then proceeds to discover the voting paradox as some commentators have sometimes felt, the transition is obvious (if one remembers his often-repeated desire for a non-manipulable scheme – unlike the Borda count). His argument can be paraphrased as follows: when I state a preference ordering as $x > y > z$ ($> \equiv$ 'preferred to'), I actually make three simultaneous 'elementary statements' (as he puts it): $x > y$, $y > z$ and $x > z$. In modern terminology

this is simply saying that a preference ordering is in fact a binary relation, i.e., a subset of $A \times A$. When an election is held between two candidates and plurality voting is used, a 55% majority for x over y *also means* a 45% minority for y over x. No confusion is possible since no other alternative is being considered. This feature should then be preserved by the election procedure he is looking for: the social outcome on (x, y) should *only* depend on individual preferences on *that pair*, no matter what the preferences may be on other pairs. This is exactly what the independence condition formally requires. The surprising point, however, is that, since Condorcet's work was rediscovered after Arrow's (and Arrow himself acknowledges being unaware of it at the time of his writing), no one seems to have examined Condorcet's rationale for requiring what is simply the independence condition. A close look at the context of Condorcet's work (especially in response to Borda, as we explained earlier) affords an insight into it. Almost two centuries later and starting from what Arrow himself thought was an entirely separate issue – the 'game view' of social choice (as he puts it) – Gibbard and Satterthwaite have established what Condorcet intuitively felt: to be strategy-proof, a voting procedure must verify the independence of irrelevant alternatives. If not, then other alternatives become very relevant since they can be used for individual preference misrepresentation.

Graduate School of Management
Northwestern University

NOTES

[1] John Wiley and Sons, New York, 1963.

[2] To be perfectly accurate, this should read 'let $C(S)$ and $C'(S)$ be the corresponding *values* of the social choice function'. This is so because $(R_1 \ldots R_n)$ and $(R'_1 \ldots R'_n)$ are two points in the domain of the social welfare function, and we are not considering a change in the social welfare mapping, but simply the *two images of* these points *under a given mapping*.

[3] This terminology is not uniform; Hansson [13] refers to 'group decision functions' and Satterthwaite [21] speaks of 'voting rules'.

[4] B. Hansson [13] provides a very complete treatment of both concepts from the standpoint of the independence condition.

[5] As we consider social welfare functions – and not social choice function – we need not consider environments larger than two-element sets; for a choice function rationalizable by a pre-ordering, we only need consider the two-element set. See Arrow [2] Lemma 2, p. 16 on this point.

[6] From this notion it is easily shown that the Arrow type social welfare functions can be described in the language of n-person simple games (as in Wilson [22] or Montjardet [16]). This approach also leads to a constructive proof of Arrow's theorem based on the theory of filters on a set (see, for instance, Hansson [12]).
[7] These functions are sometimes referred to as 'majority-type aggregation functions' (see, for instance, Guilbaud [9], and Montjardet [16]).
[8] This last point also bears on our previous remark since if the decision process calls for sequential elimination of the alternatives, this would be another reason for making the R_i's functionally dependent on A.
[9] In particular, it would be interesting to relate it to the Rawlsian notion of justice.
[10] For a discussion of the relationship between majority voting and the maximum likelihood statistical estimation method see [3].

REFERENCES

[1] Arrow, K. J.: 1959, 'Rational Choice Functions and Orderings', *Economica* **26** 121–7.
[2] Arrow, K. J.: 1963, *Social Choice and Individual Values*, John Wiley and Sons, New York, 2nd ed.
[3] Blin, J. M.: 1973, 'Preference Aggregation and Statistical Estimation', *Theory and Decision* **4**, 65–84.
[4] Borda, Jean-Charles de: 1781, 'Mémoire sur les élections au scrutin', *Histoire de l'académie royale des sciences.*
[5] Condorcet, Marquis de: 1785, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris.
[6] Friedland, E. I. and Cumbala, S. J.: 1973, 'Process and Paradox: The Significance of Arrow's Theorem', *Theory and Decision* **4**, 51–64.
[7] Gärdenfors, P.: 1973, 'Positionalist Voting Functions', *Theory and Decision* **4**, 1–24.
[8] Gibbard, A.: 1973, 'Manipulation of Voting Schemes: A General Result', *Econometrica* **41**, 587–601.
[9] Guilbaud, G. Th.: 1952, 'Les théories de l'intérêt général et le problème logique de l'aggrégation', *Economie Appliquée* **5**, 501–584.
[10] Hansson, B.: 1968, 'Choice Structures and Preference Relations', *Synthese* **18**, 443–458.
[11] Hansson, B.: 1969, 'Voting and Group Decision Functions', *Synthese* **20**, 526–537.
[12] Hansson, B.: 1973, 'The Existence of Group Preference Functions', Working Paper, Department of Philosophy, University of Lund.
[13] Hansson, B.: 1973, 'The Independence Condition in the Theory of Social Choice', *Theory and Decision* **4**, 25–50.
[14] Luce, R. D.: 1958, *Individual Choice Behavior*, John Wiley and Sons, New York.
[15] Luce, R. D. and Raiffa, H.: 1967, *Games and Decisions*, John Wiley and Sons, New York.
[16] Montjardet, B.: 1974, 'Théorie axiomatique de l'agrégation des tournois', Working Paper, University of Paris.
[17] Plott, C. R.: 1973, 'Some Recent Results in the Theory of Voting', in M. Intrilligator (ed.), *Frontiers of Quantitative Economics*, North Holland.

[18] Plott, C. R.: 1973, 'Path Independence, Rationality and Social Choice', *Econometrica* **41**, 1075–91.
[19] Radner, R. and Marschak, J.: 1954, 'Note on Some Proposed Decision Criteria', in R. M. Thrall, C. H. Coombs and R. L. Davis (eds.), *Decision Process*, John Wiley and Sons, New York.
[20] Ray, P.: 1973, 'Independence of Irrelevant Alternatives', *Econometrica* **41**, 987–991.
[21] Satterthwaite, M.: 1974, 'Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions', *J. Economic Theory* **10**, 187–217.
[22] Wilson, R.: 1972, 'The Game-Theoretic Structure of Arrow's Theorem', *J. Economic Theory* **5**, 14–20.
[23] Young, P.: 1974, 'An Axiomatization of Borda's Rule', *J. Economic Theory* **9**, 43–52.