

SOCIAL CHOICE THEORY VERSUS THE PEOPLE

N.M.L. Nathan

K.J. Arrow and other have shown that no method of aggregating individual preference rankings into a social ranking can satisfy every member of any one of various sets of individually reasonable-looking conditions.¹ It is sometimes thought that Arrow-type impossibility results should alarm the democrat even if they can be accepted with equanimity by the welfare economist. The results apply only to methods of aggregation which do not use information about interpersonal comparisons of utility; and although such information is likely to be used in the aggregation exercises of a welfare economist, it is not necessarily relevant to the democrat who wants merely to aggregate individual political judgments. A democrat is not likely to want voting rules to take into account the relative welfare levels of different voters, and will not necessarily want to assume anything at all about whether or not every voter who prefers one alternative to another prefers it with equal intensity. It looks then as if the democrat may be thrown back on those informationally deprived methods of aggregating preference rankings to which Arrow-type impossibility results really do apply. There is admittedly nothing here to alarm those who, with Nietzsche, see democratic institutions as mere 'quarantine measures against that ancient plague, the lust for power . . . very necessary, and very boring'. From the power-quarantine point of view any non-dictatorial voting rule is as good as another, so long as it is regularly employed to secure a turnover of governments. But a populist democrat, who sees a positive value in some systematic correspondence between government policy and the set of individual political judgments, must also be concerned with the intrinsic reasonableness of electoral rules. Arrow-type results have been taken to show that no electoral rules are reasonable, and hence to undermine the whole theory of populist democracy. The full Arrovian case against populist democracy has recently been

¹ K.J. Arrow, *Social Choice and Individual Values* (2nd ed, New York, 1963). A.K. Sen provides a good critical survey of subsequent results in 'Social Choice Theory: a Re-examination', *Econometrica* 45 (1977) 53-89, reprinted in *Choice, Welfare and Measurement* (Oxford 1982)

elaborated with great force and skill by William H. Riker, in his sombre study *Liberalism against Populism*.²

In what follows I argue that one very important group of Arrow-type impossibility results should not really alarm the populist democrat at all. The group consists of all those impossibility theorems whose sets of simultaneously unsatisfiable conditions contain Independence of Irrelevant Alternatives. I will try to show that the Independence condition is not one that the populist has any good reason to want electoral rules or voting methods to satisfy. Arrow-type theorems, whose sets of conditions do contain Independence, include not only Arrow's own General Possibility Theorem, which applies to methods of aggregation yielding social orderings, but also various theorems which apply to methods of aggregation which yield social rankings satisfying less stringent consistency conditions. Mas-Colell, Sonnenschein, Blau and Deb have proved impossibility theorems about methods yielding social rankings in which strict preference is treated as an acyclical rather than a transitive relation, and Bordes has proved a theorem about methods yielding rankings in which strict preference need satisfy only the condition of acyclicity over triples.³ These theorems have been taken to show that Arrowian gloom should extend even to those populist democrats who do not expect their electoral rules to satisfy Arrow's own initial stringent requirements for collective rationality. But on my argument the inference is unjustified, because even these theorems exploit an Independence condition which the populist need not want voting rules to satisfy. We can certainly find Arrow-type results which are Independence-

² (San Francisco 1982). Riker's case against populist democracy invokes theorems about strategic voting and agenda control as well as more familiar Arrow-type results. On the irrelevance of Arrow-type results to welfare economics see A.K. Sen, 'Personal Utilities and Public Judgements: or What's Wrong with Welfare Economics?', *Economic Journal* 89 (1979) 537-58.

³ Arrow's General Possibility Theorem is that no method of aggregating the preference orderings of two or more individuals over a set X of three or more alternatives can yield a social ordering of those alternatives and satisfy all of the following four conditions. U (Unrestricted Domain): the method is applicable to all logically possible n -tuples of individual preference orderings on X ; P (Weak Pareto Principle): For any x, y , in X , if everyone strictly prefers x to y , then x is strictly preferred to y in the social ranking; D (Non-dictatorship): There is no person whose strict preference over any pair $[x, y]$ is invariably reflected in social strict preference; I (Independence of Irrelevant Alternatives): The social ranking of any pair of alternatives in X depends only on individual rankings of that pair, and not on individual rankings of any other pair. The preference relation R (strict preference or indifference) gives an ordering if it is taken as transitive as well as reflexive and complete. Aggregation methods yielding social orderings from individual orderings are called SWFs and distinguished from SDFs which yield, from individual orderings, social rankings in which, in every subset of the alternatives, there

free. One of them is Sen's proof of the Impossibility of a Paretian liberal.⁴ Another is Ferejohn and Grether's theorem to the effect that extensions of majority rule methods which satisfy an expansion-consistency condition violate a version of Pareto rationality.⁵ I shall not consider how alarmed the populist democrat should be by Independence-free impossibility results.

The Independence of Irrelevant Alternatives condition is that the social ranking of any pair of alternatives depends only on individual rankings of that pair, and not on individual rankings of any other pair. In other words, if on two occasions when a method of aggregation is applied, there is some pair of alternatives such that no individual changes his rankings of that pair, then the method of aggregation does not yield a different social ranking of that pair on the two occasions. The Independence condition bans changes in social rankings when these changes are not correlated in a certain way with changes in individual rankings. Let us call the kind of change banned by the Independence condition Irrelevantly Dependent Change, and consider an actual example of Irrelevantly Dependent Change, in the normal symbolic representation. In this example, given below, we have two elections in which the same three voters A, B, C order the same three alternatives x, y, z . In each election the social ranking is derived from the individual orderings by the same rule. One rule which would yield these particular social rankings from these particular individual orderings is the Borda method. Following this, we give three points to each voter's first preference, two to his second preference

is an alternative to which no other alternative is strictly preferred. R will yield that kind of ranking if it is reflexive, complete, and the transitivity of strict preference is replaced by acyclicity, i.e. by the condition that if there is a finite sequence $x_1 P x_2, x_2 P x_3, \dots, x_{n-1} P x_n$, then not $x_n P x_1$. Acyclicity over triples is the condition that there is no triple x, y, z such that $x P y, y P z, z P x$. A. Mas-Colell and H. Sonnenschein proved that for any SDF satisfying conditions U, I, P and PR, if there are at least four individuals then someone has a veto. ('General Possibility Theorem for Group Decisions', *Review of Economic Studies* 39 (1972) 185-92) PR (Positive Responsiveness) is the condition that if x goes up in somebody's preference and does not fall in anyone's preference, then provided that x was originally at least weakly preferred to y on the social ordering, it becomes strictly preferred to y on the social ordering. N has a veto if for any pair x, y , such that N strictly prefers x to y , x is weakly preferred to y in the social ordering. Blau and Deb, and Bordes also produced veto results, with certain restrictions on the number of individuals. For details and references see A.K. Sen, 'Social Choice Theory: A Re-examination'.

⁴ A.K. Sen, 'The Impossibility of a Paretian Liberal', *Journal of Political Economy* 78 (1970) 152-7. (Reprinted in F. Hahn and M. Hollis (eds), *Philosophy and Economic Theory* (Oxford 1979))

⁵ John A. Ferejohn and David M. Grether, 'Weak Path Independence', *Journal of Economic Theory* 14 (1977) 19-31

and one to his third preference, and construct the social ranking according to the total number of points which each alternative then receives. The change in the social ranking from $xIzPy$ is Irrelevantly Dependent because although there is a change in the social ranking of x and z , no individual voter has changed his ranking of x and z .

Election E_1

$$xP_A yP_A z$$

$$zP_B xP_B y$$

$$zP_C xP_C y$$

Social ranking: $xIzPy$

Election E_2

$$xP_A yP_A z$$

$$zP_B yP_B x$$

$$zP_C xP_C y$$

Social ranking: $zPxPy$

How, more exactly, are we to interpret the symbolism in this example? Take the formulae $xP_B y$ and $yP_B x$, which occur under the E_1 and E_2 headings. Do they (α) just describe the patterns of marks which B makes on his ballot papers in the two elections? Do they (β) also give some such brief description of B 's mental states as that at t_1 , the time of E_1 , he prefers the election of candidate x to candidate y , and that at time t_2 , the time of E_2 , he prefers the election of candidate y to candidate x ? Or do they (γ) both describe the pattern of marks that B makes and also give some such fuller description of B 's mental states as that at t_1 he prefers what he believes at t_1 that candidate x will do after t_1 to what he believes at t_1 that candidate y will do after t_1 , and that at t_2 he prefers what he believes at t_2 that candidate y will do after t_2 to what he believes at t_2 that candidate x will do after t_2 ? There are then three methods of interpreting the individual preference symbolism. I will try to show, from the E_1 — E_2 example, that whichever method of interpretation we choose, α , β or γ , there is no good reason why a populist democrat should want his voting rule to satisfy the Independence Condition.

It is logically possible for Irrelevantly Dependent Change to result from successive applications of a voting rule if and only if that rule does not satisfy the Independence Condition. And one is liable to conclude from this that it is reasonable for a democrat

who dislikes Irrelevantly Dependent Change to want his voting rule to satisfy the Independence Condition. But in fact this does not follow. It is one thing for it to be logically possible for successive applications of a voting rule to produce Irrelevantly Dependent Change, another for there to be some serious likelihood of this result. And if there is no serious likelihood that Irrelevantly Dependent Change will result from successive applications of a voting rule which does not satisfy the Independence Condition, then it is not necessary for someone who dislikes Irrelevantly Dependent Change to insist that his voting rule satisfies the Independence Condition. He need not concern himself about the conditions his voting rule would need to satisfy in logically possible worlds which he does not believe are at all likely to be actual. Furthermore, he will have a positive reason for *not* wanting his rule to satisfy the Independence Condition, if he also knows that, as the social choice theorists tell us, voting rules which satisfy the Independence Condition cannot also satisfy all of various other reasonable-looking conditions. With this much accepted, and with the help of our distinction between α -, β -, and γ -interpretations of the preference symbolism, we can now construct a dilemma. For the populist to be reasonable in wanting a voting rule to satisfy the Independence Condition, two things are necessary: (i) it must be reasonable for him to dislike Irrelevantly Dependent Change; (ii) there must be a serious likelihood that successive applications of voting rules which do not satisfy the Independence Condition will result in Irrelevantly Dependent Change. But whichever interpretation of the preference symbolism we take, either (i) or (ii) is false. On the α - and β -interpretations, the populist has no reason to dislike Irrelevantly Dependent Change. On a γ -interpretation, there is no serious likelihood that Irrelevantly Dependent Change will result from successive applications of a voting rule which does not satisfy the Independence Condition.

Suppose we interpret the symbolism by the α -method. Then Irrelevantly Dependent Change becomes a kind of change in social rankings, which is not correlated in a certain way with changes in the marks which are made on ballot papers. Has the populist any good reason to dislike α -defined Irrelevantly Dependent Change? He wants social rankings to reflect what the voters really think. But why should he suppose that $xIzPy$ and $zPxPy$ are worse, as reflections of what the people really think at times t_1 and t_2 , than any other pair of social rankings, given that the α -interpreted E_1 — E_2 data tell us nothing about what the people think apart from

what we can infer from the marks they make on paper? It is perfectly consistent with the α -interpreted E_1-E_2 data that although at t_2 B marks candidate y above x on his ballot paper, thus changing the t_1 relation between his x and y markings, what B actually believes at t_2 that x will do after t_2 is quite different from what he believed at t_1 that x will do after t_1 . But then, even if we can infer from B 's markings of y above x at t_2 that he prefers at t_2 what he believes at t_2 that y will do after t_2 to what he believes at t_2 that x will do after t_2 , we will not be able to assume that he prefers at t_2 what he believes at t_2 that y will do after t_2 to what he believed at t_1 that x would do after t_1 . We will not be able to assume that two identical things have changed places in B 's estimation, rather than that he successively estimates two different pairs of things. Now if, in the E_1-E_2 case, there really is a change in B 's beliefs about what x will do in future, so that the change in his markings of x and y does not signify that two identical things have changed places in his estimation, then there will be nothing particularly objectionable from a populist point of view about the change in the social ordering from $xIzPy$ to $zPxPy$. And since the α -interpreted E_1-E_2 data do not tell the populist whether or not this is what has really happened, he is in no position to object to the kind of Irrelevantly Dependent Change whose occurrence they do inform him of. (We could reinforce the point by bringing in other possibilities which the α -interpreted E_1-E_2 data are also too meagre to rule out. For example, what B believes at t_2 that x will do after t_2 could be quite different from what A and C believe at t_2 that x will do after t_2 , and although A and C both mark x above y at t_2 they also both prefer what they believe at t_2 that y will do after t_2 to what B believes at t_2 that x will do after t_2 .)

Has the populist any good reason to dislike β -interpreted Irrelevantly Dependent Change? The β -interpretation also give him insufficient information about individuals. It follows from the β -interpreted E_1-E_2 data that at t_2 , unlike t_1 , B prefers the election of candidate y to candidate x . But this again fails to tell us whether two identical things have changed places in B 's estimation, or whether B is successively estimating two different pairs of things. 'At t_2 , unlike t_1 , B prefers the election of y to the election of x ' tells us nothing about whether what B believes at t_2 that x will do after t_2 is the same as what B believes at t_1 , that x will do after t_1 . If there is not this stability of belief, then it is not necessarily the case that two identical things have changed

places in B 's estimation, and the populist is once again unable to disapprove of the change in the social ordering from $xIzPy$ to $zPxPy$.

Suppose finally that we interpret the E_1-E_2 data in a γ way. Then we rule out at least some of those possibilities about hidden changes in the attitudes of the voters which prevent a populist confined to α - or β -interpretations from being confident that there is anything wrong with the changes in the social orderings. On a γ -interpretation, a formula like $xP_B y$ means not just that the voter marks a ballot paper in a certain way at a certain time, nor even just that in addition to this he has a preference at that time for x 's election over y 's election. It means that at a certain time he prefers what he believes at that time that x will do after that time to what he believes at that time that y will do after that time. If we are told that in election E_1 $xP_B y$ and in election E_2 $yP_B x$, then on a γ -interpretation what B prefers to y at t_1 is the same as what he prefers y to at t_2 . If there is any interpretation of the preference symbolism on which it is reasonable for the populist to dislike Irrelevantly Dependent Change, then it will be a γ -interpretation.

If I can now show that, on a γ -interpretation, there is no serious likelihood that successive applications of a voting rule will result in Irrelevantly Dependent Change, then, given my previous argument, it will follow that there is no reason for the populist to want his voting rule to satisfy the Independence Condition. And it is in fact quite easy to show that, on the γ -interpretation, Irrelevantly Dependent Change is very unlikely to occur. However we interpret the preference symbolism, Irrelevantly Dependent Change only occurs between a pair of elections if in both elections precisely the same alternatives are being ranked by all voters. That might happen if, as on the α - or β -interpretations, ranking alternatives is just making certain kinds of marks on paper, or just doing this and being in any mental state which can be described as preferring one option to another. But if ranking alternatives is interpreted in a γ -way, then it is utterly unlikely that precisely the same alternatives will be ranked by all the voters in any pair of elections. Consider the E_1-E_2 pair. If a voter believes at t_1 that candidate x will do such-and-such after t_1 , and believes at t_2 that x will do such-and-such after t_2 , then the different time references in the contents of his two beliefs might by themselves seem sufficient to guarantee that for this voter, it is not the case that precisely the same alternatives are at issue in the two elections. But even if we waive this point, it will still be quite unrealistic to suppose that for *each*

voter and *each* candidate the such-and-such that the voter believes at t_2 that the candidate will do after t_2 is the same as the such-and-such that he believes at t_1 that that candidate will do after t_1 . If there is time for B to change his preferences between E_1 and E_2 , then there is time for the world to change in a way which alters the beliefs of at least one voter about what at least one candidate will do if elected. So unless highly implausible conditions for the general stability of beliefs are satisfied, no real historical sequence will be represented by a γ -interpretation of our $E_1 - E_2$ data. And there is a similar unreality about every other γ -interpreted description of Irrelevantly Dependent Change. It would for example make no difference if it were pure policies rather than candidates for office which were being voted on in a pair of elections: there is every likelihood that between the elections somebody's belief will change about how one of the policies will work out in the future.

Tidily modelling electoral rules on functions whose domains consist of n -tuples of individual orderings of fixed and unchanging sets of alternatives, the social choice theorist invites us to agree that electoral rules should satisfy Independence, a condition which postulates successive applications of rules. But nobody would want an electoral rule to satisfy the Independence condition unless he disliked Irrelevantly Dependent Change, and if the fixed alternatives model is to be realistic individual rankings of the alternatives must be understood in such a way that Irrelevantly Dependent Change is not something he has any reason to dislike. If on the other hand we interpret individual rankings of alternatives so that a fixed alternatives model is unrealistic, then it is equally unrealistic to suppose that Irrelevantly Dependent Change is at all likely to occur, and therefore unnecessary to make it logically impossible for a rule to produce it by requiring the rule to satisfy the Independence Condition.

