Articles 2

"SOCIAL PREFERENCE" AND GAME THEORY: A COMMENT ON "THE DILEMMA OF A PARETIAN LIBERAL"

Nicholas R. Miller*

In "The Dilemma of a Paretian Liberal," John Aldrich argues persuasively that Amartya Sen's theorem concerning "The Impossibility of a Paretian Liberal" (1970) has wide-ranging consequences and is intimately connected with a number of well-known theoretical anomolies. In the main, Aldrich's arguments strike me as quite reasonable, and certainly they are of considerable interest. However, it does appear to me that his article contains several ambiguities, and I want here to address what seems to me to be the most basic one.

Sen's theorem (like Arrow's) is cast in the framework of abstract social choice theory that Aldrich sketches out on p. 1. Such theory is concerned with formal relationships between n-tuples of individual preference orderings and a "social preference relationship"; it is not directly concerned with the mechanisms, institutions, or processes that generate such "social preferences." But Aldrich's examples of applications of Sen's theorem do involve such processes and, accordingly and appropriately, are (or can be) cast in the framework of game theory. Such theory does specify a process for generating "social choice" by assigning strategic

*Department of Political Science, University of Maryland, Baltimore County, 5401 Wilkens Avenue, Baltimore, Maryland 21228.

opportunities to players who have preferences over outcomes and who are assumed to act in some rational fashion.

Now the question which seems to me to be at best only ambiguously addressed in Aldrich's article is this: how precisely do we connect these two theoretical frameworks – what exactly does "social preference" mean in a game-theoretical context. In this comment, I try to sketch out some alternative answers and note some problems with each.

As a preliminary matter, we must consider what is meant by an "outcome" in game theory (i.e., the object that corresponds to an "alternative," "social state," etc., in social choice theory). Let us follow Robin Farquharson (1969, p. 21) and call each strategy n-tuple (cell in the matrix representing the game in normal form) a *situation*. Now the question is whether every situation belongs to a distinct "outcome"? On p. 13 Aldrich suggests as much: "An 'outcome' is defined in game theory as different from another if it has a different 'history'." Apparently this means that each distinct "play" (path from the initial point to an end point in the tree representing the game in extensive form) leads, by definition, to a different "outcome." But, even so, two situations may entail the same play and thus the same outcome, since they may differ only with respect to choices they prescribe at information sets that are never reached in either situation. Beyond this, it seems natural to say that two situations, even if they entail different plays, *may* belong to the same outcome, if all players (and interested bystanders) are indifferent between them.[1]

The connection between abstract social choice theory and cooperative game theory (in which it is assumed players can make binding commitments) has been investigated in several recent articles. (See especially Wilson, 1972; Bloomfield, 1976; and the relevant sections in Plott, 1976.) Roughly "social preference" is equated with "domination." Classical (von Neumann-Morgenstern) theory assumed the feasibility of side payments; thus domination was defined over "payoff vectors" or "imputations," not over outcomes per se. We can, however, adapt the concept of domination to the (more relevant) case of games without side payments. The general approach in the articles referred to is to specify "rules of the game" that directly empower certain coalitions to make certain "social choices." Thus, "*a* is socially preferred to *b*" means "*a dominates b*" which in turn means that there is some coalition *S* all of whose members prefer *a* to *b* and which "has the power to choose *a* rather than *b*" (or "is decisive for *a* over *b*"). In the special case in which *S* includes a single player, we may have a kind of "liberalism" condition; and in the special case in which *S* includes all players, the Pareto principle is relevant.

[1]Though Aldrich cites Luce and Raiffa (1957), pp. 43-44, to support his view, I read their discussion as supporting the opposite view. Farquharson (1969, p. 6) identifies an outcome as the adoption of a particular proposal; thus many situations belong to the same outcome. Of course I do not dispute Aldrich's point that a player may care *how* a proposal is adopted, e.g., by what margin, whether he votes for or against or abstains (see Mayhew, 1974, especially p. 115, for empirical illustration); I dispute only the suggestion that this must be the case.

But what does "the power to choose a rather than b" (or "decisive for a over b") actually mean, when we start, as Aldrich generally does, with the game in normal form? Most obviously, it may mean that the coalition S has the power to impose a as the realized outcome, i.e., that there is a set of strategies, one for each member of S, such that every situation including these strategies belongs to the same outcome a[2]; thus a will be the realized outcome whatever the players not in S do. (Notice that such a coalition then has the power to choose a rather than any other outcome, not just a rather than b; notice also that the power to choose a rather than b need not entail the power to choose b rather than a.) In a majority voting game (and defining an outcome as adoption of a given proposal), for example, all domination comes about in this fashion. But in many other games, it may be that no coalition, other than the coalition of the whole, has such "power to impose." For example, in the Prisoners' Dilemma neither player can impose any outcome. Thus, if domination could come about only in this manner, only unanimity brings about domination, and "social preference" is as shown in Figure 1, where $a \xrightarrow{S} b$ means "a dominates b through coalition S" and where the outcomes are labelled as in Aldrich's Tables 4A and 4B. Clearly, if "social preference" is equated with domination, and if domination can come about only in this manner, the Prisoners' Dilemma does not illustrate the "liberal paradox" – indeed it does not illustrate "liberalism" at all.

More basically however, "(minimal) liberalism" is impossible by itself (and not just impossible in conjunction with other conditions such as U, P, and SDF) under this interpretation, for clearly two disjoint coalitions (including two individual players) cannot simultaneously be empowered to impose different outcomes. But surely this is not a fair translation of Sen's condition; as Bernholz (1974, p. 100) says, "under the rule of liberalism, no individual faces alternative social states among which he can freely choose *for* society." Rather, we can follow up on the suggestions of Bernholz (1974, p. 101; 1976, p. 27) and Gibbard (1974, p. 390) and interpret the "rule of liberalism" as empowering an individual (and, minimally, at least one other) to decide at least one "issue," e.g., (to use an example originally introduced by Sen, 1970, p. 153) whether his walls shall be painted pink or white. Putting the matter more abstractly, the set of social states is partitioned into two subsets: those in which his walls are pink and those in which they are white. The individual is then empowered to determine whether the realized outcome shall belong to the first subset or the second (he is "decisive" between the two subsets). Putting the matter more generally[3] and also into game-theoretical terms, the individual player has a strategy s such that at least one outcome, say b, does not belong to any situation including s, i.e., the player has the power to preclude (or, as Aldrich says, "to rule out") at least b as the realized outcome. (What outcome is realized depends of course on the strategy selections of the other players.)

[2] Obviously, such power to impose an outcome can exist only if several situations belong › the same outcome; cf. footnote 1.

[3] Cf. Sen (1970), footnote 2.

It can be seen quite readily that two disjoint coalitions (including two individual players) can simultaneously be empowered to preclude different outcomes. So if Sen's condition is translated in this fashion, no immediate impossibility results. But how does domination, and thus "social preference," come about as a result of such "power to preclude"? Suppose that a coalition can preclude b such that the realized outcome must be some outcome in the set A. The spirit of classical (von Neumann-Morgenstern) theory seems best to be retained if we require for domination that the coalition members prefer every outcome in A to b (in which case every outcome in A dominates b). In other words, domination comes about, not necessarily because a coalition can impose a preferred outcome, but because a coalition can preclude an outcome in such a way that whatever then happens is preferable.[4]

Now, if we accept this notion of domination, "social preference" in the Prisoners' Dilemma is filled out further (see Figure 2), though it remains incomplete. We do get a contradiction in "social preference," as we have both b P c and c P b. But d in turn is "socially preferred" to both b and c, and a is "socially preferred" to d; thus a is the predictable, and Pareto-optimal, realized outcome. So it seems that, on this interpretation, the Prisoners' Dilemma allows for "liberalism" but presents no "liberal paradox."[5]

Thus it appears that we still have not captured Aldrich's understanding of "social preference" as generated by a game. In the particular case of the Prisoners' Dilemma, this understanding seems best to be captured in a less than fully cooperative framework (in which it is assumed that players cannot make binding commitments), in which we equate "social preference" with "vulnerability." Again following Farquharson (1969, pp. 24, 51), we say that a situation t belonging to outcome b is *vulnerable* to a set S of players if, given that all players not in S continue with their present strategies, the players in S can change their strategy selections in such a way that the new situation belongs to an outcome a they all prefer to b. We might then say that a is "socially preferred" to b. Again in the special case in which S includes a single player, we may have a kind of "liberalism" condition; and in the special case in which S includes all players, the Pareto principle is relevant (and only in this latter case is vulnerability equivalent to domination).

[4]But notice that if "liberalism" is defined in this fashion, i.e., everyone (or, minimally, at least two players) has some "power to preclude," and if domination is defined in this fashion, "social preference" (i.e., domination) may still oppose at every point the preferences of a player so empowered – that is, whether such power is useful to the player depends on the nature of his preferences. Certain restrictions on individual preferences, along the lines of se arability, may assure that such power is useful.

[5]Essentially, working within a cooperative framework, we have permitted what Gibbard (1974, pp. 400-401) calls "alienable rights" and what Bernholz (1976, pp. 29-30) calls "logrolling." More generally, the game-theoretical framework naturally incorporates the distinction emphasized by Gibbard, Bernholz, and Kelly (1976) between "according a right" (having power) and "exercising a right" (using power).

FIGURES

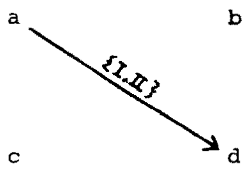

Figure 1 - Domination in the Prisoners' Dilemma Reflecting only "Power to Impose"

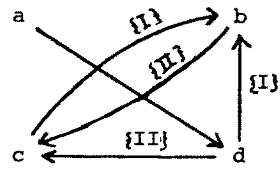

Figure 2 - Domination in the Prisoners' Dilemma Reflecting "Power to Preclude"

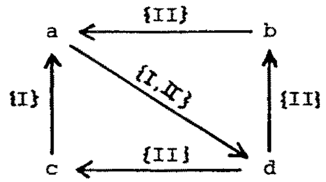

Figure 3 - Vulnerability in the Prisoners' Dilemma

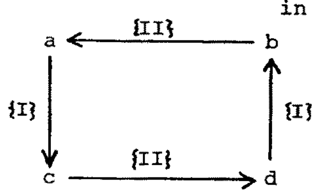

Figure 4 (a)

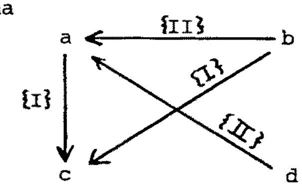

Figure 4 (b)

Figure 3 shows vulnerability relationships in the Prisoners' Dilemma, where $a \xrightarrow{S} b$ now means "*b* is vulnerable to *S* with respect to *a*." The resulting "social preference" is incomplete but nevertheless cyclical and thus illustrates the "liberal paradox" as Aldrich intends.

But is this interpretation of "social preference" generally satisfactory? We may note several problems. First, keeping the same matrix as for the Prisoners' Dilemma, let the players have the following (strictly competitive) preferences: $a\ P_I\ d\ P_I\ b\ P_I\ c$ and $c\ P_{II}\ b\ P_{II}\ d\ P_{II}\ a$. The resulting "social preference" is shown in Figure 4(a) (and, for contrast, "social preference" in the cooperative/domination sense is shown in Figure 4(b)). In this case, the Pareto principle is irrelevant and "liberalism" alone entails a "social preference" cycle, i.e., contradicts SDF. (cf. Gibbard, 1974, p. 389, and Aldrich's example of "indiscriminant liberalism" on p. 5).

Second, to this point in considering "social preference" in the vulnerability sense, we have ignored one problem: several situations may belong to the same outcome. This is a problem because vulnerability is defined over situations, not outcomes, and one situation belonging to a given outcome *a* may be vulnerable to *S* while another situation belonging to the same outcome *a* may not be. Has some "social preference" involving outcome *a* nevertheless been established? If so, it follows that two players (or disjoint coalitions) may be "decisive" for the same pair of outcomes – a direct contradiction in the social choice framework. Moreover, it also follows that every non-dummy player brings about at least one vulnerability relationship in every game, so on this interpretation every "proper" game (i.e., that has at least two outcomes and in which no individual player is all powerful) meets the "minimal liberalism" condition, including for example a majority voting game. But Sen (1970, p. 152) makes it clear that "liberalism" is a limitation on majority rule, not an aspect of it. Indeed, it seems to me that what characterizes all games is not something akin to the "liberalism" condition of social choice theory (as Aldrich tends to suggest on p. 7) but something more akin to the "positive responsiveness" condition of social choice theory.

I conclude this comment, therefore, on a question mark: the precise connection between social choice theory and game theory remains an open question. Accordingly, the significance of Aldrich's article is difficult to judge. But it certainly presents arguments that are of interest; and if it stimulates further work on the connection between social choice theory and game theory, it will prove to be especially valuable.

REFERENCES

Bernholz, Peter. "Is a Paretian Liberal Really Impossible?" *Public Choice*, 20 (Winter 1974).

__________. "Liberalism, Logrolling, and Cyclical Group Preferences." *Kyklos*, 29 (1976).

Bloomfield, Stefan D. "A Social Choice Interpretation of the von Neumann-Morgenstern Game." *Econometrica*, 44 (January 1976).

Farquharson, Robin. *Theory of Voting*. New Haven: Yale University Press, 1969.

Gibbard, Allan. "A Pareto-Consistent Liberatarian Claim." *Journal of Economic Theory*, 7 (1974).

Kelly, Jerry S. "Rights Exercising and a Pareto-Consistent Libertarian Claim." *Journal of Economic Theory*, 13 (1976).

Luce, R. Duncan, and Howard Raiffa. *Games and Decisions*. New York: Wiley, 1957.

Mayhew, David R. *Congress: The Electoral Connection*. New Haven: Yale University Press, 1974.

Plott, Charles R. "Axiomatic Social Choice Theory: An Overview and Interpretation." *American Journal of Political Science*, 20 (August 1976).

Sen, Amartya K. "The Impossibility of a Paretian Liberal." *Journal of Political Economy*, 78 (January-February 1970).

Wilson, Robert B. "The Game-Theoretic Structure of Arrow's General Possibility Theorem." *Journal of Economic Theory*, 5 (1972).