DISCUSSION PAPER NO. 212

On Preferences, Beliefs and Manipulation Within Voting Situations

by

Jean-Marie Blin and Mark A. Satterthwaite

March 8, 1976

Department of Managerial Economics
and Decision Sciences
Graduate School of Management
Northwestern University
Evanston, Illinois 60201

Forthcoming in Econometrica (May or July 1977).

*The comments of Salvador Barbera, David Baron, Leonid Hurwicz, Ehud Kalai, John Roberts, and an anonymous referee were of great help in this paper's development.

This paper supersedes Discussion Paper No. 129, "An Impossibility Theorem for Deterministic Organizations," January 1975.

On Preferences, Beliefs and Manipulation

Within Voting Situations

by

Jean-Marie Blin and Mark A. Satterthwaite

March 8, 1976

ABSTRACT

This paper shows that no nondictatorial voting procedure exists that induces each voter to choose his voting strategy solely on the basis of his preferences and independently of his beliefs concerning other voters' preferences. This necessary dependence between a voter's beliefs and his choice of strategy means that a voter can manipulate another voter's choice of strategy by misleading him into adopting inaccurate beliefs concerning other voters' beliefs.

On Preferences, Beliefs and Manipulation Within Voting Situations

by

Jean-Marie Blin and Mark A. Satterthwaite

March 8, 1976

I. Introduction

Consider a voting situation, as in a committee. Each rational member has preferences over the alternatives being considered and beliefs concerning the other members' preferences. The question we consider in this short paper is: can a voting procedure be constructed such that each member's vote depends only on his preferences, not on his beliefs concerning other individual preferences. We show, by an application of Gibbard [6] and Satterthwaite's [11] impossibility theorem for strategy-proof voting procedures, that such a voting procedure does not exist. Moreover, we show that this necessary lack of independence between a member's beliefs and his choice of voting strategy makes him vulnerable to possible manipulation by other members. Specifically, consider members one and two. Since member one partially bases his vote on what he believes member two is seeking, member two may deliberately mislead member one into adopting a false belief concerning member two's preferences. As a consequence of this inaccurate belief, member one may decide to vote in a manner that is, in fact, unfavorable to himself and favorable to member two.

Derivation of these results depends critically on the possibility that members may be uncertain concerning other members preferences. This assumption is reasonable because the purpose of legislative bodies is to reconcile conflicting preferences. If preferences were generally known with certainty, then, as Wilson [14, p. 310] has pointed out, the need for a legislative body would vanish because preferences could be aggregated directly. Therefore, a realistic analysis of voting behavior must accept that a member's true preferences are private.

Our results are consistent with the work that other researchers have reported. Dummett and Farquharson [3, pp. 34-35] and, to a lesser extent, Wilson [14] assumed the validity of our results. Harsanyi [7] in discussing bargaining situations where the two opponents are uncertain concerning the other's preferences argued that the decisive element may not be the actual preferences of the two individuals involved, but rather the societal "stereotypes" (beliefs) concerning their preferences. Schelling [12, e.g. Ch. 3] in his insightful discussion of bargaining strategy dwells extensively on the same theme.

II. Formulation

A brief statement of the formal notation that we use for analyzing voting situations follows.[1] A committee consists of a finite set $N = \{1,2,\ldots,n\}$ of n members, $n \geq 2$. The committee must select one alternative from a given, finite set X of m alternatives, $m \geq 3$. The committee makes its decision among the elements of X by voting. Each member $i \in N$ picks a strategy s_i from his set of admissible strategies S_i. A strategy, for example, might be to vote yes on the first motion, no on the second motion if the first motion passed, and no on the second motion otherwise. The voting procedure is a singlevalued function $g(s)$ that evaluates the members' strategy profile $s = (s_1,\ldots,s_n) \in S = \Pi_{i=1}^{n} S_i$ and selects one element of X as the committee's chosen alternative.

Each agent $i \in N$ has preferences R_i that form a complete, reflexive, and transitive ordering on X. The notation $x \, R_i \, y$ means that member i either prefers alternative x over alternative y or is indifferent between the two. The notation

$x \bar{R}_i y$ means that member i strictly prefers x over y. All conceivable R_i on X are admissible as a member's preferences. Let $R = (R_1,\ldots,R_n)$ be the preference profile, let $\mathcal{R}$ be the collection of all possible complete, reflexive, and transitive orderings over X, and let $\mathcal{R}^n = \Pi_{i=1}^n \mathcal{R}$, the n-fold cartesian product of $\mathcal{R}$, be the set of admissible preference profiles.

As we emphasized in the introduction, the preferences of each member i are internal to his own mind. Nevertheless, other members, while not having knowledge concerning member i's preferences, do have beliefs concerning member i's preferences. In order to avoid the complexities and unsolved problems that arise in the analysis of games of incomplete information as studied by Harsanyi [8], we limit ourself here to consideration of the two extremes of beliefs: complete uncertainty at one extreme and complete certainty at the other extreme. Agent i's beliefs are represented by λ_i. If he is completely uncertain concerning what others' preferences are, then $\lambda_i = \Pi_{i=1}^{n-1} \mathcal{R} = \mathcal{R}^{n-1}$. If he is certain, whether correctly or incorrectly, that the other members have preferences $R_{)i(} = (R_1,\ldots,R_{i-1}, R_{i+1},\ldots,R_n) \in \mathcal{R}^{n-1}$, then $\lambda_i = R_{)i(}$. Let $\lambda = (\lambda_1,\ldots,\lambda_n)$ be called the belief profile.

Based on his preferences R_i and beliefs λ_i each member chooses the strategy $s_i \in S_i$ that he thinks is best for him.[2] For each agent i let this calculation be described by the decision function $\sigma_i(R_i,\lambda_i) = s_i$, i.e. if agent i has preferences R_i and beliefs λ_i, then he picks strategy s_i. If preferences R_i, beliefs λ_i, and decision functions σ_i are given for all $i \in N$, then the outcome of the committee's vote is determined: $x = g[\sigma_1(R_1,\lambda_1),\sigma_2(R_2,\lambda_2),\ldots,\sigma_n(R_n,\lambda_n)]$. Let $\Sigma(R,\lambda) = [\sigma_1(R_1,\lambda_1),\ldots,\sigma_n(R_n)]$ be called the decision pattern for voting procedure g and let the pair (g,Σ) be called a completely described voting procedure (CDVP).

A CDVP (g,Σ) is defined to be belief independent if and only if, for every preference profile R and all pairs of belief profiles (λ,λ'), $\Sigma(R,\lambda) = \Sigma(R,\lambda')$, i.e. each member's choice of strategy must always be independent of his beliefs The effective range of a CDVP (g,Σ) is the set $Y(g,\Sigma)$ of outcomes $x \in X$ for which a preference profile R and a belief profile λ exist such that $g[\Sigma(r,\lambda)] = x$. Let $|Y(g,\Sigma)|$ be the number of elements contained within $Y(g,\Sigma)$. A CDVP (g,Σ) is dictatorial if and only if a member i exists such that, for all preferences profiles R, all belief profiles λ, and all alternatives $x \in Y(g,\Sigma)$, $g[\Sigma(R,\lambda)]\ R_i\ x$.

III. The Inconsistency Between Belief Independent Voting Procedures and Individual Rationality[3]

Consider a CDVP (g,Σ) that is belief independent. Suppose that it is not dictatorial and that its effective range contains at least three outcomes. Since (g,Σ) is belief independent, its outcome for any preference profile R and belief profile λ will be independent of λ. Therefore, the functional relation among R, λ, and the outcome x may be written as:

(1) $$g[\Sigma(r,\lambda)] \equiv g[\Sigma^*(R)] \equiv f(R) = x$$

where $\Sigma^*(R) = [\sigma_1^*(R_1),\ldots,\sigma_n^*(R_n)] = \Sigma(R,\lambda)$ and f is the composition of g and Σ^*.

Since (g,Σ) is not dictatorial, $|Y(g,\Sigma)| \geq 3$, and admissible preferences are unrestricted the impossibility theorem for strategy-proof voting procedures of Gibbard [6] and Satterthwaite [11] implies that a preference profile R^o, a member i, and a preference ordering R_i' must exist such that

(2) $$f(R^o_{)i(},R_i')\ \bar{R}^o_i\ f(R^o_{)i(},R^o_i)$$

where

(3) $$R^o_{)i(} = (R^o_1,\ldots,R^o_{i-1},R^o_{i+1},\ldots,R^o_n).$$

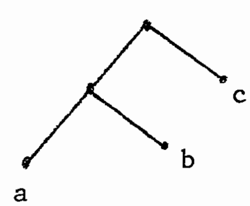
The meaning of (2) is this. Because (g,Σ) is belief independent, the members choose the strategy profile $s = \Sigma^*(R^o) = [\sigma^*_1(R^o_1),\ldots,\sigma^*_i(R^o_i),\ldots,\sigma^*_n(R^o_n)]$ whenever the preference profile is R^o. If member i, however, has enough information, then he can calculate that strategy $s^o_i = \sigma^*_i(R^o_i)$ is not optimal for him; according to (2) he can switch the outcome from $f(R^o_{)i(},R^o_i)$ to $f(R^o_{)i(},R'_i)$, which he strictly prefers, by selecting strategy $s'_i = \sigma^*_i(R'_i)$ instead of strategy $s^o_i = \sigma^*_i(R^o_i)$.

Presumably member i defined his decision function $\sigma^*_i(R_i) = \sigma_i(R_i,\lambda_i)$ to be independent of λ_i because he calculated that his beliefs λ_i were irrelevent to his choice of strategy. In particular, if he was rational, he must have decided that he should always act identically in the two polar cases concerning his beliefs about other members preferences: complete uncertainty versus absolute certainty. Therefore he must have calculated that in the case of complete uncertainty his best strategy would be $s_i = \sigma^*_i(R_i)$ when his preferences were R_i. But (2) implies that then he could have calculated that $s'_i = \sigma^*_i(R'_i)$, not s^o_i, is his best strategy whenever he has preferences R^o_i and is certain that other members have preferences $R^o_{)i(}$. In other words, the elimination of his uncertainty concerning other members' preferences gives him reason to switch from strategy s^o_i to strategy s'_i.

Consequently, if he is rational, member i will abandon his decision function $\sigma^*(R_i)$ and adopt a second function, $\sigma'_i(R_i,\lambda_i)$, that will take on the value s_i when $R_i = R^o_i$ and $\lambda_i = \mathcal{R}^{n-1}$ and the value s'_i when $R_i = R^o_i$ and $\lambda_i = R^o_{)i(}$. This change, however, means that g is no longer belief independent.

The purpose of this demonstration has been to show that uncertainty concerning other members' preferences makes every possible nondictatorial voting procedure that has at least three outcomes in its effective range dependent on beliefs. This indeterminacy exists even in those voting procedures that, if every agents' preferences are always known by every other agent, are strictly determined. The following example illustrates this point.

Suppose $N = \{1,2,3\}$ and $X = \{a,b,c\}$. The group makes a decision among the elements of X by taking two majority votes. The first vote is between the set $\{a,b\}$ and the alternative c. If c receives a majority, then it is the outcome. Otherwise a second vote is taken between a and b with the winner being the outcome. In tree form the two votes may be represented as:



If each agent is certain of every other agents' preferences, then Farquharson [5, p. 42] has shown that this method of voting is, within the terminology of von Neumann and Morgenstern [13, section 14 in Chapter III], strictly determined.

For example, suppose every member believes that the preference profile R is

$$R_1 = (a\ b\ c),$$

$$R_2 = (a\ c\ b),$$

$$R_3 = (b\ c\ a)$$

where $R_1 = (a\ b\ c)$ means that member 1 most prefers a, second most prefers b, and least prefers c. Backward induction shows that a will win the second vote between a and b if it should take place. Consequently every rational agent realizes that the first vote is not between $\{a,b\}$ and c, but is actually between a and c, a vote that alternative a wins two to one. Therefore, when each member has perfect information, the determinate outcome is a.

The outcome is different when each agent is completely ignorant of every other agents' preferences. In such a case backward induction is unavailable and agents must resort to rules for decision making under uncertainty like maximin. If each agent does follow maximin, then the outcome for profile R is c, not a. Thus, as our result predicts, even this simple voting mechanism is not belief independent if its members are rational.

IV. Manipulation of Beliefs

In the previous section we showed that the strategic choices of a rational member within a CDVP depend both on his preferences R_i and his beliefs λ_i. This dependence on beliefs opens up possibilities for manipulation because beliefs, in general, are formed on the basis of past experience. Therefore if a member within a CDVP (g,Σ) realizes before the other members that they are mutually involved in a game, then he may be able to manipulate the outcome by influencing the other members' beliefs.

For example, consider the same voting procedure that we considered in the last section. Suppose the preference profile R is

$$R_1 = (c\ a\ b),$$
$$R_2 = (a\ c\ b),$$
$$R_3 = (b\ a\ c).$$

If every member a priori knows for certain every other members' preferences, then backward induction results in a being the outcome. If, however, member one somehow misleads the other two members into believing that his preferences are $R_1' = (c\ b\ a)$, then member two, based on his backward induction, will vote for c instead of $\{a,b\}$ on the first vote, thus making c the outcome. Moreover, member one's misrepresentation cannot be detected by the other two members because on the first ballot member two changes his vote on the false belief that member one would vote for b over a on the second ballot. But because member two changes his vote on the first ballot, the second ballot never takes place and member one is not forced to reveal his preference between a and b.

This example is not an artifact unique to this particular voting mechanism. The possibility of such manipulation exists within any nondictatorial CDVP (g,Σ) where members base their choices of strategies on their beliefs and preferences. To prove this assertion consider a nondictatorial CDVP (g,Σ) for which $|Y(g,\Sigma)| \geq 3$ and suppose that every member is certain that he knows every other member's preferences accurately, i.e. $\lambda_i = R_{)i(}$ for all $i \in N$. Therefore, $\Sigma(r,\lambda) = [\sigma_1(R_1,\lambda_1),\ldots,\sigma_n(R_n,\lambda_n)] = [\sigma_1(R_1,R_{)1(}),\ldots,\sigma_n(R_n,R_{)n(})]$ may be rewritten as $\Sigma^*(R)$, i.e. the outcome is a function only of the preference profile R. Exactly as in section III, Gibbard [6] and Satterthwaite's [11] theorem applies: a R and R_i' must exist such that

$$(4) \qquad g[\Sigma^*(R_i',R_{)i(})]\ \bar{R}_i\ g[\Sigma^*(R_i,R_{)i(})].$$

The interpretation of (4) is that if member i perceives soon enough that the other members have preference profile $R_{)i(}$, then he can manipulate the outcome in his own favor by misleading, if possible, the other members into believing that his preferences are R_i', not R_i.[4] Moreover, this manipulation will not be detected, even _ex post_, by the other members because the left hand side of (4) states that member i achieves the favorable outcome $g[\Sigma(R_i',R_{)i(})]$ by actually acting as if his preferences were R_i'.

V. General Discussion of the Results

The results of sections III and IV together imply that a voting procedure's decision may be unpredictable because the roots of the decision, at least in part, lie in a past history that the members may or may not have accurately perceived and analyzed. Specifically, section III established that a rational member will base his choice of strategy on both his preferences R_i and his beliefs λ_i concerning other members' preferences and intentions. These beliefs are based on his past experience, his past observation, his prejudice, etc. Such beliefs, of course, may be inaccurate, e.g. member i may be certain that member j has preferences R_j' when in fact he has preferences R_j. Inaccuracies may then reflect themselves in inappropriate choices of strategies with consequent unsatisfactory outcomes.

The prevalence of inaccurate beliefs would seem to be high even if members related to each other in a relatively straightforward manner. The situation, however, is likely to be worse because beliefs have impact and therefore members will not necessarily relate with each other straightforwardly. As section IV showed, if a member uses a priori information in making his choice of strategy, then he opens himself to the possibliity of being manipulated through acceptance of false information. Thus the importance of beliefs gives members an incentive to induce inaccuracies in other members' beliefs. Clearly this rapidly becomes a complicated game and the possibility of miscalculation may be great. Each member may believe that he is outguessing the other, but everyone mutually outguessing everyone else is not possible. If the net result is that all members have inaccurate beliefs, then an _ex post_, grossly unsatisfactory outcome would seem quite possible.

For example, consider an expanded version of the voting procedure that we have considered as an example in the previous two sections. Let $X = \{a,b,c,d\}$ and let the first vote be between $\{a,b\}$ and $\{c,d\}$. The second vote is between either a and b or c and d depending on which set won the first vote. Let the preference profile R be

$$R_1 = (b\ a\ c\ d),$$
$$R_2 = (a\ c\ b\ d),$$
$$R_3 = (d\ b\ a\ c).$$

Note that a is unanimously preferred to c. Assume that members one and two believe with certainty that the preference profile is R. Backward induction, based on the belief that member three also believes that the profile is R , causes them to vote on the first ballot for $\{a,b\}$ and $\{c,d\}$ respectively. Assume that member three mistakenly believes with certainty that the preference profile is R':

$$R'_1 = (b\ a\ d\ c),$$
$$R'_2 = (a\ c\ d\ b),$$
$$R'_3 = (d\ b\ a\ c).$$

His backward induction leads him to case a decisive vote for the set $\{c,d\}$. Alternative c, the Pareto dominated alternative, thereupon wins the second vote. Thus member three's inaccurate beliefs lead to selection of the Pareto dominated alternative.[5]

In this particular example member three would have done better to have disregarded his beliefs and, in accordance with the maximin criterion, voted for the set $\{a,b\}$. But, as one of the examples in section III showed, the maximin criteria can also lead to an *ex post* unsatisfactory outcome. Moreover, even if a prescription to disregard beliefs were justified, members would be unlikely to accept it. If an agent believes that he knows what another agent will do, then he is going to use that information if it is relevant.[6] In fact, as Riker [9, p. 24] has pointed out, the canons of a profession like law state explicitly that such beliefs should be acted upon if such action is in the interest of the client.

We should emphasize, however, that the conclusions of this paper are qualitative. We have shown that for voting procedures the beliefs of members must have decisive effect in at least one possible situation. The important question that this analysis does not answer is empirical: how frequently do the problems that we have identified actually occur. That question, we suspect, cannot be answered in general but rather must be answered case by case through the analysis of specific mechanisms. An example of this type of analysis, which involves a market situation rather than a voting situation, is the work of Roberts and Postlewaite [10]. They have studied the incentive properties of the competitive mechanism within an exchange economy. Their results imply that with the exception of two situations every member's choice of an optimal strategy always depends on his beliefs concerning other members preferences. The first situation exists whenever no trades are possible because the initial endowments happen to be Pareto optimal. The second situation exists whenever the number of members approaches infinity and each members' impact on prices goes to zero in the limit. Therefore it appears that, even within competitive markets, beliefs may play a frequent role in determining outcomes.

Footnotes

1. For more comprehensive discussions see, for example, Gibbard [6], Satterthwaite [11], and Blin and Satterthwaite [2].

2. If a member's calculations show that two distinct strategies s_i, $s_i' \in S_i$ are both optimal, then our assumption is that he will use a nonrandom rule to pick one of them.

3. The derivation in this section is a simple adaptation of Gibbard's proof [6] that no straightforward game form exists.

4. If the other members' beliefs are such that they are already certain that member i's preferences are R_i, then his attempt to mislead will have no effect except on his own future credibility.

5. The outcome of this example is not the result of a pathological preference profile; profile R has no majority rule intransitivities.

6. Authorities on family counseling (Bach [1, e.g. Chapter 11] and Ellis and Harper [4, pp. 203-205]) prescribe that a person in dealing with his or her spouse should largely disregard her or his beliefs concerning the other's preferences. They base this recommendation on the empirical observation that a person is likely to greatly exaggerate how well he knows his spouse's preferences. As a result, acting on the basis of beliefs is likely to lead, on average, to a less satisfactory outcome than acting as if one is uncertain of the other's preferences.

References

[1] Bach, George R., and Wyden, Peter. The Intimate Enemy: How to Fight Fair in Love and Marriage. New York: Avon, 1970.

[2] Blin, Jean-Marie, and Satterthwaite, Mark A., "Individual Decisions and Group Decisions: The Fundamental Differences." Discussion Paper #183, The Center for Mathematical Studies in Economics and Management Science, Northwestern University, Evanston, IL., November 1975.

[3] Dummett, Michael, and Farquharson, Robin. "Stability in Voting". Econometrica, 29 (1961), 33-43.

[4] Ellis, Albert and Harper, Robert. A Guide to Successful Marriage. North Hollywood: Wilshire Book Company, 1971.

[5] Farquharson, Robin. Theory of Voting. New Haven: Yale University Press, 1969.

[6] Gibbard, Allan. "Manipulation of Voting Schemes: A General Result". Econometrica, 41 (1973), 587-602.

[7] Harsanyi, John C. "Bargaining in Ignorance of the Opponents Utility Function". Journal of Conflict Resolution, 6 (1962), 29-38.

[8] Harsanyi, John C. "Games with Incomplete Information Played by 'Bayesian' Players, I-III". Management Science, 14 (1967), 159-182, 320-334, 486-502.

[9] Riker, William H. The Theory of Political Coalitions. New Haven: Yale University Press, 1962.

[10] Roberts, Donald John, and Postlewaite, Andrew. "The Incentives for Price-Taking Behavior in Large Exchange Economies". Econometrica, 44 (1976), 115-128.

[11] Satterthwaite, Mark A. "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions". Journal of Economic Theory, 10 (April 1975), 187-217.

[12] Schelling, Thomas C. The Strategy of Conflict. London: Oxford University Press, 1960.

[13] von Neumann, John, and Morgenstern, Oskar. Theory of Games and Economic Behavior. 3rd ed. New York: John Wiley and Sons, 1953.

[14] Wilson, Robert C., "The Role of Uncertainty and the Value of Logrolling in Collective Choice". La Décision: Agrégation et Dynamique des Ordres de Préférence. Proceeding of an international conference organized by the Centre National de la Recherche Scientifique, held at Aix-en-Provence, 3-7 July, 1967.