BERNARD GROFMAN AND CAROLE UHLANER

METAPREFERENCES AND THE REASONS FOR STABILITY IN SOCIAL CHOICE: THOUGHTS ON BROADENING AND CLARIFYING THE DEBATE*

ABSTRACT. The standard model of collective choice looks at aggregation procedures which take individual preferences as existing for a specified set of alternatives. In this paper we propose that actors also have preferences for rules of choice or characteristics of choice processes (e.g., the perceived fairness of procedures or the popularity of outcomes) rather than simply for alternatives (outcomes) themselves. We argue that the positing of the existence of meta-preferences can illuminate a number of areas of choice theory. Here we focus on one such area: the problem of "too much" stability in majority rule decision making — a stability which belies the standard theoretical results on the generic instability of majority rule processes.

  We also show that discussion of the prevalence of stability in collective decision making needs to be clarified because there are at least six distinct types of stability which are sometimes confounded in the literature.

I. INTRODUCTION

Some of the apparent puzzles in social choice theory and evident contradictions between theoretical predictions and empirical observations arise from limitations within the usual conceptualization of the process of collective decision making. Currently, people are presumed to have preferences among well-defined elements of a choice set, and the group's social choice mechanism aggregates these preferences to produce an outcome. We propose in this paper that persons may *in addition* have preferences over characteristics of choice processes and the operation of choice processes[1] (i.e., preferences for the features of the procedures which result in outcomes, and not simply preferences for outcomes *per se*.) We call such preferences "meta-preferences". Positing the existence of meta-preferences illuminates several aspects of choice theory. One area where social choice theory and empirical reality apparently diverge is the problem of "too much" stability of majority rule processes. We shall show that the concept of meta-preferences is quite useful in helping to understand reasons for stability. We argue that the outcomes of voting processes may be stable, even though majority (pairwise preferences are hopelessly

cyclical, if committee members/voters have preferences not just over alternatives, *per se*, but also over features of choices.

A considerable body of recent work in social choice indicates that collective decision-making will lead to a well-defined majority choice only under extraordinarily stringent conditions (Schofield, 1978; McKelvey, 1976; McKelvey and Wendell, 1976; Bell, 1978. See summary in Riker, 1982; cf. Feld and Grofman, 1984). Thus, the "paradox of cyclical majorities", wherein individuals have consistent preferences but the group majority preferences cycle over a large number of alternatives, should be commonplace. Yet, as a number of public choice scholars have pointed out (see, e.g., Chamberlin *et al.*, 1984) cyclical majorities are almost never observed and do not seem to be a real problem for group decision making. Why the world demonstrates so much (unexpected) stability has become a topic of considerable interest (see, e.g., Tullock, 1981; Niemi, 1983; and Shepsle and Weingast, 1981). Moreover the issue of stability is commonly seen as tied to the legitimacy of outcomes and the viability of democratic decision making.

We begin the paper by discussing the different types of stability which can occur. We then review the major explanations for stability offered in the literature. We then go on to suggest two new lines of investigation as to the factors which affect the various types of stability. One of these deals with meta-preferences. The meta-preference perspective which we shall explain in detail in Section IV unifies several existing stands in the literature relevant to the problem of stability and also suggests new possibilities. We wish to alert the reader, however, that our use of the term meta-preferences differs somewhat from that elsewhere. A number of authors have used the term meta-preferences to refer to an individual's preferences for the kind of person he ought to be (and the kind of preferences he ought to hold). For example, Jeffrey (1974: 377) provides a discussion (in terms of a symbolic logic notation) of preferences to be the kind of person who holds certain types of preferences, and of even higher levels of preferences among preferences. Similarly, Hirschman (1980: 69) has discussed "desires about desires." We are using the term meta-preferences differently from these authors.

The second new approach to stability involves another specification problem. Stability will be puzzling in some cases because the analyst does

not perceive the same alternatives as the decision-maker. An alternative *in posse* is not the same as an outcome in place. An alternative after it has been selected for implementation by some social choice mechanism may be, in several important senses, no longer the "same" alternative that voters first considered.

The last section of the paper will discuss implications for further research.

II. TYPES OF STABILITY

We believe that in much of the recent debate on "why so much stability?" authors have been partly talking past each other because of failure to recognize that the term "stability" has been used with a number of different meanings. Clarifying these meanings will allow us better to identify sources of stability.

We identify six basic types of stability.

Global preference stability, Type 1, occurs when a transitive social (majority) preference ordering (or, at minimum, a Condorcet winner) exists among the feasible set of alternatives.

Choice set stability, Type 2, is said to occur when a transitive social (majority) preference ordering (or, at minimum, a Condorcet winner) exists in the set of alternatives which are actually voted upon.

Riker-move stability, Type 3, occurs when there does not exist any new issue dimension such that, if choices were expanded to include that dimension, the social choice among the previously defined alternatives would be altered.

Stochastic predictability, Type 4, is said to occur when the expected outcomes of a given social choice mechanism are "well-behaved" in the sense of being predictable. For example, the expected outcomes may be confined to some limited portion of the issue space.

Outcome stability, Type 5, requires that the outcomes of a given social choice mechanism are accepted as legitimate (binding) once they are announced. At minimum, voters do not immediately seek to overturn outcomes, whatever they may be.

System stability, Type 6, exists when there is general willingness to accept the basic "rules of the game".

These six types of stability are distinct, as are the empirical evidence and theoretical results bearing on each. Indeed, Miller (1983) has suggested that (in our terminology) Type 1 instability implies Type 6 stability. In the remainder of this paper we discuss what is known about the connections among these six types of stability and introduce our two new arguments about why Type 5 stability, outcome stability, is so common.

Type 1 instability, which we shall call global preference instability, is the form of instability which social choice theory most leads us to expect in the spatial context. In this form of instability, all (or almost all) potential alternatives are in the top cycle set (McKelvey, 1976; Schofield, 1978).[2] It is important to point out that because all (or almost all) potential alternatives cycle, this does not imply that all (or even most) finite subsets of these alternatives contain a majority preference cycle. Thus, Type 1 instability does not imply Type 2 (choice set) instability, since we may find mechanisms to constrain selection of the alternatives from which actual choice is made that avoid a cycle in the limited set of choices within which votes are actually taken. We shall discuss such mechanisms in Section II of the paper.

Type 3 instability, which we shall call Riker-move instability, since Riker was the first to identify it clearly (Riker, 1961, 1982), refers to situations in which some of the losers can reverse an existing choice by widening the scope of conflict to include one or more new dimensions. Riker (1982) identifies such a Riker-move by the Whigs — repeated losers on an agrarian vs. commercial dimension of conflict — as the introduction of anti-slavery as an issue in U.S. politics in the middle 19th Century.[3]

Conditions leading to the possibility (or impossibility) of a Riker-move have not been theoretically examined, but we share Riker's view (personal communications, July 1984) that it is almost always possible *in principle* to find a Riker-move which will destabilize the status quo. However, for reasons that we identify in Section IV (e.g., uncertainty avoidance), such potential Riker moves may be resisted even by some actors who we might think could expect to gain from them. Moreover, even when Riker-moves are theoretically possible, cognitive and institutional barriers are likely to limit their frequency. The same barriers that limit the dimensionality of any conflict also inhibit later Riker-moves. Also, were all dimensions always present in conflict, no new dimension could be introduced.

There may be a Riker-move available even if we have Type 1 (or Type 2) stability, since, by definition, a Riker move involves the introduction of a *new* issue dimension. We should also note that there may be Riker-moves which promote stability, i.e., ones that take voters out of a cycle into a new space which has a core solution, albeit such a situation is probably quite unlikely.

Type 4 stability, stochastic predictability, can occur even if we have instability of the first three types. All that is required is that there be a social choice mechanism whose *probable* (albeit not necessarily certain) choices can be identified in advance for a given set of alternatives and preference profiles. There has been considerable recent work done on conditions sufficient for this type of stability. (See Section III below.) We believe this work to be of great promise, since, in our view, even when the core is empty, some alternatives are still apt to be much more probable than others.

Type 5 stability, outcome stability, and Type 6 stability, system stability, refer to post-decisional behaviors. An outcome of a social choice mechanism is said to be stable if, whenever a result is announced, voters accept it as binding, even if there may be alternatives which they prefer over that which was adopted. A social choice rule is said to be system-stable when voters regard it as legitimate mechanism for making collective choices. These two forms of stability are logically distinct. We might have an outcome accepted but revolt against future use of the rule; or we might have a rule which was generally held to be valid, but whose particular application was being protested as unfair. In either case, however, the test for these two types of stability lies in observation of post-decision behavior, and not in the structure of preferences *per se*.

Before discussing the relevance of meta-preferences to the study of stability, we shall briefly review nine approaches to stability found in the current literature.

III. PREVIOUS RESEARCH ON STABILITY

Various arguments have been made about both the theory and the empirical incidence of stability. Many of these arguments are, in our view, far more relevant to some one or two of the six types of stability we have

identified than to the others. The public choice literature contains nine main lines of attack on the question of stability. We note that these lines of attack are not mutually exclusive. In discussing each we shall indicate to which of our types of stability the research is most relevant.

(a) *Restrictions on the set of feasible preference orderings.* The properties of restriction which could give rise to situations in which there is a certainty (or at least a very high probability) of there being a clear majority winner have been extensively investigated (see e.g., Sen, 1966–1977). The standard interpretation is that the conditions sufficient to avoid cycles, e.g., single-peakedness, are too strong to be satisfied in most real-world group decision making (see Reiker, 1982; cf. Niemi, 1969; Feld and Grofman, 1984). Such conditions are particularly relevant to Type 1 stability (global preference stability) which cannot in general occur without some such preference domain restrictions. Such restrictions can also guarantee Type 2 stability, but that type can arise in other ways as well.

(b) *Restriction to the top cycle set.* If there is instability (i.e., cycles) a natural question to ask is how many alternatives are in the (top) cycle set. If that set is small, or if all of its members are "close together" (in a sense that can be made precise), it may not matter much if there isn't a unique majority winner. This line of approach appears most relevant to Type 4 stability, the predictability of outcomes. Since the (top) cycle set will in general be very large or even infinite (Bell, 1978; McKelvey, 1976; Schofield, 1978), this source of Type 4 stability is apt to be rare.

(c) *Agenda control.* There are features of parlimentary procedure and group dynamics that may either act to constrain the number of alternatives (as long as alternatives are few the probability of a majority winner remains high), or to delimit choice to a set of alternatives such that choice from among this set will be a acyclic or at least "well-behaved" or "predictable". Among the features that have been looked at are the parliamentary rule that requires any motion to defeat the status quo before it can be approved, the germaneness rule, and committee structures which impose a division of respectability across multiple issue-dimensions (see e.g., Shepsle, 1979; Shepsle and Weingast, 1981, 1982, 1984). Such features promote Type 2 and especially Type 3 stability. More generally, attention

has been called to the role of leaders and parties in structuring the agenda and/or forcing choices into a dichotomous mode (see, e.g., Downs, 1957; Roemer and Rosenthal, 1978).

(d) *Cognitive limitations.* Cognitive biases may constrain the domain of available alternatives and dimensions. These limitations arise from the proclivity of people to think in customary fashions and to mentally fit new or continued conflicts into pre-existing categories or pigeonholes. Institutions, including the press, reinforce the difficulty by providing publicity in well-established terms. A candidate whose positions are unclassified in standard fashion, such as a fiscal conservative who is also a social liberal, generally finds him or her self ignored into an ill-fitting compartment. One might interpret Gary Hart's 1984 primary battle in part as an ultimately unsuccessful attempt to define a new hole. This area of research has been almost entirely explored. Such cognitive constraints enhance the likelihood that potential Type 3 instabilities will never be permitted to surface.

(e) *Coalitional or other structural features of the top cycle set.* Recent work on the uncovered set (Miller, 1980; Shepsle and Weingast, 1984; McKelvey, 1983; Banks, 1984; Miller *et al.*, 1985; Feld, *et al.*, 1985) has shown that even if the top cycle set is large, a centrally located subset of it may have various desirable properties and may arise "naturally" from various agenda processes (including sophisticated voting). Thus, the internal structure of the majority preference relation may strongly delimit choices in a fashion which can be thought of as relatively "institution-free", even if all elements cycle.

(f) *Logrolling.* Gordon Tullock (1970) has hypothesized that (in our terminology) Type 2 stability may be achieved if voters trade votes across different vote decisions or across multiple issue-dimension. A considerable amount of formal research has been done on this question, with the general conclusions being negative (see, e.g., Schwartz, 1981), but Shepsle and Weingast (1982) have looked at rational choice explanations for the development of reciprocity norms in legislatures and other decision-making bodies whose members can expect to have long-term interactions.

We believe that if vote-trading is a stabilizing factor, it is most important in terms of Type 5 stability, not Type 2 stability, because of the existence of meta-preferences, such as those for honoring commitments (see below). Miller (1983: 738) has argued, though for somewhat different reasons, that logrolling may foster Type 6 stability. Note also that the

notion of logrolling is distinct from that of Riker-moves, though both involve multiple issue dimensions.

(g) *Appearance of stability masking instability.* If voters lack complete knowledge of the preferences of the other voters in the group, there may be a cycle without anyone being aware of it. This is, of course, directly relevant to the appearance of Type 2 stability. This line of research has not been much investigated (see, however, Riker, 1961), although there are a few articles on how easy it is to detect the presence of a paradox of cyclical majorities from observable roll-call (or other ballot) data (see, e.g., Bowen, 1972).

(h) *Supramajoritarian rules and veto rules*: If we shift from a simple majority game to a supramajoritarian game, there may well be a core in the latter game, although there is none in the former (Kramer, 1977). Such supramajoritarian elements may be either formally or informally imposed. Such forms of structure-induced stability are most applicable to Type 2 stability.

(i) *Stochastic approaches*: Even if most or all items are in the top cycle, there will be one or more sets of alternatives (Von Neumann–Morgenstern solutions) such that at least one of the members of such a set can defeat every alternative not in the set. In spatial voting games, a set of alternatives with this property may be found in a rather narrowly circumscribed area of the issue space, and with the further property that, if alternatives are posed with a probability related to the number of voters who favour them to the then existing status quo, then "most of the time" (in a sense that can be made precise) outcomes will be confined to a neighborhood of this set (Ferejohn *et al.*, 1980; Packel, 1981; Ferejohn *et al.*, 1984; Glazer *et al.*, 1984; and Grofman *et al.*, 1984). This new and promising line of research shows how Type 4 stability (stochastic predictability) can be fully compatible with *in*stability of the first three types.

As can be seen from the above discussion, arguments in the public choice literature about reasons for stability apply to some, but not all, of the types of stability we have identified. Recognition of the differences among these types of stability clarifies and enriches the debate.

IV. SPECIFYING THE REALM FOR CHOICE: META-PREFERENCES

A limitation common to all of the approaches described above is failure

to pay sufficient attention to the problem of specifying the alternative space. If we are truly to understand the nature of real world choices we must specify the set of alternatives from among which choices are to be made *as these alternatives are seen by the participants*. Some of the confusion over the nature of stability may arise from disparity between the outcomes set seen by the analyst and that perceived by participants. In particular, participants may consider as "different" outcomes classified as "the same" by the observing analyst.

   In many cases, a potential outcome changes its nature as it moves through the choice process until finally being chosen (or rejected). In particular, once an outcome has come to pass there is associated with it the entire history of its passage.[4] The outcome subsumes the whole set of trades, negotiations, debates and votes which went into its final passage. Indeed, in the U.S., courts may take this legislative history as a vital clue in interpreting the "meaning" of legislation. The concept of "face" also involves a recognition of some of this sense in which the process of decision contributes elements to the choice set. In addition, as the decision process continues, new information is usually made available, for example, on the amount (and intensity) of support for various alternatives. Once an outcome has become a reality, preferences may shift in its favor because of a cognitive dissonance reduction effect, or because the outcome now has enhanced legitimacy, or because of extreme difficulty and cost to undo what has already been decided upon, or because of generally agreed upon "don't rock the boat" norms. In other words, the process of choice itself can alter the characteristics of "alternatives," and furthermore, a correctly specified alternative space includes preferences across features generally excluded entirely from analysis. These features include the characteristics of choice mechanisms themselves and also the history of the social interaction which resulted in the choice. (As we shall discuss below, modelling these neglected preferences entails more than simply adding an additional attribute to a utility function.) Such considerations lead us to expect outcome stability (Type 5) far more often than global preference stability (Type 1) or even choice set stability (Type 2).

   We shall identify seven features of decision outcomes and decision processes (other than simply preferences for the alternative which is to be chosen) about which individuals may have preferences and which affect the specification of the outcome space. Such preferences we shall call

meta-preferences. Meta-preferences are preferences for rules for choice or for the characteristics of choice processes, rather than simply for alternatives (outcomes) themselves. Decision makers are posited to have simple preferences across the elements of choice and, in addition, meta-preferences across what may loosely be thought of as different "mechanisms" for choice. These two types of preference may interact. How a voter will ultimately choose will depend upon the characteristics of this interaction. The importance of meta-preferences can be seen in the ubiquitousness of procedures (i.e., cloture rules, germaneness, alteration of pro and con speakers) which appear to implement the types of meta-preferences we discuss below.

Meta-preferences can help us account for both Type 5 and Type 6 stability. More generally, meta-preferences affect outcomes by introducing considerations which may alter choices. Thus, stability creation is only one potential effect — although an important one — of meta-preferences. (Under some circumstances the effects of a given set of meta-preferences may in fact be destabilizing.) In any event, explicit consideration of meta-preferences enriches theories of choice by introducing features of choice important to decision makers yet incorporated insufficiently in present models.

We list below seven types of meta-preference. The list below is not intended to be exhaustive nor mutually exlusive. We do not believe it captures the most common sorts of meta-preferences. It is intended as an heuristic device.

(i) *Procedural fairness*, i.e., is the outcome the result of a procedure which is thought to be *basically* fair and reasonable? If it is, the fact that such a procedure may occasionally lead to perverse results may be of little relevance. For example, individuals may be generally willing to accept outcomes of rules which are voluntarily chosen and unwilling to accept outcomes of rules which lack that status, regardless of other characteristics of those outcomes. Individuals may care more about the face legitimacy of a decision procedure than about the specific choices produced by it. A "rational" choice argument can be given for the existence of a meta- preference for procedural fairness since the short-run gain of overturning a decision which is disliked may readily be outweighed by the long-run cost of disturbing agreed upon norms of how to do business

(cf. Rae, 1969).[5] This type of meta-preference may be especially relevant for system (Type 6) stability.

(ii) *Consensus*, i.e., does the outcome have (or appear to have) substantial popular support? Even if voters do not change their own preference ordering upon learning of the preferences of other group members (and some voters will, via conformity, persuasion, or the desire to belong, change their preferences to more nearly fit those of the group); voters may believe that outcomes which have garnered some specified level of wide support should, as a matter of principle, be supported. Other voters may simply have low tolerance for disagreement. In either case, a voter's preference for consensus or for agreement may outweigh his or her preference for any specific outcome. Here, too, a "rational" basis for such a meta-preference can be argued. If large majorities favor an alternative, the cost of trying to overturn it can be considerable, expecially if it eventually loses to an alternative which only barely defeats it (cf. Kramer, 1977). A preference for consensus could also arise where the specific object of choice has less salience or importance for the voter than does maintaining group cohesion or avoiding dissension. This latter preference could be rational if the cohesion facilitates obtaining other valued (material or psychological) goods.

(iii) *Uncertainty avoidance*. If there is enough potential instability, then the ultimate effect of tampering with the status quo is apt to be quite uncertain. If preferences are cyclical, most voters will have alternatives which they prefer to that status quo and which could receive the support of a majority in pairwise contest with the status quo; but if there are cyclical majorities, there will also be (many) alternatives which might be chosen which would leave the voter worse off than if he simply reconciles himself to the status quo. Thus, in situations of complexity with cyclical majorities across multiple dimensions, voters may be very reluctant to "rock the boat", since there is no guarantee that they will be better off doing so.

This line of reasoning is further strengthened if, instead of looking at individual voters, we look at potential majority coalitions. While such coalitions can find alternatives preferred by all members of the coalition to the status quo, there is always the fear of members being lured away into other coalitions by alternatives which, to them as individuals, are even

more attractive. Further, we suggest that even obtaining a preferred alternative may not in fact be preferred should that require intermediate uncertainty which prevents planning or greater expenditures on information gathering transactions, costs, insurance, etc. (See discussion of decision costs below.) Thus, we suggest that enough instability (of preferences) may imply stability (of outcomes)![6]

Decision-makers may place a high value upon maintenance of existing decision-making institutions and procedures, for reasons which may include custom, habit, reduction of information costs, and uncertainty avoidance. (If the institution fails, an unknown new one would replace it.) For example, debate over resolution of a policy issue may also entail (in a federalized polity) debate about the relative powers of a central federal government and the governments of the federation members. Conceivably, an advocate of increased federal funding for, say, elementary school science education might, nonetheless, not wish to alter the balance of powers between governments by increasing federal jurisdictional authority over education, with the attendant unforeseen consequences in this and other policy areas. Similarly, someone opposed to a particular reapportionment, such as the most recent in California, may still prefer maintaining the current system of reapportionment to risking the uncertain changes from the status quo produced by creating a reappointment commission.[7]

(iv) *Decision costs.* Decision costs are reduced by simplicity and by timeliness (reaching a decision which can implemented within a reasonable time period). Concern for decision costs leads voters to care about the simplicity of procedures (accounting, in part, for the wide popularity of simple plurality voting) and to prefer timeliness (manifested, for example, in parliamentary rules which make difficult or impossible reconsideration of previously defeated motions) and thereby avoid the spending of time in rehashing "already decided" issues. In addition, timeliness has a desirability that goes beyond the decision costs that it saves. For example,the rules that litigation and criminal charges must be decided within fixed time periods reduce uncertainty (see above), and also are perceived as "fairer" to the litigating parties. Moreover, insofar as time (and information gathering) is costly, voters may settle for less than ideal alternatives, because continuing search would be more expensive than the anticipated

gains are worth (cf. Hoffman and Packel, 1982). Finally, there are many types of real-world decisions where there are externally imposed time constraints, so that it becomes *necessary* to reach a decision (which may then, after a short time period, become essentially irrevocable). Thus, in general there can be considerable costs to changing your mind and considerable advantages to making your mind up quickly. These forms of meta-preference have a greater probable impact on outcome (Type 5) stability than on system (Type 6) stability.

(v) *Universalism norm.* Regardless of the formal characteristic of any particular procedures, if groups operate under a norm of "if I get my way this time, then you can have your way next time", committee members may be willing to reconcile themselves to particular undesirable outcomes, knowing that in the long run "things will even out". Shepsle and Weingast (1982; cf. Grofman, 1983) have shown how a meta-preference for universalism may arise from egoistically rational calculations even if the universalism norm is socially inefficient.[8] Like Miller (1983) we believe that such a norm can facilitate Type 5 and indeed Type 6 stability. Moreover, it may well be, as Miller (1983) suggests, that such a norm is facilitated when there are multiple issue dimensions — a condition conductive to Type 1 instability

(vi) *Civility norms.* Regardless of the formal characteristics of decision rules, most groups have strong norms about how business is to be conducted. Legislative rituals, for example, preserve the appearance of respect toward one's fellow legislators (Matthews, 1957). Whatever the outcome, people want to feel good and to be able to transact future business with one another. We believe that the widespread discontent with Carter's early concession in 1980 (before the polls had closed in California) stems in part from a violation of the norm that voting is a "sacred" act and that each voter's vote *is* to be counted (even if it is known to be irrelevant). Anything less would be disrespect. (Similarly, sports teams keep playing even though it may be obvious which team must win the championship bid.)

(vii) *Preference for decision-maker's-image.* There are two special cases of interest here. One is for being (or *not* being) "on the record" as in favour of a given outcome. The other is for being seen as the kind of person who honors commitments.

We may wish something to pass, but not want our name associated (e.g., a legislative pay hike), or, on the other hand, we may propose something which we know has no serious chance of passage in order to appease constituency demands (e.g., in proposing a constitutional amendment mandating a balanced budget). Similarly, we may be unwilling to engage in sophisticated voting maneuvers that will enhance the likelihood of the outcome we most prefer occurring, if doing so requires us to go on the record on some amendment in seeming opposition to what we truly favour (Riker, personal communication, July 1984). Again, of course, a rational choice argument for such a meta-preference can readily be constructed. We may not wish to have to explain away such seemingly anomalous votes to our constitutents or to the folks who put together roll-call indices for special interest groups. More generally, we cannot treat outcomes as being identical regardless of who does the implementing. For example, politicians may prefer having, say, a busing scheme, implemented by the courts, even though that scheme is (from their standpoint) inferior to the compromise they might have been able to come to among themselves which could have avoided court intervention. By leaving it to the courts, politicians can disclaim responsiblity. (An analogous argument can be made for the willingness of labor leaders to submit to "binding arbitration", even if the anticipated outcome is less favourable than what they could get by compromise (Debra Friedman, personal communication, July 1984)).

As for the meta-preference for honoring commitments even if it is costly to do so, the second case, a rational-choice argument for such a meta-preference, has been given by a number of authors based on life as a repeated game. Such meta-preferences may be especially relevant to type 5 stability.

V. DISCUSSION AND CONCLUSIONS

By distinguishing among six types of stability we have, we think, contributed to clarifying the debate about "why so much stability?" and have suggested new ground for research on empirical and theoretical links among the six types of stability. Further, the notion of meta-preferences that we have introduced provides an analytical tool that may aid in

modelling the circumstances leading toward or away from the various types of stability we have identified. It is important to recognize that decision-maker's conception of outcomes may include more than a simple listing of alternatives and may take into account features such as the nature of the deliberations, the size of the majority and the procedures by which choices are made.

Any specific choice will reflect not only the preferences across the outcomes, but also preferences across the characteristics of choice, that is, preferences about the choice process itself, broadly conceived. The idea of meta-preferences may enable us to integrate several diverse findings about group decision-making. For example, Hoffman and Packel's (1982) experiment, in which an imposed exogenous cost on the introduction of new notions led to the choice of outcomes not in the core, can be seen as an instance of an imposed meta-preference for timeliness. The Shepsle and Weingast (1982) work on porkbarrel coalitions can be seen as providing a rationale for a meta-preference for universalism. Buchanan and Tullock's (1962) analysis of decision costs provides a rationale for individuals' willingness to tolerate decision rules in which they do not have a veto. Similarly, the generic instability results of Schofeld, McKelvey and others can be thought of as implying a powerful status quo preserving incentive for collective decision-making, because of the likelihood of group members having meta-preferences for uncertainty avoidance (cf. Miller, 1983). Meta-preferences can provide theroetical guidance for an analytical framework within which preferences for particular alternative actions can be seen as only one of the elements which influence preference formation, broadly conceived.

The formal modelling of meta-preferences presents some interesting problems. One obvious approach would be to evaluate choice alternatives in multiattribute terms. Appended to the description of outcomes would be measures such as degree of support, locus of support, time needed to achieve agreement, and so forth. Choice, then, involves tradeoffs (as in Keeney and Raiffa's (1976) discussion of multiattributed utility functions) between "ordinary" preferences for outcomes and various types of meta-preferences. In the case of perfect information, the value for each person of the meta-preference component of his or her utility function may be specifiable. (If everyone's preferences are known, a measure of the

achieved degree of consensus can be attached to each outcome.) In that event, standard results for choice theory with multiattribute utility functions apply immediately; we are done; and the concept of meta-preferences adds little.

In the more common case, without perfect information, the values of the meta-preference components of the utility functions are not known at first because their values depend upon the preferences and choices of other decision makers. For example, the amount of time needed to reach a decision given a simple majority voting rule could be short if the preference profiles yield a Condorcet winner and longer if they contain cycles. The actual time will also depend upon participants' willingness to trade away preferred alternatives in order to shorten the process. In another example, if some individuals place a high value upon consensus, anything they learn through the decision process about the preferences of others may affect which alternative they support. Thus, in an environment of imperfect information, each individual's evaluation of alternatives in terms of meta-preferences depends upon the preferences and meta-preferences of others which can be ascertained only via the process of decision. Since some of the components of the utility vector depend upon the operation of the choice function, the standard multivariate utility approach cannot be applied.

A more promising alternative, at least with moderate or small sized decision-making groups, would be a game theoretical model. For example, if some group members are known to have strong meta-preferences for timelines while others are indifferent to time but strongly prefer some outcome, the latter can use delaying tactics to get the former to support the latter's preferred outcome.[9] The game structure enables preservation of the dependence of payoffs upon the choices made by others.

These comments are meant simply to indicate that the idea of meta-preferences could be formalized. it is not simply a catch-all term. Meta-preferences offer some possible new directions to the ongoing debate over why actual collective decision making does not degenerate into chaos more frequently.

We believe that recognition of meta-preferences and the modelling of them may lead to derivation of new results on the conditions for stability and provide insight into the dynamics of choice. However, this essay is

only intended to lay the groundwork for such an endeavor. In the future
we hope to formalize the implications of a theoretical approach which
encompasses meta-preferences.

NOTES

* An earlier version of this paper was prepared for delivery at the Public Choice Institute,
Halifax, Nova Scotia, July, 1984. We are indebted to Esther Martin and other staff at the
Word Processing Center for typing earlier versions of this manuscript and to Sue Pursche and
Dorothy Gormick for bibliographic assistance.

[1] Note that ours is a very general definition of choice process, e.g., one choice process would
be "use a two-thirds rule as long as the majority is willing to continue to debate, then switch
to a plurality rule".

[2] Technically we should distinguish between McKelvey's (1976) global instability (every-
thing in the top cycle set, even though the path may be a convoluted arc) and Schofield's
(1978) local instability (everything in a small neighborhood of a given point being in a cycle
including that point.) If we were to make this distinction this would lead to a seventh form
of instability which we might call "local preference instability".

[3] Our colleague Professor Wuffle has identified as a Riker-move the introduction of hot
fudge into a world without ice cream toppings. In the world of just ice creams, chocolate
beats vanilla. In the new two dimensional space, vanilla with hot fudge may be the Condorcet
winner.

[4] For example, a man or woman may have an abstract preference in principle not to become
a parent. When actually faced with a pregnancy, however, their preference in the concrete
may not only be different but also may be affected by the process. For instance, were deceit
or force involved, one might expect different feelings than were there honesty and no force.
The relevant choice set is more complex than simply (child, no child).

[5] Various procedural rules — and normative inhibitions upon the use of formal power — can
be understood in this light. Thus, although Sir John Kerr, Australia's Governor General, had
the formal authority to dismiss Gough Whitlam's Labour government in 1975, when he
exercised that power and thereby precipitated an election won by the Liberal Country Party,
even opponents of Whitlam (who would have desired the government fo fall) found unfair
the violation of the previously extent norms of procedure. In Australia, considerable effort
is now being spent rethinking the roles of the Senate and of the Governor-General.

[6] Miller (1983) has put forward a similar claim, though for quite different reasons (cf. our
discussion of the universalism norm below).

[7] A preference for certainty in the form of institutional maintenance may underly the
contemporary reluctance of losing presidential candidates to request recounts. This is not,
of course, the only plausible explanation for why a recount is not sought. "A candidate for
the Virginia House of Delegates lost a special election by one vote and did not ask for a
recount. His reasoning was simply that he *knew* who was counting the votes and that one vote
was as close as he could ever get". (Richard Smolka, personal communication, February
1981). (Consider also in this regard the Indiana congressional election still unresolved five
months after the 1984 election and after a series of straight party-line votes.)

[8] A universalism norm is not the same as an explicit global log-roll since the payoff to some voters may be simply in the form of a promise/commitment over a vaguely defined set of issues with no definite time of realization.

[9] A similar tactic is well-known among political activists, especially in a variant in which the activists prolong the length of meetings so that time-minimizers are motivated to maximize utility by departing from the meeting, thereby abstaining from the decision and guaranteeing the policy choice of the activists.

REFERENCES

Banks, Jeffrey S.: 1984, 'Sophisticated Voting Outcomes and Agenda Control', California Institute of Technology Social Science Working Paper 524 (April).

Bell, Colin E.: 1978, 'What Happens When Majority Rule Breaks Down? Some Probability Calculations', *Public Choice* 33 121–126.

Bowen, Bruce D.: 1972, 'Toward an Estimate of the Frequency of Occurence of the Paradox of Voting in U.S. Senate Roll Call Votes', in Richard D. Niemi and Herbert F. Weisberg (eds.), *Probability Models of Collective Decision Making*, Charles E. Merrill, Columbus, Ohio, pp. 181-203.

Buchanan, James and Tullock, Gordon: 1962, *The Calculus of Consent*, University of Michigan Press, Ann Arbor, Mich.

Chamberlin, John R., Cohen, Jerry, L., and Coombs, Clyde H.: 1984, 'Social Choice Observed: Five Presidential Elections of the American Psychological Association', *Journal of Politics* 46, 479–502.

Downs, Anothony: 1957, *An Economic Theory of Democracy*, Harper and Row, New York.

Feld, Scott and Grofman, Bernard: 1984, 'Net Preferences and Transitive Social Choice: A Little Single-Peakedness Can Go a Long Way'. Unpublished manuscript.

Feld, Scott, Grofman, Bernard and Miller, Nicholas: 1985, 'Cycle Lengths and other Features of Majority Preference in the Spatial Context', Paper Delivered at the Weingart Conference on "Models of Voting", California Institute of Technology, March 1985.

Ferejohn, John, Fiorina, Morris, and Packel, Edward W.: 1980, 'Nonequilibrium Solutions for Legislative Systems', *Behavioral Science* 25, 140–148.

Ferejohn, John A., McKelvey, Richard D., and Packel, Edward W.: 1984, 'Limiting Distribution for Continuous State Market Models', *Social Choice and Welfare* 1, 45–67.

Fiorina, Morris and Shepsle, Kenneth A.: 1982, 'Equilibrium, Disequilibrium, and the General Possibility of a Science of Politics', in P. C. Ordeshook and K. A. Shepsle (Eds.), *Political Equilibrium*, Kluwer-Nijhoff, Boston, pp. 49–64.

Glazer, Amihai, Grofman Bernard, and Owen, Guillermo: 1984, 'Imperfect Information Models of Two Candidate Competition', Presented at the Annual Meeting of the Public Choice Society, New Orleans, March, 1985.

Grofman, Bernard: 1983, 'The General Irrelevance of the Zero Sum Assumption in the Legislative Context', in Manfred Holler (Ed.), *Coalitions and Collective Action*, Physica-Verlag, Wuerzburg, pp.100–112.

Grofman, Bernard, Owen Guillermo, Glazer, Amihai and Noviello, Nicholas: 1984, 'A New Look at an Old Solution Concept' Presented at the Annual Meeting of the Public Choice Society, New Orleans, March 1985.

Hirschmann, Albert O.: 1980, *Shifting Involvements: Private Interest and Public Action*, Princeton University Press, Princeton, N.J.

Hoffman, Elizabeth and Packel, Edward: 1982, 'A Stochastic Model of Committee Voting with Exogenous Costs: Theory and Experiments', *Behavioral Science* **27**, 43–56.

Jeffrey, Richard: 1974, 'Preferences Among Preferences', *J. Philosophy* **71**, 377–392.

Keeney, Ralph L. and Raiffa, Howard: 1976, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley, New York.

Kramer, Gerald: 1977, 'A Dynamic Model of Political Equilibrium', *Journal of Economic Theory* **16**, 310–334.

Matthews, Donald R.: *U.S. Senators and their World*, University of North Carolina Press, Chapel Hill, N.C.

McKelvey, Richard D.: 1976, 'Intransitivities in Multi-Dimensional Voting Models and Some Implications for Agenda Control', *Journal of Economic Theory* **12**, 472–481.

McKelvey, Richard D.: 1983, 'Covering, Dominance, and Institution Free Properties of Social Choice', California Institute of Technology Social Science Working Paper 494 (September).

McKelvey, Richard D. and Wendall, Richard E.: 1976, 'Voting Equilibria in Multi-dimensional Choice Spaces', *Mathematics of Operations Research* **1**, 144–158.

Miller, Nicholas R.: 1980, 'A New Solution Set for Tournaments and Majority Voting: Further Graph-Theoretic Approaches to the Theory of Voting', *American Journal of Political Science* **24**, 68–96.

Miller, Nicholas R.: 1983, 'Pluralism and Social Choice', *American Political Science Review* **77**, 734–747.

Miller, Nicholas R., Grofman, Bernard, and Feld, Scott: 1985, 'Cycle Avoiding Trajectories and the Uncovered Set', Paper Delivered at the Weingart Conference on 'Formal Models of Voting', California Institute of Technology, March, 1985.

Niemi, Richard: 1969, 'Majority Decision Making with Partial Unidimensionality', *American Political Science Review* **63**, 489–497.

Niemi, Richard G.: 1983, 'Why So Much Stability?: Another Opinion', *Public Choice* **41**, 261–270.

Packel, Edward W.: 1981, 'A Stochastic Solution Concept for N-Person Games', *Mathematics of Operations Research* **6**, 349–362.

Rae, Douglas: 1969, 'Decision Rules and Individual Values in Constitutional Choice', *American Political Science Review* **63**, 40–56.

Riker, William H.: 1959, 'A Method of Determining the Significance of Roll Calls in Voting Bodies', in John Wahlke and Heinz Eulau, (eds.), *Legislative Behaviour* Free Press, Chicago, pp. 377–384.

Riker, William H.: 1961, 'Voting and the Summation of Preferences', *American Political Science Review* **55**, 900–911.

Riker, William H.: 1982, *Liberalism Against Populism*, Freeman, San Fransisco.

Roemer, Thomas and Rosenthal, Howard: 1978, 'Political Resource Allocation, Controlled Agent, and the Status Quo', *Public Choice* **33**, 27–45.

Schoefield, Norman: 1978, 'Instability of Simple Dynamic Games', *Review of Economic Studies* **45**, 575–594.

Schwartz, Thomas: 1981, 'The Universal Instability Theorem', *Public Choice* **37**, No. 3, 487–501.

Sen, Amartya, K.: 1966, 'A Possibility Theorem on Majority Decisions', *Econometrica* **34**, 491–499.
Sen, Amartya K.: 1977, 'Social Choice Theory: A Reexamination', *Econometrica* **45**, 53–88.
Shepsle, Kenneth A.: 'Institutional Arrangemewnts and Equilibrium in Multidimensional Voting Models', *American Journal of Political Science* **23**, 27–59.
Shepsle, Kenneth A. and Weingast, Barry R.: 1981, 'Structure-Inducement Equilibrium and Legislative Choice', *Public Choice* **37**, 503–519.
Shepsle, Kenneth A. and Weingast, Barry R.: 1982, 'Institutionalizing Majority Rule: A Social Choice Theory With Policy Implications', *American Economic Review* **72**, 367–367.
Shepsle, Kenneth A. and Weingast, Barry R.: 1984, 'Uncovered Sets and Sophisticated Voting Outcomes with Implications for Agenda Institutions', *American Journal of Political Science* **28**, 49–74.
Tullock, Gordon: 1970, *Toward a Mathematics of Politics*, University of Michican Press, Ann Arbor.
Tullock, Gordon: 1981, 'Why So Much Stability?', *Public Choice* **37**, 184–202.

School of Social Sciences,
University of California,
Irvine,
California,
U.S.A.