Discussion Paper No. 1196

Social Choice Theory, Game Theory, and Positive Political Theory[1]

by
David Austen-Smith[2]
and
Jeffrey S. Banks[3]

September 1997
Revised, October 1997

[1] Paper prepared for Nelson W. Polsby (editor), *Annual Review of Political Science, Volume 1*, Palo Alto: Annual Reviews Inc.

[2] Department of Political Science, Northwestern University, Evanston, Illinois 60208.
[3] Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125.

Abstract

In this paper, we consider relationships between the collective preference and the non–cooperative game–theoretic approaches to positive political theory. In particular, we show that an apparently decisive difference between the two approaches — that in sufficiently complex environments (e.g. high dimensional choice spaces) direct preference aggregation models are incapable of generating any prediction at all, whereas non–cooperative game–theoretic models almost always generate predictions — is indeed only an apparent difference. More generally, we argue that there is a fundamental tension when modeling collective decisions between insuring existence of well–defined predictions, a criterion of minimal democracy and general applicability to complex environments: while any two of the three are compatible under either approach, neither collective preference nor non–cooperative game theory can support models that simultaneously satisfy all three desiderata.

1. Introduction

Positive political theory is concerned with understanding political phenomena through the use of analytical models which, it is hoped, lend insight into why outcomes look the way they do and not some other way. Examples of such phenomena include which parties or candidates are elected at certain times, the bills adopted by legislative bodies, and when and how wars are fought between countries. Most of the models begin with the presumption that these phenomena are the result of the decisions made by the relevant individuals, be they voters and candidates in the first example, elected representatives and appointed ministers in the second, or heads of state in the third. Furthermore, these decisions are to a large extent a consequence of the preferences, beliefs and actions of these individuals.

Most models within positive theory are members of one of two families, although the demarcation line between the two is at times opaque;[1] indeed, one of the goals of the current paper is to provide our perspective on how these two families fit together and to argue that, at times, this line *should be* opaque. One class of models is motivated by the canonic rational choice theory of individual decision-making. In its simplest form, this theory assumes an individual has well-defined preferences over a given set of alternatives and chooses any alternative with the property that no other alternative in the set is strictly more preferred by the individual, that is, the individual chooses a "best" alternative. In politics, however, it is rarely the case that only one individual's preferences are relevant for any collective choice; even dictators are sensitive to at least some others in the polity. Consequently, the first family of models in positive political theory, which we associate with the methods of social choice, examines the possibility that individual

[1]We confine attention in the essay to formal models built around rational choice theory. Although by no means the only possible or extant sort of formal model for studying politics, rational choice models are far and away the modal sort.

preferences are directly aggregated into a collective, or social, preference relation which, as in the theory of individual decision–making, is then maximized to yield a set of best alternatives (where "best" is here defined as being most preferred with respect to the collective preference relation). If a set of best alternatives for a given method of aggregation necessarily exists, then we have an internally consistent model of observed collective choices as being elements from this set analogous to the model of individual choice, and it is in principle possible to ascertain whether the model does or does not provide a good explanation for what is observed in the real world of politics.

One missing piece of the direct aggregation story is the appropriate method by which the aggregation of individuals preferences into social preference is made, for example majority rule, unanimity, or dictatorship. Although this is typically dictated by explicit features inherent in the political phenomenon in question (e.g. plurality–rule elections), there are occasionally more amorphous situations in which the choice might best be considered in terms of a class of rules, all of which satisfy some critical properties of the situation (e.g. while the specific rules governing within–committee decisions may be more or less fluid, it is reasonable to suppose they all satisfy some notion of monotonicity in that more support for an alternative improves that alternative's chance of selection). The key here is to think of the aggregation rule as a particular feature of the model itself, and so appropriately left to the analyst to decide depending on what exactly she is attempting to explain. But whatever rule is appropriate for any given model, the model itself is well–specified as an explanatory model of political outcomes only to the extent that the rule yields best alternatives. Thus, for the direct preference aggregation approach to work as a general theory of politics, we need to determine the extent to which different aggregation methods insure the existence and characterization of best alternatives.

It is important to emphasise that the direct aggregation of individual preferences is not in general equivalent to indirect aggregation of preferences through the aggregation of individual actions. For example, an individual may have well–defined preferences over a set of candidates but choose to abstain in an election, or to vote strategically. Consequently, there is no *a priori* reason to suppose that elections lead to the same outcome that would occur if aggregation were directly over given preferences rather than indirectly over recorded votes. Of course, we expect the actions of purposive individuals to be intimately connected to their respective preferences, and such connections are the subject of the second principal family of models within positive political theory.

In this second family of models individuals are no longer passive participants in the collective decision making, but rather make individual choices of behavior which then jointly determine the collective choice of outcome. These models then naturally fall into the methodology of game theory. Here the fundamental moving parts of the model include the set of possible behaviors or strategies available to each of the participants, as well as a description of how any list of strategies relates to the set of outcomes. As with the preference aggregation rules in the first family of models, the appropriate choice by the analyst of the appropriate moving parts may be influenced by explicit features of the political phenomenon in question. Examples of such features would be the closed rule in parliamentary decision–making; presidential veto power; floor recognition rules in legislative debate and agenda–setting; germaneness rules for amendments; and party primaries for selecting electoral candidates. At other times, in contrast, there may not exist such explicit features to provide a roadmap to the "correct" model. Perhaps the quintessential example is the modeling of any sort of bargaining process for, say, within–committee or within–party decision making: questions the analyst must decide include who has the right to make what proposals and when, how to treat non–binding communication or the possibility of renegotiation, and so on and so forth. And the extent to which one model is "better than" another here depends in part on the empirical evaluation of their various predictions, the relative degree to which they generate insights into the workings of the institution, and so on.

Unlike with direct preference aggregation models, in the game theory models there is no presumption that collective outcomes are best elements relative to some underlying social preference relation. Rather, they are the consequences of a set of mutually consistent individual decisions within a given game. It is thus the composition of preferences and game structure that explains collective choices in this family of models and not, as in the first family, the application of an aggregation rule to preferences *per se*. A simple example illustrates the two classes of models.

There are three individuals, 1, 2 and 3, who must come to some collective choice from a set of three mutually exclusive alternatives, x, y and z. Let z be a given status quo policy and assume x and y are the only feasible alternatives to z. Individual 1 is assumed to prefer x most, followed by y and finally to consider z the worst option; individual 2 strictly prefers z to x, and strictly prefers x to y; finally, individual 3 prefers alternative y best, ranks x next best, and considers z to be the worst outcome. Then under simple majority rule, a direct preference aggregation model predicts a collective choice of x (since x is pairwise majority preferred to both y and z) and the explanation for such a prediction is in terms of x being uniquely best relative to the underlying aggregation rule.

Now suppose that instead of direct preference aggregation by majority rule, the choice of an alternative from the list x, y and z is determined according to the following game form, or set of rules. Individual 3 has the sole right to propose a take–it–or–leave–it change in policy away from the status quo, z. So individual 3 can either make no proposal, in which case z remains the collective choice, or can propose alternative x or y as the new policy; if individual 3 does offer a proposal, then the collective decision is reached via majority vote between z and the proposal. Under this institutional arrangement, it is clear that individual 3 offers alternative y, y defeats z, and y becomes the collective choice. Thus, in this example, the institutionally explicit model of indirect preference aggregation offers a distinct prediction of a collective choice, y, and this is supported by reference not only to individuals' preferences and the aggregation rule, but also to the institutional rules governing the choice process and the particular choices individuals make.

Prima facie, it seems reasonable to infer that results from the direct collective preference models have little, if any, relevance for those from the indirect game theory models. We consider such an inference inappropriate. In particular, we shall argue that the choice between collective preference and game–theoretic models cannot be predicated on a claim that the former typically fails to predict any choice, but the latter almost always does yield a prediction. Indeed, such a claim is true only to the extent that game–theoretic, in contrast to collective preference, models do not insist that all collective choices satisfy a certain normative requirement. First, however, we review in more detail the structure of the two families of models. In doing this the aim is not to provide any sort of comprehensive survey of positive political theory. Instead, we offer a (likely idiosyncratic) perspective on what goes into a formal model and how the two methodological approaches within positive political theory hang together. Our arguments are illustrated with four examples from the literature on legislative behavior.

2. The basic environment

The primitives of any rational choice–theoretic formal model of politics include a specification of a set of the relevant *individuals*, denoted N, a set of feasible *alternatives* or *outcomes*, X, and, for each individual in N, a description of her *preferences* over the set X. Examples of N are the members of a congressional district, of an interest group, of a parliamentary committee, of a jury, and so on. Corresponding examples of the alternatives from which such groups are, respectively, to choose are candidates for legislative office, alternative policies to the status quo, the guilt or innocence of a defendant, and so forth. Although many of the results below have analogues when the set of feasible alternatives is finite, for pedagogic reasons we will restrict most of our attention to the *spatial model*, in which the set X of feasible alternatives constitutes a nicely shaped geometric object, in that it is a closed, bounded, and convex subset of d–dimensional Euclidean space. Thus X could be the unit square, with an element of X then describing two numbers between zero and one (for instance, two tax rates); or X could be the two–dimensional unit simplex, with an element in X describing how one dollar is to be divided among three groups.[2] In general we will interpret the parameter d, the number of issues to be resolved, as a crude measure of the complexity of the collective decision problem at hand, in that for example single–issue decision problems are in a certain sense easier to solve than multi–issue decision problems.

We observe particular policies or outcomes being selected at different points in time and by different polities, as well as a variety of contemporaneous and exogenously–given parameters (more on the latter below). As described in the introduction, one of the goals of positive political theory is to explain these observed policy choices as functions of the observed parameters. A maintained hypothesis in most positive theory models is that these observed collective choices somehow reflect the underlying preferences (tastes, values, opinions, etc.) of some or all of the individuals in the relevant group. To formalize the idea of individual preferences, each individual in N is assumed to possess a binary preference relation on X, denoted

[2]Although there are three groups, the feasible set of alternatives is two–dimensional, since the three numbers have to add up to one; thus any two amounts completely determines the third.

R_i, where, for any two alternatives x and y, "xR_iy" is a shorthand for the statement "according to individual i, x is at least as good as y". From R_i one can define i's strict preference relation, P_i (where "xP_iy" reads "according to individual i, x is strictly better than y") and i's indifference relation, I_i (where "xI_iy" reads "according to individual i, x and y are equally good"). For any given group of individuals $N = \{1,...,n\}$, the list of all individuals' preferences is described by a *preference profile*, denoted $R = (R_1,...,R_n)$.

Restrictions of various sorts are placed on individual preferences to keep the models relatively tractable. The most basic of these are imposed to guarantee that individual–level decision problems are well–defined; these restrictions then imply that any subsequent negative results concerning collective "rationality" are not due to any individual–level irrationality, but are rather the consequence of the interaction between individuals with disparate preferences, as summarized by the profile R. Typical assumptions for the spatial model are given by the "four Cs", namely, that individual preferences are *complete* (for all x,y in X either xR_iy or yR_ix, or both); *consistent*, e.g. transitive preferences (if both xR_iy and yR_iz then xR_iz), *continuous* (if xP_iy then for any alternative z sufficiently close to y and w sufficiently close to x, wP_iz), and (strictly) *convex* (if xR_iy then for any distinct alternative z lying on the straight line between x and y, zP_iy). Together these four assumptions imply that for each individual the set of preference–maximizing alternatives or *ideal points*, $m(R_i)$, is necessarily non–empty and single–valued and that preferences decline as outcomes move away from this ideal point in any direction.[3] In the special case where the outcome space X is one–dimensional, such preferences are known as

[3]These assumptions are useful for keeping various models comparable. Strictly speaking, however, they are stronger than necessary for some of the results below and a little weaker than necessary for others (see later). Also, the assumptions are emphatically *not* the same as assuming "Euclidean" preferences, where outcomes equidistant from i's ideal point are judged to be indifferent by i. Euclidean preferences are a (relatively small) subset of those allowed for here.

single–peaked. Let $\mathcal{R}$ denote the set of preference profiles $R = (R_1,\dots,R_n)$ such that each R_i satisfies the above four assumptions.

3. Social choice

Suppose we hypothesize that the observed collective choices are in fact the best alternatives from, say, individual 1's perspective. We could then develop a model of how various parameters, for example 1's income or socio–economic background, influence 1's preferences and hence her best alternative in any set of choices, and subsequently test the model using the observables (i.e. realized choices and given parameter values). The basic premise in most positive political theory models, however, is that more than one individual's preference matter. Consequently, even if we have a firm grasp of how individual preferences depend on some list of exogenous parameters, we still need a theory of how the possibly different and conflicting preferences of individuals get translated into policy choices. Any such theory can be represented as a mapping from the set of preference profiles $\mathcal{R}$ into the set of outcomes X. Let c denote a generic theory of this sort, and refer to the mapping c as a *social choice correspondence*; thus $c(R) \subseteq X$ denotes the outcomes selected by the collective N when the preference profile is R, where in general we allow c(R) to be empty (that is, the theory does not make a prediction at R).

The difficulties in aggregating heterogeneous preferences have been known for quite a while, at least reaching back to Condorcet (1785), who formulated his famous paradox associated with majority rule: let there be exactly three individuals, $N = \{1,2,3\}$, and suppose their respective preferences are given by the profile

$$R^\circ = \begin{cases} xP_1yP_1z \\ yP_2zP_2x; \\ zP_3xP_3y \end{cases}$$

Here each alternative has a majority which prefers some other alternative, thereby causing a fundamental problem for the method of majority decision. More generally, we can think of the problem inherent in this profile as being that each alternative has, by symmetry of the preference profile, an identical claim on being the chosen outcome. Equivalently, we could say that *none* of the alternatives has such a claim, since for each there exists another which is preferred by all but one of the individuals. Therefore modeling how the collective decides in this case requires more structure than simply an assumption of aggregation by majority preference.

Say that a preference profile R exhibits the *Condorcet problem* if it is the case that for all x in X, there exists another alternative y in X such that the number of individuals strictly preferring y to x is at least $n-1$. Thus, when a profile exhibits the Condorcet problem, no one alternative stands apart as being a natural selection since for every alternative one can find another such that the latter is preferred by all but at most one individual to the former. Our first result shows that the existence and prevalence of such profiles depends on the "complexity" of the decision problem at hand:

Theorem 1 *(a) If $d \geq n-1$ then there exist a preference profile in $\mathcal{R}$ which exhibits the Condorcet problem. (b) If $d \geq 3(n-3)/2$ then almost all preference profiles in $\mathcal{R}$ exhibit the Condorcet problem.*[4]

[4]These results are simply re–statements of the non–existence results for q–rules, whereby x is ranked better than y if and only if at least q individuals strictly prefer x to y (with $n \geq q > n/2$), when $q = n-1$. Part (a) follows from Greenberg (1979) and part (b) follows from Banks (1995) and Saari (1997). Strictly speaking, (b) requires an additional technical assumption that individual preferences be representable by continuously differentiable utility functions and $n \geq 5$. Moreover, although a version of (b) holds for n equal to 3 or 4, the precise statement is a little more involved due to some special cases and, in the interest of continuity, we omit it here and in what follows: see Saari (1997) for details. Finally, while the qualifier "almost all" has a precise mathematical meaning, it suffices here to interpret

Therefore the only way to avoid such unpleasant profiles is to assume a simple enough decision problem, or else assume them away directly by positing a tighter restriction on individual preferences. Alternatively, for complex environments with unrestricted preferences any social choice correspondence must necessarily come to grips with the question of how collectives decide on outcomes when the preferences of the individuals exhibit a high degree of heterogeneity.

Theorem 1 has an immediate implication for social choice correspondences, once we identify a suitable analog to the Condorcet problem. Say that the correspondence c satisfies *minimal democracy* at the profile R if it is the case that x is not in c(R) whenever there exists an alternative y such that all but at most one individual prefers y to x.[5]

Theorem 1′ *(a) If* $d \geq n-1$ *then there exists a preference profile* R *such that either* c(R) *is empty or else* c *does not satisfy minimal democracy at* R. *(b) If* $d \geq 3(n-3)/2$ *then for almost all preference profiles* R *either* c(R) *is empty or* c *does not satisfy minimal democracy at* R.

Thus there exists a fundamental tension in preference–based theories of collective decision making, in that either existence of solutions or the notion of minimal democracy must be sacrificed at some, and at times most, preference

---

the caveat as meaning "except in some extremely unlikely circumstances".

[5]This definition derives from Ferejohn, Grether and McKelvey (1982). And it is worth pointing out that "minimally democratic" rules as defined here (such as majority rule) may not be very palatable in other respects. For instance, a minimally democratic rule might endorse a transfer of all individual 1's possessions away from 1 to everyone else in society if all individuals other than individual 1 strictly prefer such a reallocation to leaving 1 alone. See Sen (1970:ch.6) and the subsequent literature on the "liberal paradox" for discussion of how collective decision–making and individual rights can conflict.

profiles. We shall see below that the principle "negative" result of the collective preference approach can be viewed as sacrificing existence in the name of minimal democracy, whereas the principle "positive" result of the game theory approach sacrifices minimal democracy in the name of existence. A distinguishing feature of the two approaches therefore is the trade-off they make between these two concepts.

4. Collective preference

The collective preference approach to politics seeks to understand the properties of various methods for taking preference profiles into some sort of collective or social preference relation; such methods we term *preference aggregation rules.* As the basis for a positive model of political phenomena, the premise would then be that, analogous to models of individual decision making, the observed outcomes of the collective decision making are those that are judged to be optimal from the perspective of this social preference relation. If such optimal outcomes necessarily or frequently exist for different aggregation rules, it would then be an empirical question for the positive models as to which is the "right" rule for a given circumstance (and a philosophical question for the normative model).

Formally, a preference aggregation rule, denoted f, specifies for every profile R in the set of admissible profiles $\mathcal{R}$ a social preference relation, R_S (with strict aspect P_S and indifferent aspect I_S), where analogous to the interpretation of individual preference relations, the statement "xR_Sy" is read "based on the individual preferences R, and the method of aggregating them into a social preference relation f, x is judged to be at least as good as y" (where for notational simplicity we keep the dependence of R_S on R and f implicit). Thus a preference aggregation rule takes the individuals' preferences and generates a social or collective preference relation according to some procedure or rule, and so we can think of an aggregation rule f as a mapping from the set $\mathcal{R}$ of admissible preference profiles on X into the set, call it

B, of complete binary relations on X.[6] Common examples of preference aggregation rules include (1) *majority rule*: xP_sy if and only if the number of individuals strictly preferring x to y is more than half the population, (2) *unanimity*: xP_sy if and only if every individual strictly prefers x to y, and (3) *dictatorship*: there is some specific individual in N, say individual i, such that if i strictly prefers x to y then xP_sy.

Consistent with the theory of individual choice, models of direct preference aggregation posit that the alternatives selected from X are those that are best or maximal with respect to the underlying (social) preference relation; as earlier, we let m(b) denote the set of maximal elements with respect to an arbitrary element of the set B. Thus the mapping f takes preference profiles as input and generates elements in the set of binary relations, B, while the mapping m takes binary relations as input and generates elements in the set of outcomes X; see Figure 1.

FIGURE 1

$$\mathcal{R} \xrightarrow{f} B \xrightarrow{m} X$$

(Note that, unlike individual preferences, we have not imposed any structure to guarantee m(b) is a singleton). Composing the mappings f and m, we label the set of best elements given a profile R and an aggregation rule f as the *core* of f under R. Predictions from a direct preference aggregation model are thus core alternatives, and we can think of the core as itself defining a social choice correspondence, i.e. a mapping from preference profiles into outcomes.

A natural question to ask, then, is, when is the core of an aggregation rule f guaranteed to be non–empty? That is, when are there best elements in X as judged

[6]Despite the language, however, it is important to note that there is no sense in which it is presumed that societies *per se* have "preferences" in the same way as individuals; social "preferences" depend not only on the given list of individual preferences, but also on the aggregation rule in use.

by the social preference relation derived from R via f? From the Condorcet example above we know that this question is non–trivial when X is finite, since there the majority rule core is seen to be empty. Hence in this instance one in principle cannot explain any choice from this set it terms of it being the best alternative according to majority rule.

Two additional features of the Condorcet example deserve emphasis at this point. The first is that for other preference profiles the majority rule core *is* non–empty, as for instance when all individuals have the same preferences. Thus the non–existence of a majority rule core requires a sufficient amount of preference heterogeneity, as exemplified in the profile R° above. For a wide class of preference aggregation rules we can identify precisely when such heterogeneity exists and hence when core alternatives do and do not exist. Further, this characterization depends, as in Theorem 1 above, on the complexity of the decision problem at hand. Say that an aggregation rule f satisfies *monotonicity* if it is the case that for any x,y in X, if x is socially preferred to y at the profile R and the profile R' is such that x does not fall relative to y in any individual's ordering, then x remains socially preferred to y at the profile R'; and satisfies *neutrality* if the rule is symmetric with respect to alternatives (i.e. the names of the alternatives are immaterial): see Sen (1970) for formal definitions. Then we have the following:[7]

Theorem 2 *For any neutral and monotonic aggregation rule* f *there exists a number* $d(f) \geq 1$ *such that if* $d \leq d(f)$ *then the core of* f *is non–empty for all* R *in* $\mathcal{R}$.

In particular, when the outcome space is one–dimensional any neutral and

[7]See Greenberg (1979), Strnad (1984) and Schofield (1984). Austen–Smith and Banks (1997, ch.5) offer a self–contained development of the theory.

monotonic aggregation rule, including majority rule, will have a non–empty core.[8] Additionally, from Black (1958) we know that when n is odd the majority rule core point is unique, and possesses the well–known "median voter" characterization: if we align the individuals' ideal points along the one dimension from left to right, then the (unique) core alternative is simply the ideal point of an individual, say k, at which k and all those with ideal points at or to the left of k's ideal point constitute a majority, and k and all those with ideal points at or to the right of k's ideal point constitute a majority.

Thus settings in which an assumption of a unidimensional outcome space can be justified are readily amenable to empirical analysis under majority rule, in that we have all we could conceivably hope for: existence, uniqueness, and a straightforward characterization.[9] Alternatively (as we shall see below), often a theoretical model will presume a unidimensional outcome space to justify use of Black's theorem and thereby conveniently summarize a particular collective choice process which is but a part of the larger theoretical enterprise.

On the other hand, we know from the logic of Condorcet's example that the majority rule core will be empty whenever the preference profile R exhibits the Condorcet problem, and from Theorem 1 such profiles necessarily exist when the decision problem is sufficiently complex. In this sense, then, majority rule cannot provide a *general* theory of social decision making. Furthermore, Theorem 1′ suggests that any such general theory under the collective preference approach, i.e. any aggregation rule which has a non–empty core for all R in $\mathcal{R}$ and for any d, must

---

[8]Technically, this implies that there is insufficient preference heterogeneity in one dimension to construct a cycle as in the Condorcet example, whereas in two or more dimensions there does exist such heterogeneity: Schofield (1983).

[9]Romer and Rosenthal (1979) provide an excellent review of much of the empirical work based on the median voter theorem up to 1979, and argue that the results are equivocal with respect to whether it is indeed the median voter's preferences that determine political decisions. For a more positive assessment of the theorem's predictive value, see Bueno de Mesquita and Stokman (1994).

come at a price. This can be seen by the second important feature of the Condorcet example, which is that for certain aggregation methods core points exist even with the preference profile R°. For example, if we use the dictatorship rule and simply take the social preference relation to be the same as (say) individual 1's preference relation, then (since 1's most preferred alternative is x) there will necessarily be a core. Alternatively, under the unanimity method each alternative is judged to be indifferent to every other, and so all three alternatives are in the core (note that the latter prediction is useless from an empirical perspective).

Of course, both of these aggregation rules generate core correspondences which, when viewed as social choice correspondences, fail to satisfy minimal democracy. In fact, any aggregation rule for which existence is guaranteed invariably involves some combination of normatively unappealing (as in the case of dictatorship) and empirically unappealing (as with unanimity) qualities. To state this negative result formally, say that an aggregation rule is *minimally democratic* if x is judged socially preferred to y by the rule whenever all but at most one individual strictly prefers x to y. Then the following is a straightforward consequence of Theorem 1.

Corollary 1 (to Theorem 1) *For any minimally democratic aggregation rule,* (*a*) *the core is empty for some preference profiles when* $d \geq n-1$, *and* (*b*) *the core is empty for almost all preference profiles when* $d \geq 3(n-3)/2$.

From a modeling perspective, therefore, Corollary 1 tells us that any minimally democratic aggregation rule (e.g. majority rule) possesses a serious shortcoming as the basis for a general theory of politics, in that such a rule *in principle* cannot explain collective decision making in certain environments while simultaneously allowing some modest amount of latitude in the specification of individual preferences. Put another way, any explanatory theory of collective choice in complex environments based on a model of direct preference aggregation under minimal democracy must describe how individual preferences consistently live along the razor's edge of profiles that admit non–empty cores.

Now it is reasonable to consider the direct preference aggregation theory of collective choice as a formal theory of political decision making in terms of some notion of a "collective will", where the latter is reflected in the desiderata (including minimal democracy) defining the particular aggregation rule in use. As such, Corollary 1 renders any such conception of political decision making suspect (Riker, 1982): when there is no core, the view that observed policy choices embody a "collective will" seems hard to maintain since for any alternative there exists a policy that is socially preferable according to that same "will". And this inference is reinforced by the (somewhat unfortunately termed) "chaos theorems". Specifically, while we know that when there does not exist a core there must exist a social preference cycle, as in the Condorcet paradox above, McKelvey (1976, 1979) demonstrates further that in the spatial model such cycles are essentially all–inclusive for aggregation methods like majority rule. Thus his result implies that when social preference breaks down, in the sense of not admitting a core alternative, it breaks down completely: social preference cycles fill the space, and one can get from any alternative to any other (and back again) via the social preference relation. In general, individual preferences *per se* place almost no constraints on collective preference.[10]

Some have interpreted McKelvey's Theorem as predicting that anything can happen in politics (Riker 1980), meaning that political behaviour under minimally democratic institutions (in the technical sense of the term used here) is necessarily

[10]It is perhaps worth emphasising here that McKelvey's theorem says nothing about whether the core is empty; it only concerns the properties of social preferences (not choices) *given* the core is empty. See also Schofield (1984).

chaotic or unpredictable. We do not agree with this interpretation. The theory, as exemplified by Corollary 1 above, does not predict that anything can happen, *it does not predict anything at all*, which of course is the fundamental problem in employing the theory as a positive model of politics. The chaos result of McKelvey simply emphasises the impossibility of any general theory of political behavior based solely on the notion of preference aggregation under the constraint of minimal democracy and, from a normative perspective, implies that any hope of finding substantive content in the idea of a "collective will" with respect to policy choice is slender indeed. The "chaos" that McKelvey's Theorem speaks to is not from our perspective an equilibrium phenomenon, as for instance is found in various recent macroeconomic models (e.g. Grandmont 1985), but rather demonstrates how badly any minimally democratic social preference relation can behave. Hence attempts to render this an empirical prediction, and then ask questions such as "Why so much stability?" (Tullock, 1981) are moot.[11]

5. Structure and strategy

An alternative to the direct preference aggregation approach to understanding political behavior, is the class of *choice* aggregation models. All such models require a theory of how individuals make choices in collective decision making settings, and the most widely employed of these theories is non–cooperative game theory.

As before, we begin with a set of alternatives X and a list of individuals'

---

[11]One possible escape route from this argument is to weaken the second mapping in Figure 1; that is, rather than look solely for maximal elements with respect to the social preference relation, identify some other (with luck non–empty) set of alternatives. The two most well–traveled routes with respect to majority rule are the *top cycle set* (Schwartz 1972) and the *uncovered set* (Miller 1980) (actually, one could equivalently treat this as modifying f to generate the transitive closure relation and the covering relation, and then still use m). However McKelvey's Theorem shows that the former is just about the entire set when the majority rule core is empty, thereby nullifying any empirical content. The second option holds a higher promise; see McKelvey (1986).

preferences, summarized by the profile R. But now it is necessary to specify exactly what choices are available to individuals, to describe the outcome in X resulting from any given list of possible choices, and to offer a theory of how individuals' preferences and choices are related. Formally, the primitives of the model include, for each individual i in the polity N, a set of available *strategies* S_i, where a strategy is understood as a complete description of how an individual behaves conditional on every logically possible circumstance that might confront the individual. Analogous to a preference profile, we will label a specific list of strategies $s = (s_1,\ldots,s_n)$, one for each individual, as a *strategy profile* and we will let S denote the set of all possible strategy profiles. For the second component, an *outcome function* g specifies which alternative in X is chosen when a given list of strategies is chosen by the individuals; that is, the function g takes as input elements of S and gives as output elements of X, and so g(s) is an element of the set X. Taken together, G = (S,g) is known as a *game form*. Adding the list of individuals' preferences to the game form G yields the *game* (G,R).

Although the idea of a game form is somewhat abstract, a natural interpretation within political science is that game forms are succinct descriptions of institutions: strategy sets define the choices available to individuals under the rules of the institution, and outcome functions specify the collective decisions consequent on any list of such choices. Examples include various electoral procedures (e.g. plurality rule vs. proportional representation), voting procedures (e.g. amendment agendas vs. successive, open vs. closed rules), committee systems, executive vetoes, and the like that have been the focus of attention for much of the formal analysis in the recent past. Indeed, to the extent that these elements are readily observable game theoretic models permit a test of both outcomes *and* behavior. Furthermore, they allow the analyst to compare and make judgments about different institutions with respect to both the behavior they induce and the outcomes they generate.

Note that an individual's preferences over the set of outcomes X, together with the outcome function g, induce preferences for the individual over the set of strategy profiles S by equating a strategy profile with the outcome it ultimately produces. That is, a strategy profile s is judged to be at least as good as another s' if the outcome associated with s, namely g(s), is at least as good as that associated with s', $g(s')$. This allows for a more concise description of a game as simply the available strategy sets, $(S_1,...,S_n)$, together with the individuals' induced preferences over S; this is what is known as a game in *normal form.* We wish to maintain the "spatial" structure on the decision problem as before, so assume each S_i is a closed, bounded, and convex subset of d_i–dimensional Euclidean space, and that an individual's induced preferences on S satisfy the "4 Cs" from above.[12]

Finally, it remains to specify how individuals select strategies, where such a selection can be predicated on the specifics of the game form. Notice first that if one were to fix the strategy choices for all individuals other than, say, individual i, then i's problem of finding an optimal or preference–maximizing strategy is well–posed; in particular, for any list of others' strategies there would exist a uniquely optimal strategy for i. As well, for a certain class of games this optimal strategy for i is actually independent of the others' strategies, and so constitutes (in game theory parlance) a *dominant strategy.* The most well–known example of such a game is the Prisoners' Dilemma. For those games in which each individual has a dominant strategy there is then a straightforward theoretical prediction, namely, that each will adopt their dominant strategy.

For many games, however, dominant strategies do not exist. Consequently, any individual's optimal strategy will depend non–trivially on the choices made by others

[12]Note that these implicitly place continuity and convexity assumptions on the admissible set of outcome functions. Also, as before these assumptions are somewhat stronger than necessary.

and we therefore need a richer behavioral theory for describing how individuals within a game choose their respective strategies. The fundamental concept describing strategy choices under such circumstances is that of a *Nash equilibrium*: in a Nash equilibrium each individual's chosen strategy constitutes her optimal strategy, given the strategy choices of all of the other individuals. Thus a Nash equilibrium is a profile of strategies $s^* = (s_1^*,...,s_n^*)$ with the feature that no individual i in N can unilaterally change her strategy to something else, say s_i', and generate a strictly better outcome. Largely for this reason, any list of Nash equilibrium strategies is self–enforcing: if s^* is a list of Nash equilibrium strategies, then no individual i has any incentive to do anything other than use the prescribed strategy s_i^* when all others are likewise using their respective strategies under s^*.

For any game form G and behavioral theory h (e.g. dominant strategy or Nash), we can compose the mappings h and g to identify the set of *equilibrium outcomes*, i.e. those outcomes x in X such that x = g(s), where s is an equilibrium strategy profile at R according to the theory h. This composition then generates a particular social choice correspondence, in much the same way that a preference aggregation rule generated a social choice correspondence via the core. Notice however that, in contrast to the collective preference approach, here the influence of individual preferences on collective outcomes is occurring somewhat more indirectly, through the individuals' strategic choices under the constraints imposed by the game form. Thus we can think of the collective choices as being co–determined by the individuals' preferences and the specifics of the game form, and so any test of a game–theoretic model is a joint test of the behavioral theory embodied in the equilibrium concept and the institutional assumptions defining the game form. (Similarly, any test of a direct preference aggregation model is a joint test of the rule f and the presumption that choices reflect core outcomes under f.)

Letting h denote any arbitrary behavioral theory (mapping preference profiles into strategy profiles), we have an analogous diagram to that for direct preference aggregation; see Figure 2.

FIGURE 2

$$\mathcal{R} \xrightarrow{h} S \xrightarrow{g} X$$

Our next result shows that well–defined equilibrium outcomes for normal form games are the rule rather than the exception.[13]

Theorem 3 *Any normal form game satisfying the above assumptions has a Nash equilibrium.*

Indeed, the main problem with most game–theoretic analyses is not that Nash equilibrium fail to exist for any profile R but rather that there are too many of them. In this respect the concept of a Nash equilibrium is too weak in that it places few restrictions on what outcomes might be observed. For example, consider a plurality voting game in which any individual's strategy is a vote choice for one out a large finite set of alternatives and the outcome function selects the alternative that receives the largest total of votes. In this game any outcome can be supported as a Nash equilibrium outcome irrespective of individuals' preferences when there are at least three individuals: if everyone votes the same way at every stage then, under plurality rule, no single individual can change the outcome by switching their voting strategy and so may as well vote with the crowd. This is a silly prediction and it is ruled out by the additional requirement on the choice of equilibrium strategies that they be "perfect". "Perfection" is one example of an *equilibrium refinement.* Over

[13]Nash (1950), Debreu (1952). Theorem 3 can be used to prove existence of *mixed strategy* Nash equilibria in games with finite strategy sets.

the past twenty years or so, there has developed a literature on equilibrium refinements in which further assumptions are made on how individuals select strategies, thereby imposing further constraints on predictions generated by any model using these assumptions. Since the issues here, although important, are quite technical we do not pursue them further; the interested reader can consult, for example, Morrow (1994) or Fudenberg and Tirole (1991).

Returning to the notion of political institutions as game forms, suppose we have a set of possible institutions, $\mathcal{G}$ (with common elements $G = (S,g)$, $G' = (S',g')$, etc.), and suppose for ease of exposition that associated with each game form is a unique equilibrium, and hence (through the outcome function g) a unique element of the outcome set X. Then it is meaningful to compare institutions via the equilibrium outcomes that they support (Myerson, 1995, 1996). Such an approach has generated a rich set of empirical predictions regarding how institutional constraints influence political behavior and outcomes (e.g. Huber, 1996; Krehbiel, 1990). In addition, by focusing solely on the equilibrium outcomes this method also provides a foundation for making normative arguments regarding institutional choice (Austen-Smith and Banks, 1988; Cox, 1990; Myerson, 1993; Diermeier and Myerson, 1995; Diermeier and Feddersen, 1996; Persson, Roland and Tabellini, 1996).[14]

All of this is predicated to a certain extent on the existence result found in Theorem 3, namely, that regardless of the heterogeneity in individuals' preferences, a Nash equilibrium (or a refinement) will exist for a wide class of games. But the Nash equilibrium outcome correspondence is "just" another example of a social choice correspondence as defined in Section 3, and hence that the negative implications of

[14] A different perspective on games and institutions is provided by Calvert (1995). He argues that an appropriate way to view social institutions is as *equilibria* of a fairly loosely specified game played over time, rather than as game forms *per se*. This approach seems particularly useful for developing a theory of institutional evolution and stability.

Theorem 1 must hold true for the game theory approach to politics as it does for the collective preference approach. In fact, as with Corollary 1 for the latter, Theorem 1 has an immediate implication for game theory:

Corollary 2 (to Theorem 1) *Let* G *be any game form such that, for all profiles* R *in* $\mathcal{R}$, *the set of equilibrium outcomes is non-empty.* (*a*) *If* $d \geq n-1$ *then there exists a preference profile and alternatives* x *and* y *in* X *such that* x *is an equilibrium outcome and* y *is strictly preferred to* x *by at least* $n-1$ *individuals.* (*b*) *If* $d \geq 3(n-3)/2$ *then statement* (*a*) *is true for almost all preference profiles.*

(Note that as stated Corollary 2 holds for *any* behavioral theory, not just Nash.)

For sufficiently complex problems, therefore, game-theoretic models of indirect preference aggregation avoid the implications of Theorem 1 only by giving at least one individual some veto power, or dictatorial control, over at least one collective decision. Less prosaically, any appearance that Theorem 3 avoids the consequences of Theorem 1 is illusory; if surrendering minimal democracy is deemed acceptable to obtain equilibrium existence in game-theoretic models, then *a fortiori* it is acceptable to give up the condition to obtain core existence in social choice models. And as Figure 3 (we hope) makes clear, the collective preference approach and game theory approach should be considered two sides of the same coin, two complementary methods for generating social choice correspondences.[15]

[15]Figure 3 is a minor variation on the familiar Mount–Reiter diagram in economic theory: Reiter (1977).

FIGURE 3

$$\begin{tikzcd}
& B \arrow[dr, "m"] & \\
\mathcal{R} \arrow[ur, "f"] \arrow[rr, "c"] \arrow[dr, "h"'] & & X \\
& S \arrow[ur, "g"'] &
\end{tikzcd}$$

Where Corollary 1 and Theorem 3 differ is not so much in their respective existence claims as in their relative adherence to the minimal democracy condition, suggesting that to insist on this condition *a priori* is unproductive for development of a positive political theory. Any selection of game theory over collective preference as a method of political analysis, therefore, cannot be predicated on the issue of existence, but must depend on the problem of concern. For example, it is sensible to use game theory to understand the behavioural incentives induced by, and strategic properties of, various political institutions (e.g. voting with amendment agendas). On the other hand, collective preference theory is better suited for normative analysis of such properties and for decision problems with little or obscure detailed institutional structure (e.g. open rules in Congress).

6. Some more connections

The preceding argument centered on whether preference–based models invariably provide a prediction, and tapped the common underlying properties of the collective preference and game theory approaches to politics. It is also the case that in certain circumstances the two approaches are observationally equivalent, in the sense that the predictions of the collective preference approach and the game theory approach coincide.

Suppose we have a social choice correspondence c which is non–empty for all R in $\mathcal{R}$, and recall from Theorem 2 that the majority rule core *is* such a mapping when the outcome space X is one–dimensional. Say that a game form (S,g) *implements c in dominant strategies* if it is the case that for all R in $\mathcal{R}$, each i in N has a dominant strategy in S_i, and that the set of outcomes supported by such strategies is the same as the set of outcomes selected by c. In such an instance we can think of the game form (S,g) as performing indirectly the operation that c performs directly on preferences. We then have the following (Moulin 1980):

Theorem 4 *When $d = 1$ there exists a game form which implements the majority rule core in dominant strategies.*

As a simple example of such a game form, let $S_i = X$ for all i in N, with the outcome function g then being the selection of the median of the chosen alternatives. Then each player has a dominant strategy to choose her ideal point, since choosing any other alternative can only move the median away from her ideal point (and by convex preferences she prefers alternatives closer to her ideal point).[16]

Thus we can think of the median voter theorem as *either* an exercise in direct preference aggregation as in Section 4, or as the equilibrium outcome associated with a specific game form as in Section 5. In other words, the median voter theorem has a non–cooperative strategic foundation. Of course, from an empirical perspective it does not matter which model one adopts, since as constructed the models arrive at precisely the same predicted outcome.

[16]Alternatively, consider the following (intuitively described) game: there is a given status quo policy, q, and each individual proposes an alternative to q, say p_i, as the collective choice. Once all individuals have made a proposal, the collective decision is made from the set $\{p_1,\ldots,p_n,q\}$ according to the "amendment" procedure that first determines the majority vote between p_1 and p_2, then puts the winner against p_3, etc., until eventually the collective choice is given by the majority winner in a contest between the surviving proposal and the status quo q. Given the assumptions (X one–dimensional, etc.) it turns out that the unique (perfect) Nash equilibrium outcome to this game is the median voter's most preferred outcome.

Theorem 4 starts with the core concept, namely majority rule, and then shows the existence of a game form for which the outcomes of the two coincide when the outcome space is one–dimensional. For a class of game forms it is possible to go in the other direction as well; that is, for a given game form one can find a preference aggregation rule such that the core of this rule and the Nash equilibria of the game form coincide. Say that the outcome function g is *one–to–one* if g(s) is different from g(s′) whenever s is different from s′.

Theorem 5 *If the game form* (S,g) *is such that* g *is one–to–one, then there exists a monotonic preference aggregation rule* f *such that for all* R *in* $\mathcal{R}$ *the core outcomes under* f *are equivalent to the Nash equilibrium outcomes under* (S,g).[17]

(A proof for this theorem is given in Section 9, below.) Thus, when the outcome function is one–to–one, the distinction between the collective preference approach and game theory approach is again attenuated, in that one could have just as readily taken the former approach while employing a particular monotonic aggregation rule (the first model of the next section provides an example of a game form with such an outcome function). Of course, if $d \geq n-1$ and the game form G is such that Nash equilibria exist for all preference profiles in $\mathcal{R}$, then we know something more about this associated aggregation rule, namely, that this rule must fail to satisfy the criterion of minimal democracy (by Corollary 1).

---

[17]Even if g is not one–to–one in a game form G, it is possible to construct a preference aggregation rule f_G for which the core coincides with the set of Nash equilibrium outcomes: for any game (G,R) and for any x,y such that x is, and y is not, a member of the equilibrium set of outcomes, let f_G rank xP_sy; and for any other pair of alternatives, let f_G rank xI_sy. This rule is not very nice. In particular it violates both monotonicity (as defined in the text) and the weaker Arrovian condition of Independence of Irrelevant Alternatives.

7 Examples from legislative politics

In this section we review four theoretical models drawn from the literature on legislative politics, to illustrate some of the issues we have raised. The first two models have as their environment the basic multidimensional structure found in Section 2; the third is a one–dimensional model with the added twist of incomplete information; and the fourth is a distributional problem. The first three employ observed institutional structures as the foundation for their models, whereas the fourth is one in which no such observable structure exists.

7.1 Ministers and policy portfolios

Laver and Shepsle (1990) and Austen–Smith and Banks (1990) model governmental decision making in parliamentary democracies as being predicated on a decomposition of complex, multidimensional choice problems into a family of smaller, "dictatorial" choice problems. Each issue of the policy space is associated with exactly one cabinet minister, who has complete control over the outcome along this dimension. A minister is allowed to hold multiple issues or "portfolios", and so we define an *issue allocation* as an assignment of the set of policy issues to individuals such that every issue is assigned to some individual and no issue is assigned to more than one individual. Assuming for convenience that the set of outcomes X is "rectangular" or separable and so is itself decomposable, any issue allocation generates a well–defined game among the ministers, i.e. those individual legislators holding portfolios.[18] For example, if X is the unit square, and the issue allocation assigns dimension 1 to the first individual and dimension 2 to the second, i selects a point s_i in [0,1], with the collective outcome then being simply (s_1,s_2). Given our earlier assumptions on individual preference, Theorem 3 guarantees the existence of a Nash

[18]Banks and Duggan (1997) show how to get around this "rectangular" assumption.

equilibrium for any issue allocation. And notice that, by definition of an issue allocation and the rules governing how issue–by–issue choices are made, the outcome function for the game form described here is one–to–one. So, by Theorem 5, any equilibrium outcome in the portfolio allocation model is also a core alternative for a naturally defined monotonic preference aggregation rule (see the proof in Section 9 for a construction).

It is not hard to see that the location of the equilibrium outcomes will depend (*inter alia*) on the particular issue allocation, in that different allocations typically give rise to different equilibrium outcomes. Thus we might expect individuals in N to have preferences over the allocations *per se*, in which case the structure induced explanation of particular policy decisions is in a certain sense incomplete. Austen–Smith and Banks (1990) and Laver and Shepsle (1990) explore various concepts of the core applied to issue allocations; that is, given that policy outcomes are determined by the relevant allocations, individuals' induced preferences over allocations can be derived and we can look for cores with respect to these induced, rather than the primitive, preferences. Since the set of possible equilibrium outcomes under issue allocation models is much smaller than the set of all outcomes, "allocation" cores can exist more often than in the direct preference aggregation models. Further, allocation cores then yield predictions on both the distribution of decision making responsibility in a legislature and on the policy outcomes supported by such allocations. In particular, for some distributions of preferences the model predicts minority coalition governments.[19]

The use of the core here to study the choice of issue allocation among legislators is largely to avoid modeling the specifics of government formation. That is, there is a presumption that in this instance the core adequately captures the

[19]Laver and Shepsle (1996) take these and other predictions to the data on post–WWII coalitional governments in parliamentary systems with some degree of success.

possible outcomes from a variety of conceivable game forms describing how legislative bargaining over policy responsibilities. An alternate approach is to model this government formation process explicitly (e.g. Austen–Smith and Banks 1988; Baron 1991).

7.2 Committees

The motivation for the preceding structural approach to legislative decision making comes from Shepsle's (1979) model of decision making in the US Congress. Although the principal formal result is anticipated by Black and Newing (1951) and Kramer (1972), Shepsle (1979) was the first to provide a rigorous model of policy choice via legislative committees. Shepsle's insight was that a committee system, such as that in the US Congress, essentially decomposes a complex high–dimensional choice problem into a sequence of simpler low–dimensional problems. At this point, something akin to Theorem 2 can be invoked for *within* committee decision making, with the overall outcome then being (as in the model of 7.1) the cumulation of committee decisions.

A *simple committee system* is an institutional arrangement whereby, for each of the d issue dimensions, there is a unique committee, or subset of legislators, responsible for determining the collective choice on the dimension. A *committee allocation* is then an assignment of the set of individuals (in this case, legislators) into the d committees such that (a) no committee is empty, and (b) no individual is on more than one committee. For simplicity, assume again that the policy set X is separable, and that there are an odd number of individuals in each committee. Now suppose that each committee plays "Nash" against the others, taking the decisions on the other issues as given, and uses the majority method of preference aggregation to make its own decision. An equilibrium outcome, say, $x^* = (x_1^*,\dots,x_d^*)$ is an element of X such that for each dimension j, x_j^* constitutes the ideal outcome along dimension j for the median member of the j^{th} committee given the choices $\{x_i^*\}$ of all committees i other than j. Shepsle (1979) shows that an equilibrium outcome exists for any committee allocation in a simple committee system, thereby providing an institutionally predicated explanation for legislative policy choices in multidimensional spaces.

Although we used the phrase "Nash" in the above description, it should be noted that Shepsle's analysis is not strictly speaking game–theoretic in the sense of Section 4 above: individuals do not have strategy sets, Nash equilibria are not identified, etc. As such, his equilibria might be better understood as an example of core outcomes with respect to a particular aggregation rule (Diermeier, 1997). On the other hand, since Shepsle's model employs Theorem 2 and the median voter theorem, it appears possible to invoke a result along the lines of Theorem 4 and provide an explicit non–cooperative foundation for the equilibria his model generates. Indeed, when individual preferences are separable across the d dimensions or issues (i.e. an individual's most preferred alternative on any one dimension is independent of choices on other dimensions), one can use *exactly* the game form following Theorem 4 to implement Shepsle's equilibria in dominant strategies. When individual preferences are not separable such dominant strategy implementation will not occur (Zhou 1991). However, it is easily seen that any equilibrium in Shepsle's sense is a *Nash* equilibrium outcome of the aforementioned game, an equilibrium in which each individual again selects her ideal point (but where now this point depends on the choices of others). Therefore it is possible that this (or some other) game form implements Shepsle's equilibria in Nash equilibrium (as opposed to dominant strategies).[20] The existence of such a game form would then provide an explicit

[20]The remaining issue here concerns whether there exist other Nash equilibria to the game being used to Nash implement Shepsle's equilibrium set of outcomes. That is, Shepsle's equilibria may form a strict subset of the set of Nash equilibrium outcomes on the (implementing) game, whereas Nash implementation requires these sets to

non–cooperative foundation for Shepsle's equilibria, thereby eliminating the distinction between being a model of the social choice sort or one of the game theory sort; as with the median voter theorem, we can think of the model as both, and hence as either.

7.3 Open and closed rules

The committee system analyzed above rationalizes the use of such structures, in particular the deference paid to a select subset of individuals on certain issues, in terms of their equilibrium–generating properties. An alternative explanation for this type of deference centers on specialization and information. In fact, in the game–theoretic model of Gilligan and Krehbiel (1987) discussed here such deference cannot be explained as in Shepsle (1979), for in their model the outcome space has but a single dimension. Rather, this deference comes about as a rational response to a presumed informational advantage held by the committee.[21]

As mentioned let the set of outcomes, X, be one–dimensional, and let a committee of individuals exist whose purpose is to offer a proposed alternative to the current status quo policy. Here we wish to distinguish between *policies* and *outcomes*, so as before let x denote a typical outcome in X and let p denote a typical proposal or bill. Assume bills are also one–dimensional objects, and that the link between bills and outcomes is simply $x = p+t$, where the parameter t is some number between zero and one. The motivating assumption of the model is that t is known to the committee members but unknown to everyone else. Thus if the policy

coincide.

[21]The game Gilligan and Krehbiel analyze is one of *incomplete information*, and so technically falls outside the bounds of those discussed in Section 5. However Harsanyi (1967–8) shows how to extend the concepts of normal form games and Nash equilibrium to such environments (i.e. Bayesian games and Bayesian Nash equilibrium). We will not worry about such matters here: the interested reader should consult Fudenberg and Tirole (1993).

p is adopted, then the actual outcome is given by $x = p+t$ and this consequence is known surely only to the committee.

Both the committee and the remaining individuals in the legislature are modeled as single actors, so let C denote the former and F (for "floor") denote the latter; the theoretical legitimacy of such a modeling choice is discussed below. Two procedures are employed to determine a final policy, thereby yielding two different game forms: under a *closed rule* C is permitted to make a take–it–or–leave–it proposal whereby, if this proposal is rejected by F, the status quo policy p_o remains in effect. In contrast, under an *open rule* C again makes a proposal but now F can select any policy it wants; in particular, F is not restricted to choosing between C's proposal and the status quo. Thus under the open rule the proposal by the committee has no *substantive* content, in that it will not directly affect the floor's chosen policy. However the committee's proposal may have *informational* content due to the fact that C has ability to make its proposal depend on, or a function of, the true value of the parameter t. In particular, if the floor speculates that the committee is offering different proposals for different values of t, then upon observing one of these proposals the floor can make a better inference about what the value of t is (that is, better than F's *prior* belief). Under the closed rule, on the other hand, the committee's proposal can have both substantive *and* informational content.

Specific functional forms are assumed for C's and F's preferences, as well as a uniform prior belief concerning t for F. Even with these niceties, however, multiple equilibria exist under either procedure, and so Gilligan and Krehbiel are forced to make certain selections from the set of equilibria. Given these, they are able to show that, for certain values of the parameters, the floor actually has a preference (in terms of its *ex ante* expected payoff) for the closed rule over the open, even though the former allows the committee to bend outcomes in their preferred direction due to the monopoly control over the agenda C commands under this rule. The logic of this result follows from the fact that at times the loss to F from surrendering some control over the agenda is outweighed by the presence of an informational gain. That is, when the committee is assured some distributional gain under the closed rule, the proposals it offers signal more information regarding t to the floor. Since the floor is assumed risk–averse, therefore, the more information it has about the consequences of legislation, the better off it becomes. Note that this induced preference for the closed rule over the open rule would never occur under complete information, since then the floor can always do better by maintaining a greater amount of control over the final policy. Hence, informational asymmetries are seen as an alternative explanation for the existence of deference to committees, here in the form of closed rules.

Although perhaps not immediately apparent, the legitimacy of treating the committee and the floor as unitary actors rests largely on Black's median voter theorem. That is, because of the assumption of one–dimensional outcome and bill spaces, the presumed relationship connecting bills to outcomes preserves single–peakedness of preference profiles. Consequently, under majority rule within the committee and the legislature as a whole it is legitimate to identify the committee and the floor with their respective medians. And without these assumptions what occurs under open rule given a proposal or even what proposals are offered, and therefore what the relevant welfare comparisons across rules are, is unclear. One way to see this is to imagine the same set–up as above but where the issue space is *two*–dimensional rather than one, and where we break the floor into its constituent parts (i.e. a set of individuals with well–defined preferences over X). Then under the open rule it is not clear exactly what would happen once a proposal has been made by the committee, or rather, a well–defined game form remains to be posited. Similarly, under *either* rule it is not immediately obvious how to model the formation of a committee proposal. Such troublesome issues are conveniently side–stepped in

Gilligan and Krehbiel by the assumption of a one–dimensional policy space, a policy space with (as Theorem 2 notes) a majority rule core.

7.4 Bargaining

Finally we consider a pure distributional problem: there is a single dollar to be divided among n individuals, (so we let X denote the set of all such divisions), where each individual possesses "selfish" preferences in the sense of only caring about their own amount. Baron and Ferejohn (1989) model the determination of an element in X as occurring through the following dynamic bargaining game: in period 1 an individual is randomly selected from N, and makes a proposal x^1 in X, after which all individuals vote to accept or reject x^1. If x^1 is accepted by a majority of voters the game ends with x^1 as the outcome; otherwise the game moves to period 2, in which an individual is randomly selected to offer a proposal x^2 in X, and so on. The process continues until a proposal is accepted. In each period individuals are equally likely to be selected proposer, and individuals impatient and share a common discount factor δ, $0 < \delta < 1$, so that if the t^{th} proposal, x^t, is accepted individual i's payoff is worth simply $\delta^{t-1}x_i^t$ as evaluated at the start of the game.[22]

There are many Nash equilibria to this game. Baron and Ferejohn focus on a particularly simple class of equilibria, perfect stationary (essentially, history–independent) equilibria, which they show to exist and have the following qualitative properties: when an individual is selected, she proposes a split of the dollar in which she keeps the lion's share, (n–1)/2 others receive equal smaller shares,

[22]The central institutional feature of the Baron and Ferejohn model, sequential bargaining, has subsequently been applied to a wide variation of problems: see for example, Baron (1994) or Diermeier and Feddersen (1996). Alternative non–cooperative models with distributional and policy dimensions include Groseclose and Snyder (1996) and Snyder (1990). Note that the actual outcome space is X together with {1,2,...}, with the latter representing the time at which a proposal is accepted.

and the remaining (n–1)/2 individuals receive nothing; this proposal is then accepted by the individuals receiving a positive amount. Therefore the very first proposal is accepted, avoiding any (costly) delay, and *ex post* only a bare majority of individuals receive positive amounts (although *which* majority is uncertain *ex ante*).

The purely distributive politics game is the least tractable from a direct preference aggregation approach: since any distributional problem has dimensionality d = n–1 when there are n individuals (once n–1 shares are determined, the n^{th} share is given by the residual) and since preferences are selfish, the core is surely empty under any minimally democratic preference aggregation rule and so offers no prediction. In contrast, the Baron and Ferejohn sequential bargaining model supports a well–defined prediction that only minimal majorities will garner positive amounts of the dollar. At the same time, it is worth noting for any equilibrium allocation x^* there is a distinct allocation x' that n–1 individuals strictly prefer to x^*. Consequently, as we argued more generally in Section 5 must be the case, equilibrium collective choice through the Baron and Ferejohn bargaining process *prima facie* violates minimal democracy.

As we have already observed, empirical observation provides little in the way of structure that points to the "right" game form for modeling multilateral bargaining processes. Given this, analysts lean toward specifying the most parsimonious strategic model capable of supporting equilibria. Any judgement regarding the value of such a model then rests on the extent to which the equilibrium predictions yield empirical and conceptual insight regarding the forces at work. On the other hand, there is a question regarding why we might expect legislators to adopt stationary strategies. The importance of the "perfect stationary" equilibrium refinement lies in the fact that, as Baron and Ferejohn demonstrate, if legislators are sufficiently patient their model is subject to a folk theorem under which *any* allocation of the dollar can be supported as a perfect (albeit not stationary) Nash equilibrium.

Therefore, although no minimally democratic direct preference aggregation model can make a prediction (and not, as we keep insisting, that such a rule predicts anything can happen), the presence of a folk theorem in the absence of stationarity really does say that anything can happen in the Baron and Ferejohn bargaining model.

8. Conclusion

This brief essay makes no claim to be a general review or survey of positive political theory as a whole. Rather, we have tried to articulate some connections between the two main approaches to rational actor model–building in political science: direct preference aggregation (social choice theory), and indirect preference aggregation through the aggregation of choices in strategic settings (non–cooperative game theory). And in so doing, we have implicitly argued that the historical shift away from direct preference aggregation models toward institutionally more explicit strategic models of collective choice cannot reflect any methodological discontinuity.

Our main argument is that an apparently decisive difference between the two approaches — that in sufficiently complex environments direct preference aggregation models are incapable of generating any prediction at all, whereas non–cooperative game–theoretic models almost always generate predictions — is indeed only an apparent difference. In fact the distinction between the two sorts of model in this regard turns out to hinge critically on the extent to which a property of minimal democracy is required. If we insist that all choices must be minimally democratic (i.e. if at least all but one member of the polity strictly prefers an alternative x to another y, then y should not be chosen when x is available), then no game–theoretic model incorporating the requirement will fare any better in regard to yielding some kind of prediction than any similarly constrained collective preference model. On the other hand, if we wish our collective choice models, whether direct preference aggregation or game–theoretic, to yield predictions in all environments, then *necessarily* the models must violate minimal democracy. Equivalently, if we require our collective choice models to yield predictions and satisfy minimal democracy, then necessarily the environment must be kept relatively simple (i.e. low dimensional).

9. A proof for Theorem 5

Let $G = (S,g)$ and let $Im\{G\} = g(S)$ be the image of S under g. Define f by: (i) if $x \in Im\{G\}$ and $y \notin Im\{G\}$, then $xP_s y$ for all R; and (ii) if $x,y \in Im\{G\}$, then $xI_s y$ for all R. For $x,y \in Im\{G\}$, say that x and y are *comparable* if there exists $i \in N$, $s_i, s_i' \in S_i$, $s_{-i} \in S_{-i}$ such that $x = g(s_i,s_{-i})$ and $y = g(s_i',s_{-i})$ (where s_{-i} denotes the profile of all individuals' strategies except for individual i, etc.). Since g is one–to–one, this individual is unique, so for comparable x,y let i(x,y) denote this individual. For all non–comparable $x,y \in Im\{G\}$, let $xI_s y$ for all R, and for all comparable $x,y \in Im\{G\}$ let $xR_s y \Leftrightarrow xR_{i(x,y)}y$. Then R_s is a complete binary relation for all preference profiles R. To see that f is monotonic, note that if $x \in Im\{G\}$ and $y \notin Im\{G\}$ then $xP_s y$ for all R and hence monotonicity holds here. If $x,y \in Im\{G\}$ and $xP_s y$ then it must be that x and y are comparable, and that $xP_{i(x,y)}y$. But then under any new profile R′ satisfying the antecedent we still must have $xP'_{i(x,y)}y$, and hence x socially preferred to y remains true.

Fix a profile R arbitrarily. Let $\mathcal{C}$ denote the core of f at R and let $\mathcal{E}$ denote the set of Nash equilibrium outcomes of G at R.

(a) To see: $\mathcal{C} \subseteq \mathcal{E}$. If $x \in \mathcal{C}$ then $xR_s y$ for all $y \in X$, in which case it must be that $x \in Im\{G\}$, or $x = g(s)$ for some $s \in S$. For any $j \in N$ and $s_j' \in S_j$, $z \equiv g(s_j',s_{-j})$ is comparable to x, and so $x \in \mathcal{C}$ implies $xR_s z$ and hence $xR_j z$. But then s is a Nash equilibrium of G at R, and hence $x \in \mathcal{E}$.

(b) To see: $\mathcal{E} \subseteq \mathcal{C}$. Let $x = g(s)$; then $xR_s y$ for all $y \notin Im\{G\}$ and all non–comparable y. If in addition s is a Nash equilibrium, then for all comparable y $\in Im\{G\}$, $xR_{i(x,y)}y$, and thus $xR_s y$ for these outcomes as well. Therefore $xR_s y$ for all $y \in X$, and hence $x \in \mathcal{C}$.

Because the profile R was chosen arbitrarily, (a) and (b) together complete the proof.

<u>References</u>

Austen–Smith D, Banks J. 1988. Elections, coalitions, and legislative outcomes. *American Political Science Review* 82:405–422

Austen–Smith D, Banks J. 1990. Stable governments and the allocation of policy portfolios. *American Political Science Review* 84:891–906

Austen–Smith D, Banks J. 1997. *Positive Political Theory I: Collective Preference.* Forthcoming with University of Michigan Press, Ann Arbor.

Banks J. 1995. Singularity theory and core existence in the spatial model. *Journal of Mathematical Economics* 24:523–536

Banks J, Duggan J. 1997. Existence of Nash equilibria on convex sets. *Discussion Paper, Department of Political Science, University of Rochester.*

Baron D. 1991. A spatial bargaining theory of government formation in parliamentary systems. *American Political Science Review* 85:137–64.

Baron D. 1994. A sequential choice perspective on legislative organization. *Legislative Studies Quarterly* 19:267–296

Baron D, Ferejohn J. 1989. Bargaining in legislatures. *American Political Science Review* 83:1181–1206

Black D. 1958. *The Theory of Committees and Elections.* London: Cambridge University Press

Black D, Newing R. 1951. *Committee Decisions with Complementary Valuations.* London: William Hodge

Bueno de Mesquita B, Stokman F. (eds) 1994. *European Community Decisionmaking: Models, Applications and Comparisons.* New Haven: Yale University Press.

Calvert R. 1995. The rational choice theory of social institutions: cooperation, coordination, and communication. In *Modern Political Economy: Old Topics, New Directions*, ed. J Banks, E Hanushek, 216–28. New York: Cambridge University Press

Condorcet, M. 1785 *Essai sur l'Application de l'Analyse a la Probabilite des Decisions Rendues a la Pluralite des Voix.* Paris.

Cox, G. 1990. Centripetal and centrifugal incentives in electoral systems. *American Journal of Political Science* 34:903–35.

Debreu G. 1952. A social equilibrium existence theorem. *Proceedings of the National Academy of Science* 38:886–893

Diermeier D. 1997. Explanatory concepts in positive political theory. *Discussion Paper, Graduate School of Business, Stanford University.*

Diermeier D, Feddersen T. 1996. Disciplined coalitions and redistribution: the effect of the vote of confidence procedure on legislative bargaining. *Discussion Paper 1171, Math Center, Northwestern University.*

Diermeier D, Myerson R. 1995. Lobbying and incentives for legislative organization. *Discussion Paper 1134, Math Center, Northwestern University*

Ferejohn J, Grether D, McKelvey R. 1982. Implementation of democratic social choice functions. *Review of Economic Studies* 49:439–46.

Fudenberg D, Tirole J. 1993. *Game Theory.* Cambridge: MIT Press

Gilligan T, Krehbiel K. 1987. Collective decision–making and standing committees: an informational rationale for restrictive amendment procedures. *Journal of Law, Economics and Organization* 3:287–335.

Grandmont J–M. 1985. On endogenous competitive business cycles. *Econometrica* 53:995–1045.

Greenberg J. 1979. Consistent majority rules over compact sets of alternatives. *Econometrica* 47:627–636

Groseclose T, Snyder J. 1996. Buying supermajorities. *American Political Science Review* 90:303–315

Harsanyi J. 1967–8. Games with Incomplete Information Played by Bayesian Players, I–III, *Management Science* 14: 159–182, 320–334, 486–502

Huber J. 1996. *Rationalizing Parliament.* Cambridge:Cambridge University Press.

Kramer G. 1972. Sophisticated voting over multidimensional choice spaces. *Journal of Mathematical Sociology* 2:165–180

Krehbiel K. 1990. *Information and Legislative Organization.* Ann Arbor:U.Michigan Press.

Laver M, Shepsle K. 1990. Coalitions and cabinet government. *American Political Science Review* 84:873–890

Laver M, Shepsle K. 1996. *Making and Breaking Governments.* New York: Cambridge University Press

McKelvey R. 1976. Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory* 12:472–482

McKelvey R. 1979. General conditions for global intransitivities in formal voting models. *Econometrica* 47:1086–1112

McKelvey R. 1986. Covering, dominance, and institution free properties of social choice. *American Journal of Political Science* 30:283–314

Miller N. 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

Morrow J. 1994. *Game Theory for Political Scientists.* Princeton NJ: Princeton University Press

Moulin H. 1980. On strategy–proofness and single–peakedness. *Public Choice* 35:437–55

Myerson R. 1993. Incentives to cultivate favored minorities under alternative electoral systems. *American Political Science Review* 87:856–69.

Myerson R. 1995. Analysis of democratic institutions: structure, conduct and performance. *Journal of Economic Perspectives* 9:77–89.

Myerson R. 1996. Economic analysis of political institutions: an introduction. *Discussion Paper 1155, Math Center, Northwestern University.*

Nash J. 1950. Equilibrium points in N–person games. *Proceedings of the National Academy of Sciences* 36:48–49

Persson T, Roland G., Tabellini G. 1996. Separation of powers and accountability. *Discussion Paper 612, Institute for International Economic Studies, Stockholm.*

Plott C. 1967. A notion of equilibrium and its possibility under majority rule. *American Economic Review* 57:787–806

Reiter S. 1977. Information, incentives and performance in the $(\text{new})^2$ welfare economics. *American Economic Review* 67:219–37.

Riker W. 1980. Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review* 74:432–446

Riker W. 1982. *Liberalism vs Populism.* San Francisco: Freeman.

Romer T, Rosenthal H. 1979. The elusive median voter. *Journal of Public Economics* 12:143–170

Saari D. 1997. The generic existence of a core for q–rules. *Economic Theory*

Schofield N. 1983. Generic instability of majority rule. *Review of Economic Studies* 50:695–705

Schofield N. 1984. Social equilibrium and cycles on compact sets. *Journal of Economic Theory* 33:59–71

Schwartz T. 1972. Rationality and the myth of the maximum. *Nous* 7:97–117

Sen A. 1970. *Collective Choice and Social Welfare.* Amsterdam: North–Holland

Shepsle K. 1979. Institutional arrangements and equilibrium in multidimensional voting models. *American Journal of Political Science* 23:27–59

Snyder J. 1990. On buying legislatures. *Economics and Politics* 3:93–109.

Strnad J. 1984. The structure of continuous–valued neutral and monotonic social functions. *Social Choice and Welfare* 2:181–196

Tullock G. 1981. Why so much stability? *Public Choice* 37:189–202

Zhou L. 1991. Impossibility of strategy–proof mechanisms in economies with pure public goods. *Review of Economic Studies* 58:107–20