Games and Economic Behavior **37,** 40–78 (2001)
doi:10.1006/game.2000.0827, available online at http://www.idealibrary.com on IDEAL®

Voting for Voters: A Model of Electoral Evolution[1]

S. Barberà[2]

Departament d'Economia i d'Història Econòmica and CODE, Universitat Autònoma de Barcelona, Barcelona, Spain

M. Maschler[3]

Department of Mathematics and Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem, Jerusalem, Israel

and

J. Shalev[4]

107 Hamelachim Street, Ramat Hasharon, Israel 47286

Received March 31, 2000

THIS PAPER IS DEDICATED TO DORIT BAHIR MASCHLER (1972–2000). DORIT SOUGHT BEAUTY AND WAS GENUINELY INTERESTED IN HUMAN SOCIETIES.

We model decision problems faced by the members of societies whose new members are determined by vote. We examine a simple model: the founders and the candidates are fixed, the society operates and holds elections for a fixed number of periods, one vote is sufficient for admission, and voters can support as many candidates as they wish. We show through theorems and examples that

[1] The authors thank R. J. Aumann, C. d'Aspremont, D. Granot, S. Hart, and A. Perea for various comments and suggestions that helped improve this paper.

[2] E-mail: sbarbera@cc.uab.es. This author gratefully acknowledges partial support from Spanish Ministry of Education DGICYT Grant PB-98-0870 and Comissionat per a Universitats i Recerca Grant SGR98-62.

[3] E-mail: maschler@vms.huji.ac.il. This author gratefully acknowledges DGICYT, Spanish Ministry of Education, for Grant SAB95-0276 that enabled him to participate in this research.

[4] E-mail: shalev@witty.com. This author gratefully acknowledges the support for his research given by the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. This research was done when the author was working at CORE.

0899-8256/01 $35.00

interesting strategic behavior is implied by the dynamic structure of the problem. In particular, the vote for friends may be postponed, and it may be advantageous to vote for enemies. We characterize all pure strategy Nash equilibria outcomes and show that they can also be obtained as subgame perfect equilibria. We present conditions for existence of pure strategy (trembling hand) perfect equilibrium profiles and show that they always exist in a two-stage scheme under appropriate assumptions on utilities. We discuss the need for further refinements and extensions of our game theoretic analysis. *Journal of Economic Literature* Classification Numbers: C7, D7, D71.

Key Words: voting; elections; clubs; game theory; noncooperative games; pure-strategy equilibrium profiles; equilibrium refinements.

1. INTRODUCTION

Human societies evolve, grow, and shrink, as the result of exit and entry. We are interested in the evolution of those societies where entry is regulated by the use of formal voting procedures: new members are admitted only if they receive enough support from inside, according to well specified rules.

Clubs and learned societies are examples of human groups that fit our description exactly. Others may only meet part of the features we require here. For example, parliaments are elected according to well specified rules, but their size is fixed, while our focus will be on the forces that determine the growth or the stagnation of groups. In other cases, entry and exit are the result of informal procedures, whose description as voting rules may just be an approximation. In Section 2 we describe societies that essentially fit the model that we present in this paper.

Election rules are social constructs: they may come from an agreement among different founders, they may reflect the will of a unique founder, or they may be the result of successive amendments. However, once the rules for elections to a society are set, participants in the elections are bound to engage in strategic considerations that involve non-myopic behavior. In particular, voters cannot overlook the fact that newly elected members will become voters in later elections: this may lead to postpone the election of individually attractive candidates who might vote in unattractive ways, or to accelerate the election of a poor candidate whose early election may prove useful. We analyze a simple model, where the evolution of groups is clearly influenced by considerations of this type. To keep the analysis at a manageable level, we introduce many simplifying assumptions. Yet, it will be apparent that the phenomena we highlight would not disappear in more complicated situations.

The founders and the rules of election of a society are fixed in advance (we don't explain why they join to create the society or why they agree on these rules). The candidates to enter the society are fixed as well (we don't explain why they don't try to create other societies, or any other process by which eligible candidates could change from election to election). We assume that all candidates will always agree to enter the society, if invited, and that nobody leaves the society once admitted, thus concentrating on entry and not on exit. We study finite horizon situations where members of the society know at all times when it will be dissolved and voting takes place at a finite number of periods (even though in fact many societies operate under an uncertain horizon). We assume a specific voting method, whereby each member can vote for as many candidates as he wishes, and it is enough for a candidate to receive one vote in order to be admitted (this is the method of "voting by quota one"; many other methods are worth consideration). We postulate that agents' preferences are defined over streams of members in the society. Under these assumptions, we provide examples and theorems on the existence and the characteristics of different types of equilibrium profiles of the games generated in such dynamic voting contexts.

We have been explicit about the many restrictions imposed in this paper, because we think of it as a first step, pointing at new and interesting dynamic issues that will then need to be examined from different angles. We would like to provide some further justification for our assumption that the admission rule is quota one. Clearly, many other rules of admission are used by actual societies. Quota one is not among the most usual ones, even if we later provide some examples of societies that use it. But discussing quota one will allow us to separate variables and deal only with essentials in this first attempt to explain the reasons why a voter may decide to support candidates for admission. Here are some reasons why a member may vote for, or abstain from, admitting a candidate:

(1) He may vote to admit a candidate because he feels that this candidate will contribute positively to the welfare of the society, as measured by the voter's utility increase if this candidate (or a group containing him, in more complex situations) is admitted.

(2) He may vote to admit the candidate in the hope that he will help to recruit other desirable candidates in the future. If we had wanted to pursue this dynamic aspect, we should have chosen a rule different than quota one, one requiring the consent of several voters for admission.

(3) He may abstain from voting to admit a highly desirable candidate, fearing that this candidate may recruit undesirable members in the future.

(4) He may vote to admit a candidate because he hopes that such a candidate will deter others from undesirable actions that they could otherwise have taken.

It is these last two dynamic issues, not independent of each other, that we wanted to concentrate on, without being burdened by other issues. Since under quota one any member can admit any candidate he wants, he does not need the help of others and nobody else can prevent him from bringing in any desired candidate. Thus, the motivation mentioned in item (2) above is not relevant under quota one, and will not influence our results. The main dynamic considerations for admitting a member under quota one will be: What additional candidates will the newly admitted one bring in the future, and how will his presence affect the voting behavior of others? The role of these considerations can be singled out here and separated from others thanks to our choice of rule. Of course, it is important to study other rules, especially those used by actual societies that one may be interested in, and indeed we hope that this paper will encourage such studies. The dynamics of societies that function under different rules will often exhibit similar phenomena, but these will be mixed with others that result from the other rules.

When it comes to examples, the simplicity of our model becomes an asset: whatever counterintuitive results we exhibit are robust, since they happen even in simple situations. For instance, we shall prove that agents may want to vote for their enemies. This would not be surprising if they needed the votes of others in order to advance their friends to membership. But it is quite striking under our extreme assumption of vote by quota one, where each voter alone can assure his friends' admission! Also, many of our examples postulate a very simple structure of preferences: each voter is assumed to classify candidates as enemies or friends, and streams of elected members are valued as the sum of utilities derived from elected friends—one unit per period—plus the sum of disutilities derived from having enemies elected—essentially minus one per period. This reinforces our point that dynamic considerations induce strategic behavior, even under the simplest conditions.

Our paper connects two streams of literature, relating to group formation and voting systems, respectively. Explaining why groups form and remain together is a central theme in the social sciences. Thus, many existing papers have some connection with our concerns. Location theory, the theory of clubs, and the determination of jurisdictions are examples of general topics in economics, each of which has given rise to many works where groups play an explicit role. Recent papers involving a game theoretic analysis of such issues, are, e.g., Milchtaich and Winter (2001), Jehiel and Scotchmer (1997). Coalition formation is also a classical concern of game theory, which is often studied in parallel with the question of how to split the benefits of cooperation (see, e.g., Aumann and Myerson, 1988). Recent work on hedonic games, which concentrates on coalitions and abstracts from distribution, is close in spirit to the kind of questions we are interested in, but does not explicitly deal with dynamical aspects (see, e.g., Drèze and Greenberg, 1980; Banerjee *et al.*, 2001; Bogomolnaia and Jackson, 2000). In fact, many concepts that have been proposed to study coalition formation implicitly build upon some "forward looking" analysis of the consequences that would derive from joining a coalition (see, e.g., Chwe, 1994). But none of the above literature, even if related by subject, addresses explicitly and in general terms the central theme of our paper: namely, the intertemporal dynamics of groups whose membership is endogenously determined by vote. This topic is the direct concern of a few papers we know. Sobel (1998, 2000) is interested in the causes why the "quality" of the members in a group may decline (or rise) over time. Roberts (1999) analyzes the possibility of endogenizing group size in a model where policy decisions are taken by majority voting among the members. Each of these three papers reaches interesting specific conclusions after imposing demanding restrictions on the preferences and the behavior of agents. Our analysis takes a complementary direction: rather than making a specific prediction, we want to exhibit the many subtleties that lie ahead when one tries a full fledged game theoretic analysis of dynamic voting situations.

The paper is also connected to social choice theory. Strategic behavior is pervasive under many voting rules, even in contexts where decisions are one-shot, because of the phenomenon of manipulation. But in some limited contexts, involving restricted domain and one-shot decisions, manipulations can be avoided by the use of strategy-proof social choice rules. In "Voting by Committees," Barberà *et al.* (1991) study the question of electing members for a society in a static context, involving only one period. That paper characterizes classes of voting rules that can be strategy-proof under appropriate domain restrictions. We shall not describe the general classes, but simply say that they contain an interesting subclass of methods, which in addition to the preceding properties, will also respect anonymity and neutrality, i.e., will treat all voters and all candidates alike. This subclass consists of the methods based on voting by quota: each agent can vote for as many candidates as he wishes, and all candidates who get at least q votes are elected, where q is fixed a priori.[5] Our main interest in the present paper is in phenomena that only arise when the society's horizon is greater than one period, and this is why we

[5] A variant of this method is used to elect new fellows of the Econometric Society. The only departure from the above description is that q is not fixed a priori but is set equal to one-third of the number of fellows who actually vote.

have chosen to work with multi-period models whose one period version takes the form of voting by quota.[6] Since these methods are strategy-proof in their one-shot version (and thus game theoretically trivial), but not in the multi-period case, all the interesting strategic behavior that we describe will be due to dynamic considerations.

As already mentioned, our ambition is to study the evolution of societies who resort to voting as a means to include new members. It has both a normative and a positive viewpoint. On the one hand one would like to learn whether certain voting procedures and habits lead the evolution of certain societies in one direction rather than another. On the other hand one would like to provide recommendations to various groups of agents how to vote, if they want their society to reach their goals. This ambition must be tempered by the fact that the game theoretic analysis quickly becomes complex and presents several alternative routes. Accordingly, the paper contains examples, which point at the complexities of the analysis, as well as technical results on how to solve for equilibrium profiles and for what types of profiles to look for. It is structured as follows. In Section 2 we present the model, based on a gallery of assumptions. Section 3 contains examples. These examples show that the simplicity of the one period model is immediately lost if we have several periods. They also prove that some counterintuitive phenomena, like strategic voting for enemies, can occur if the number of periods is not too small. They also indicate that it will be worth analyzing not one but several solution concepts, because each one of them can provide some insight into the phenomena we try to model. One example shows that, although we concentrate on pure-strategy equilibrium profiles, the use of mixed strategies, or even correlated strategies, may be most reasonable in some cases. In Section 4 we provide a game theoretic analysis of our model, focusing on subgame perfect equilibrium profiles, quasi-strong profiles,[7] and Pareto undominated profiles. Section 5 is devoted to the existence of perfect equilibrium profiles in pure strategies: we provide a sufficient condition under which there will exist such profiles and examples showing that the condition is not necessary. We are able to show that if certain additivity assumptions are satisfied, every game that is generated by a generic 2-stage voting scheme has a pure-strategy perfect equilibrium.

[6] We work with the special case of quota one. It can be shown that all of our examples would also apply for other quotas and most of the results could be extended to other quotas. This, however, will introduce a new issue, namely, the need to admit an undesirable candidate who is needed to admit desirable persons in the future.

[7] I.e., equilibrium profiles that have the additional property that no deviator can benefit if the set of deviators does not include the set of all voters at the start of a deviation.

2. THE MODEL

Our goal is to analyze the results from imposing some electoral rules on the evolution of societies. The necessary elements to describe the rules, which we call *(finite horizon) voting schemes*, are the following:

(1) A nonempty set of *original founders*, denoted F^0, who belong to *society* at the initial stage and from stage to stage vote to bring in other members and/or to remove members. "Society" may be an organization, a club, a foundation, or similar enterprises.

(2) A set of *candidates* from whom new members can be chosen. This population may vary from stage to stage.

(3) A set of *voters* for each stage. Often, all elected members can vote at all stages following their election for as long as they belong to the society.

(4) A set of *rules* which specify under what conditions a person is admitted to the society, or is expelled, or resigns.

(5) A number of stages k during which the society operates. After k stages the society dissolves, having concluded its tasks, and *the play* is over.

An important part of the outcome of the voting scheme is the resulting *stream of members*, denoted $\mathscr{F} := \{F^1, F^2, \ldots, F^k\}$, where F^t represents the members at stage t, after the elections, expulsions, and resignations at that stage. Another part may be information concerning who voted at each stage and for whom. Some of the above may be unknown to some, or all the agents. All of the information that is available to agent i until stage t constitutes his $(t-1)$-stage history.

The decision on how to vote at each stage, for every voter i faces, should take into consideration the priorities that each agent has over the various streams.[8]

The above setup gives rise to many voting schemes, each possessing different properties. In order to obtain meaningful results one has to specialize to particular models in which the population is specified, the exact rules for admittance, expulsion, and resignation are given, and some information is provided about the priorities of the participants. In this paper we restrict ourselves to a particularly simple model—as simple as we could think of—that still yields insight on some dynamic issues that are involved. We hope that the analysis of this particular model will encourage

[8] One can think of complicated priorities on events that may even be concealed. For example, a voter might not like an agent j, if he knew that agent p also voted for j, but otherwise he might have loved to have j in the society. Perhaps he does not even know who elected j. We shall not consider such complications in this paper.

investigation of other, more complicated ones and that eventually insight, both descriptive and normative, will be gained about real societies.

Specification of the Voting Scheme

1. *Fixed Population.* We assume that *the population is finite and fixed and includes the nonempty set of the original founders* F^0. Therefore, we can denote the set of *agents* by N. $N \setminus F^0$ is called *the set of the original candidates* and is denoted by C^0. Similarly, we write C^t for $N \setminus F^t$. Members of C^{t-1} are the candidates from whom the voters F^{t-1} can choose at stage t.

2. *No Firing*. *We assume that an elected candidate will stay in the society all the time*. There are no provisions to fire him.

3. *No Resignation*: *Normalization. Once an agent is admitted to the society, he will stay there throughout the performance of the society. Staying alone in the society has a zero utility*.[9]

The no resignation requirement makes sense if staying in the society is highly prestigious. Nevertheless, even then it is a restriction. For example, it rules out strategies involving threats to resign, as punishments, if deviations occur.

In this paper we take the position that an agent becomes a player only after he joins the society. We shall rarely compare his utilities while in the society to his utilities before he joined the society.

4. *The* 1-*Quota Voting*. The rule for electing a candidate into the society is simple: *every voter can bring any number of candidates into the society at any stage*, *simply by casting a vote for them at the beginning of that stage*. This rule is known as *voting by quota* 1.[10]

5. *Streams of Members Are All That Matter*. We assume that *each agent cares only about the streams of members in the society* and does not care, for example, about who voted, or who did not vote for each member. This allows us to require that all his actions are based only on what he knows about the developing streams.

6. *Common Histories*. We assume that *at each stage the elected candidates are known to everyone*. Thus, for every agent i the relevant $(t-1)$-

[9] Sometimes we change the normalization, so that a zero utility corresponds to a situation where the agent stays in the society together with the original founders F^0. The reader will have no difficulty in deciding to which normalization we refer in each instance.

[10] See Section 5 for qualifications about the restrictiveness of this specification. Many aspects of our analysis would hold, or could be extended, for other quotas.

stage histories are the same,[11] namely, subsequences of the streams terminating at F^{t-1}. These will be denoted h^t, $t = 1, 2, \ldots, k$. Thus, $h^t := \{F^0, F^1, \ldots, F^{t-1}\}$.

We now have all the ingredients to convert the above setup into a *game form:* The set of *players* is N; the *pure strategies* available to player i are choices of sets that specify at each stage t the candidates that he votes for at that stage only as a function of the history at that stage. Thus, we do not allow the strategies to depend on who voted for whom in the past.

With this notation, a pure strategy for agent i, can be expressed as $\sigma_i := (\sigma_i^1, \sigma_i^2, \ldots, \sigma_i^k)$, where $\sigma_i^t(h^t)$ denotes the set of agents chosen by agent i, given[12] the history h^t.

From this description one can realize that we formally allow a player at each stage to vote even for agents that were already elected (including himself) and we allow an agent to vote even if he is not elected. This is done merely for mathematical convenience. Of course such votes will have no effect on the stream of members. Given a strategy profile $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, the stream of members is given by

$$F^t = F^t(\sigma) = F^{t-1} \cup \left(\bigcup_{i \in F^{t-1}} \sigma_i^t(h^t) \right) \qquad (t = 1, 2, \ldots, k). \quad (2.1)$$

Most of this paper will deal with pure strategies. Since the game is of perfect recall, by Kuhn's (1953) theorem (see also Selten, 1975), even when we do employ mixed strategies we can restrict ourselves to behavioral strategies, in which case it is sufficient to consider the probability distribution on the various histories.

Discussion. Modeling a complex real-life society was not a primary goal of this paper. In fact, we purposely tried to omit some ingredients that occur in life but that could blur the dynamic issues involved. Nevertheless, the present voting scheme may be a reasonable approximation to different real life cases. In addition to the vote for new members in learned societies (see footnote 9), we can propose some other examples. One that comes to mind is *a society of party goers:* A group of people that meet every weekend simply to have fun. They may sing together, or hold debates, or engage in

[11] Actually, if ballots are not secret, histories may be more complicated than simply the past stream of members. They may include information such as who voted for whom, and when. In this paper we shall not employ such histories.

[12] Note that we allow a strategy of an agent to depend on the part of the stream that existed before he entered the coalition. Usually, however, this may not be the case. Note also that there is a redundancy in this notation: What agent i "votes for" in stages before he was admitted into the society has no effect on the resulting stream of members. We use this notation for the sake of brevity.

sport activities. Each one of them can invite outsiders to come to the party, if he thinks that these persons can add to the fun. There is no authority among the people that can prohibit a member from bringing whomever he wants.

To convert our game form into a game we now introduce priorities and utilities.

7. *Known Utility Functions*. We assume that *the priorities of agent i are given by complete and transitive binary relations on the set of outcomes* and therefore they can be represented by a *utility function* u_i. Later, when we deal with mixed strategies, we shall assume that *these utility functions are*, *in fact*, *von Neumann Morgenstern utility functions*.[13]

Once an agent is in the society, every stream that is better for him than staying alone is assigned a positive utility. Every stream that is worse for him is assigned a negative utility (still larger than the utility of not being in the society).

We now present several classes of *utility functions*. It should be stated at once that the simple utility functions are adopted only in order to exhibit examples, where it is shown that even under these assumptions, quite complicated dynamic phenomena occur. For the main results, concerning the characterization of pure-strategy Nash equilibrium profiles and their refinements, quite general utility functions are assumed (Case 8b below).

The simplest class of utility functions considered in this paper assumes that for every pair of distinct agents i, j, either i *likes* j, or i *dislikes* j. Expressing it differently, we say that either j is a *friend* of i or he is an *enemy* of i, where friendship and enmity merely mean that he would like or would not like the person in the society. This does *not* imply that a voter will always vote for his friend. He may be reluctant to do so if, for example, he thinks that his friend may bring enemies to the society.

We do not assume that the "friendship" relation is either symmetric, or transitive: Agent j can be a friend of i, yet i is regarded as an enemy by j. Also, a friend of a friend need not be a friend.

"A friend" may be interpreted in several ways, such as, "the voter enjoys his company," "the voter thinks he will be useful for the workings of the society," "that his opinion should be heard, because it is relevant," etc. Likewise "an enemy" can have opposite interpretations.

We then assume that *each agent wishes to spend as much time as possible with friends and as little time as possible with enemies and that this is all he cares for*. If the stages are equally spaced in time, it then makes sense to denote by 1 the utility of having a friend in the committee for one stage

[13] This of course involves more assumptions on the binary priority relations.

and by $(-1-\varepsilon)$, the utility of having an enemy for one stage, where ε is a small positive number, added to break ties.[14]

If we adopt the, perhaps naive, assumption that the agents merely wish to spend as much time as possible with friends and as little time as possible with enemies, we can take time as a measure of utility or disutility. This yields additive utility functions, because time is always additive.

We summarize the above formally:

8a. *Pure Friendship and Enmity. The utility for a stream of members, given by* (2.1), *for an agent who succeeds in entering the society is given by*

$$u_i(\mathscr{F}) = \sum_{\{t \geq 1:\ i \in F^t\}} |F^t \cap \mathrm{fr}(i)| - (1+\varepsilon) \sum_{\{t \geq 1:\ i \in F^t\}} |F^t \cap \mathrm{en}(i)|, \quad (2.2)$$

where $|S|$ *denotes the cardinality of* S, $\mathrm{fr}(i)$ *denotes the set of friends of* i, *and* $\mathrm{en}(i)$ *denotes the set of enemies of* i. Here, $\mathrm{fr}(i) \cup \mathrm{en}(i) = N \setminus \{i\}$ for each agent i.

Our general results do not require such simple utility functions. They still hold for the much more general class that is described as follows:

8b. *General Stream Dependence. The utility of an agent who became a member of the society is only a function of the stream of members that occurred*,[15]

$$u_i(\mathscr{F}) = u_i(F^0, F^1, \ldots, F^k). \quad (2.3)$$

Some other, intermediate classes of preferences are also attractive. We present one of them: It assumes that each agent has a time-independent and society-independent *weight function* which reflects the utility of the agent for one stage, so that one can still talk about a friend (positive weight) and an enemy (negative weight).[16]

If each voter wants to spend as much time as possible with friends and as little time as possible with enemies, this yields the following utility function:

8c. *Friends and Enemies: Additivity within Each Stage and across Stages. Every agent i has a weight function* $w_i : N \to \mathfrak{R}$. *His utility* $u_i(\mathscr{F})$ *for a stream*

[14] The decision to require a positive ε is arbitrary and was chosen in order to be specific in some examples. It will play no role when the general results of this paper are proved.

[15] We even allow his utility to depend on events that occurred before he entered the society.

[16] One can talk about friends and enemies also if the weaker condition of separability prevails, under which a friend of a member is a person who will always increase the utility of the member, if he is added to the society.

of members $\mathscr{F}$, serving in the society is given by

$$u_i(\mathscr{F}) = \sum_{\{t \geq 1:\ i \in F^t\}} \sum_{a \in F^t} w_i(a). \tag{2.4}$$

In every example, theorem, or application we shall be precise about the assumptions we make about the individual preferences.

To complete the descriptions above we make a last assumption:

9. *Common Knowledge*. *All utility functions, as well as all the descriptions above, are common knowledge*.[17]

Who Are the Players? We have set up the society protocol and we have converted it into a game. Clearly, the way we formulated it, the set of players is N. Yet, we can also regard the situation as a *sequence of several games*, one starting at each stage, with different players, where the players at each stage t are the set of voters F^{t-1} and the other agents are considered extraneous entities. Indeed, agents do not really become players until they enter the society. The only votes that count are those of agents who are members by that stage. They create the continuation and it is their interest that matters.[18] Thus, if we want to talk about refinements of Nash equilibria,[19] we sometimes prefer to make them relative to the set of voters at each stage. Accordingly, we shall employ the following definition:

DEFINITION 2.1. An equilibrium strategy profile σ is called *sequentially Pareto-undominated*, if for every $t \in \{1, \ldots, k\}$ there does not exist another equilibrium strategy profile which coincides with σ up to stage $t-1$, whose outcome is weakly preferred by all voters in F^{t-1} and strongly preferred by at least one of them. The payoff that such a strategy yields is called a *sequentially Pareto-undominated outcome*.

The concept of "strong equilibrium" was introduced in Aumann (1959). We shall encounter in the next section games for which strong equilibrium profiles do not exist. Nevertheless, we shall show in Section 4 that it is

[17] This assumption is essential, although, often weaker requirements may be sufficient. The reader is invited to examine Example 3.2 and determine what knowledge each voter needs to have about the candidates and about the knowledge of the other voters, in order that the analysis makes sense.

[18] There are two ways of looking at it. On the one hand, the voters at a stage make their own decisions. They can even dictate to the elected candidates how to vote in the future, threatening not to bring them into the society if no agreement is reached. On the other hand they also have to take into account that the people who are going to participate are pursuing their own interests and will not abide by the agreement if they can benefit by violating it.

[19] In this paper "equilibrium strategy profiles" always mean Nash equilibrium profiles.

often possible to achieve "quasi-strong equilibrium profiles" as defined below:

DEFINITION 2.2. An equilibrium strategy profile σ is called *quasi-strong*, if at no stage can any voter of that stage benefit by a deviation that involves a proper subset of the voters.

This concept is in a sense weaker than Aumann's, because it does not allow for deviations involving all the voters. In another sense it is stronger, because it assures every voter that without his consent nobody can gain (even if others lose).

3. SOME INTERESTING SIMPLE EXAMPLES

In this section we demonstrate by examples some of the dynamic issues that may occur in voting schemes.

A Universal Equilibrium Profile. One equilibrium profile always exists in pure strategies:[20]

If there is more than one founder, each founder votes at stage 1 *for every candidate—friends and enemies* and (off the equilibrium path) every voter votes always for every candidate. This is certainly an equilibrium point, as nobody can change the outcome.

If there is only one founder he chooses that stream that maximizes his utility given that as soon as there are at least two voters, each will vote for every candidate. For example, under pure friendship and enmity (Assumption 8a),[21] he will vote for all his friends in the first stage, if he has more friends than enemies (and every candidate will be brought in at the second stage) and if the number of friends does not exceed the number of enemies he will vote for nobody until the last stage, whereupon he will bring all his friends.

A Transitive Friendship Relation. Here we assume additivity within each stage and across stages (Assumption 8c). If friendship is transitive, then the following is an equilibrium profile: *Each founder votes for all his friends at the first stage* and (off the equilibrium path) each voter votes for all his friends. Indeed, under this strategy, a founder need not be afraid that any of his candidates will bring anybody later and no voter can gain by deviation, neither by voting for fewer friends nor by bringing in enemies.

[20] This was first observed by Hans Reijnierse (private communication).
[21] Assuming that ε is small enough.

This equilibrium profile is perfect (see Selten, 1975), because the strategy for each player remains a best reply against any possible trembles of the others. Surprisingly, it is not necessarily a sequentially Pareto-undominated equilibrium profile (see Example 3.2 below).

The Case $k = 1$. This case is quite clear under additivity within a stage (Assumption 8c): Having each founder voting for his friends is certainly an equilibrium profile. It is perfect and Pareto-undominated, but it is *not necessarily strong*. For example, under pure friendship and enmity (Assumption 8a), if there are several founders, each having one and a different friend then the set of all founders can all benefit by all voting for nobody. *This example, which can easily be extended to any number of stages, demonstrates that one cannot always obtain a strong equilibrium profile.*

We remark that under friends and enemies and additivity within each stage (Assumption 8c), every voting profile that produces the set of all friends of all the original founders as an outcome and in which each founder votes at least for his friend constitutes also an equilibrium profile. These profiles produce the same outcome, so they are all Pareto-undominated but they need not be perfect: voting for one's friends only is a best reply against any tremble.

Complications can occur if additivity does not prevail, as the following example shows:[22]

EXAMPLE 3.1. $F^0 = \{1, 2\}$, $k = 1$, $C^0 = \{a, b\}$,

$$u_1(\varnothing) = 2, \quad u_1(a) = 3, \quad u_1(b) = 1, \quad u_1(ab) = 0,$$
$$u_2(\varnothing) = 3, \quad u_2(a) = 0, \quad u_2(b) = 2, \quad u_2(ab) = 1.$$

Here, $u_i(S)$ stands for the utility of founder i for $S \cup \{1, 2\}$. A similar convention will be used throughout.

Possible Scenario. Founder 1 likes to stay alone. He thinks it is a good idea to bring a to the society and it is a bad idea to bring b. It is a disaster to bring both, because the two will fight all the time. Founder 2 does not like a's views. He somewhat prefers b, but would above all like to stay alone. Bringing both is a "compromise" between the previous two undesirable events.

The pure-strategy equilibrium points are (b, b), (a, ab), and (ab, ab). None of them is trembling-hand perfect—they are all eliminated by weak domination. The only perfect equilibrium is mixed, in which Founder 1

[22] Here, and in the sequel, we sometimes omit curly brackets and commas. We write, for example, $u_1(ab)$ instead of $u_1(\{a, b\})$.

votes for $\emptyset$ and a with equal probabilities and Founder 2 votes for $\emptyset$ and b with equal probabilities.

This example demonstrates that sometimes one has to resort to mixed strategies if one wants a perfect equilibrium profile. We return to this issue in Section 5.

The Case $k = 2$. This case gives rise to other types of complications as is manifested by the following two examples. These complications appear already under pure friendship and enmity (Assumption 8a). This assumption will prevail for the rest of this section.

EXAMPLE 3.2. $N = \{a, b, c, d, e, f\}$; $F^0 = \{a, b\}$, $\text{fr}(a) = \{c\}$; $\text{fr}(b) = \{d\}$; $\text{fr}(c) = \text{fr}(d) = \text{fr}(e) = \text{fr}(f) = \emptyset$. $k = 2$. (It does not matter who the candidates e and f have as friends.)

Since friendship here is vacuously transitive, the following is a perfect equilibrium profile: *a votes for c at both stages and b votes for d at both stages, regardless of the histories*. Nevertheless, there is another equilibrium profile that is preferred by *both* players: *players a and b bring their friends only in the second stage and if anyone deviates in the first stage, both a and b invite all the remaining candidates in the second stage*. In this strategy each founder ties the hands of the other founder: "If you do not abide, we shall punish you by bringing in all the enemies." This is even a subgame-perfect equilibrium and sequentially Pareto-undominated,[23] but it is *not* perfect: Whatever the action of the other person, voting only for one's friend in the last stage is never worse and in some cases better than the prescribed action.

We see already in this simple example the dilemma of which equilibrium to recommend: A perfect equilibrium which yields small but "safe" profits, or an equilibrium which maximizes profits but uses threats of whose credibility is questionable?

EXAMPLE 3.3. $N = \{1, 2, 3, a, b, c, d, e, f, g, p, q, r, s\}$; $k = 2$; $F^0 = \{1, 2, 3\}$; $\text{fr}(1) = \{g\}$; $\text{fr}(2) = \{e, f\}$; $\text{fr}(3) = \{a, b, c, d\}$; $\text{fr}(a) = \{p, q\}$; $\text{fr}(b) = \{q, r\}$; $\text{fr}(c) = \{p, r\}$; $\text{fr}(d) = \{p, q, r\}$; $\text{fr}(e) = \{s, p\}$; $\text{fr}(f) = \{s, q\}$; $\text{fr}(g) = \{s, p, q\}$; $\text{fr}(p) = \text{fr}(q) = \text{fr}(r) = \text{fr}(s) = \emptyset$.

[23] Another variant, in which the deviator is punished only by the other person, in case of deviation, is *not* subgame-perfect but is more convincing: Why should the deviator agree, and abide by punishing himself? This is another manifestation of the known dilemma: Why should one trust a promise of a person, who already proved that he does not keep his promises, because he deviated in the first stage. Note that formally the strategies in this variant are not functions of the stream only. They also depend on knowing who deviated in the first stage. Of course, this knowledge can be *deduced* if one remembers for whom he himself voted and what was the outcome of the first stage.

We reach a conclusion by the following heuristic arguments: At first one thinks that 1 should not invite g at stage 1, because inviting him would bring about three enemies of 1 in the second stage. Similarly, 2 should apparently not invite any of his friends, because that would bring him more enemies in the last stage. Player 3, however, should invite *all* his four friends (not less!) in the first stage, because that will bring him only three enemies in the next stage, with a net profit of $1 - 3\varepsilon$, compared to not inviting any friend in the first stage.

Realizing that p, q are going to be in the society in the last stage anyhow, player 2 should not hesitate to vote for his friends in the first stage: He gets two friends at that stage but suffers from only one additional enemy in the next stage.

Realizing that also s will be present in the last stage anyhow, it now follows that 1 can only gain by bringing his friend in stage 1.

Thus, the following is an equilibrium profile: *Every voter brings all his friends as soon as he is allowed to vote*.

It can be checked that this is indeed an equilibrium profile and, moreover, it is perfect.[24]

This is *not* a sequentially Pareto-undominated equilibrium. Like in the previous example, there is a sequentially Pareto-undominated, subgame-perfect but not perfect equilibrium that will be strictly preferred by all original founders, and in fact, by everyone who will find himself eventually in the society, namely, to invite nobody in the first stage, invite one's friends in the second stage, and punish deviations by each voter inviting everyone in the second stage.

To sum up: We exhibited here a "safe" equilibrium outcome that does not yield much to the founders and another "not so safe" that brings about higher utilities to the founders, and moreover brings about a society with much fewer frictions in it. Which one (if any) should be chosen has to be decided by the members. Do they trust their co-founders to honor the "agreement" in the second case? Do they believe that the "punishment" will be carried out in case of a breach? The answer to such questions, we feel, is beyond the scope of the theory.

When Many Common Enemies Exist. We have seen in the previous example how a punishment can force an equilibrium. In fact, if there are enough common enemies, then *any agreement* between the current founders, at any stage other than the last, can be enforced by a strategy that stipulates that out of the agreement all voters will vote for all common

[24] Any "tremble" can be observed only in the last stage when it is still to one's advantage to bring all his friends. The considerations of this example can be employed to produce classes of 2-stage voting schemes for which pure-strategy perfect profiles always exist.

enemies as soon as they recognize that they are off the equilibrium path. This is even subgame-perfect.

The question then becomes: Which agreements are the players likely to sign? Realizing that almost all agreements can be made binding as explained above, this case should be handled with the tools of cooperative game theory and this is outside the scope of the present paper.

We keep the above in mind but we wish to make the following two observations: (1) In real life one can usually extend the set of candidates so as to include as many common enemies as one "wishes." (2) Nevertheless, a threat to bring these common enemies is often not credible as a general procedure. It often would be considered unthinkable, because it would undermine the very foundations upon which the society rests. Thus, although such threats may be feasible, often they are not viable, which brings us again to the recognition that a model does not usually capture all the intricacies of a real situation.

The Helpful Enemy. We have seen how voting for an enemy may be beneficial off the equilibrium path. The following example will show that voting for an enemy may be beneficial also along the equilibrium path.

EXAMPLE 3.4. $N = \{a, b_1, b_2, \dots, b_5, c_1, c_2, \dots, c_5, d, e\}$; $F^0 = \{a\}$; $\mathrm{fr}(a) = \{b_1, \dots, b_5\}$; $\mathrm{fr}(b_i) = \{c_i\}$, $i = 1, \dots, 5$; $\mathrm{fr}(c_i) = \{d\}$, $i = 1, \dots, 5$; $\mathrm{fr}(d) = \{e\}$; $\mathrm{fr}(e) = \emptyset$; $k = 4$.

The founder would like to bring all his friends, but if he simply does so at the first stage then each b_i will bring c_i in the next stage. This is because the b_i's will not fear[25] that c_i will bring d before the last stage, knowing that if c_i does so, d will bring e. To prevent this from happening, the founder can vote for e in the first stage. A complete strategy profile is

$$\sigma_e^t = \emptyset \qquad (t \in \{2,3,4\}), \quad \forall F^{t-1};$$

$$\sigma_d^t = \{e\} \qquad (t \in \{2,3,4\}), \forall F^{t-1};$$

$$\sigma_{c_i}^4 = \{d\} \qquad (i \in \{1,\dots,5\});$$

$$\sigma_{c_i}^t = \begin{cases} \{d\}, & \text{if } e \in F^{t-1} \\ \emptyset, & \text{otherwise} \end{cases} \qquad (i \in \{1,\dots,5\}) \quad (t \in \{2,3\});$$

$$\sigma_{b_i}^4 = \{c_i\} \qquad (i \in \{1,\dots,5\});$$

[25] We are using the fact that because ε is positive (Assumption 8a), a voter will prefer to postpone a vote for a friend if this friend will bring an enemy at the next stage. He will gain an ε by postponing one stage.

$$\sigma_{b_i}^3 = \begin{cases} \{c_i\}, & \text{if } d \in F^2 \\ \varnothing, & \text{otherwise} \end{cases} \qquad (i \in \{1,\dots,5\});$$

$$\sigma_{b_i}^2 = \begin{cases} \{c_i\}, & \text{if } d \in F^1 \\ \{c_i\}, & \text{if } e \notin F^1 \\ \varnothing, & \text{otherwise} \end{cases} \qquad (i \in \{1,\dots,5\});$$

$$\sigma_a^t = \varnothing, \qquad (t \in \{2,3,4\});$$

$$\sigma_a^1 = \{b_1,\dots,b_5,e\}.$$

One can verify that this is indeed an equilibrium profile.

Discussion. Inviting an enemy is quite common in real-life, when rules of admittance are different. One often votes for an undesirable person, because he needs his vote to bring in other candidates. This argument is not relevant if the rule of admittance is the quota 1 rule. The voter can bring to the society any person he wishes. The example here shows that even in this case it may be beneficial to bring an undesirable person, because he can *deter* others from bringing in many undesirable persons. This aspect would have been harder to appreciate without a formal analysis.

EXAMPLE 3.5. THE GAME OF CHICKEN. In this example, $F^0 = \{1,2\}$, $C^0 = \{x_1, x_2, y_1, y_2\}$, $k = 3$. Founder 1 likes only x_1, who likes only y_1. Founder 2 likes only x_2, who likes only y_2. Agents y_1 and y_2 like only each other.

Skipping formalities, each founder can essentially either vote for his friend in the first stage, or refrain from doing so. (He strictly loses by voting for an enemy at this stage.) Unfortunately, if player 1 votes for his friend at the first stage, player 2 will lose if he too votes for his own friend. The reason is that in this case it is clear that both y_1 and y_2 will be present in stage 3, so there will be no reason for both x_1 and x_2 to refrain[26] from voting for their friends in stage 2. These friends are enemies of the founders. Putting together the relevant information and ignoring ε,

[26] If, say, y_2 were not present at stage 3, x_1 would not have invited y_1 at stage 2, as ε is positive and x_1 knows that y_1 will bring y_2 (an enemy of x_1) at the last stage.

we get the following payoff as functions of the choices in the first stage:

	$\varnothing$	x_2
$\varnothing$	0, 0	-3, 1
x_1	1, -3	-4, -4

This is the famous game "chicken." It has two pure-strategy equilibrium points, $(x_1, \varnothing)$ and $(\varnothing, x_2)$, yielding payoffs $(1, -3)$ and $(-3, 1)$, respectively. In addition, the players can each use a mixed strategy $(1/2, 1/2)$ that yields a more sensible payoff $(-1.5, -1.5)$. All these are undominated and therefore perfect (see Kohlberg and Mertens, 1986, Appendix D).

Even more sensible for the players is to employ a correlated strategy under which a mechanism chooses one of $(\varnothing, \varnothing)$, $(\varnothing, x_2)$, $(x_1, \varnothing)$ with equal probabilities, informing founder 1 what row was chosen and informing founder 2 what column was chosen. The founders then choose whatever row/ column they were told, thereby reaching an equilibrium expected payoff equal to $(-2/3, -2/3)$. *This example comes to show that mixed and correlated strategies should not be ignored*.

Discussion. In this paper we are interested in pure-strategy profiles. This is because in real-life voting procedures it is rarely the case that people perform lotteries in order to determine what votes to cast. The example above has an educational value: It shows that opportunities may be missed if people restrict themselves to pure strategies and that even correlated equilibria are sometimes beneficial.

4. A CHARACTERIZATION OF PURE-STRATEGY EQUILIBRIUM STREAMS

At the beginning of this section we study *common voting* profiles; namely, profiles under which, at each stage, all voters vote for the same set of candidates. We show that every equilibrium outcome that can be reached by a pure-strategy profile can also be reached by a common-voting profile that generates the same stream of members. These profiles have the additional advantage that they are quasi-strong equilibria (Definition 2.2). A quasi-strong equilibrium gives each voter the assurance that, without his participation, no subgroup of the other players will agree to deviate, because none of them will gain, and some may even lose.

We then proceed to characterize and construct all the equilibrium streams, and therefore all equilibrium outcomes that can be achieved by pure strategies. As many of these outcomes are undesirable one would like to find out those that are sequentially Pareto-undominated. We provide a sufficient, albeit not necessary, property of such outcomes, which eliminates many outcomes that are sequentially Pareto-dominated.

An important refinement of Nash equilibrium profiles is the set of subgame-perfect equilibria. In Example 3.2 we encountered a case where a profile which is not subgame-perfect should be preferred to the subgame-perfect one. Nevertheless, one wants to know if *outcomes* that are generated by pure-strategy subgame-perfect profiles form a proper subset of those generated by pure-strategy Nash profiles. We show that this is not the case! Every pure-strategy Nash equilibrium outcome can be achieved using subgame-perfect profiles.

A key role in reaching some of these results is expressed in the following:

Remark 4.1. Quota one implies that whoever the voters bring in can also be brought by one voter. Consequently, if a set S of candidates is chosen in an equilibrium profile of a 1-stage game, this set has the property that, if elected, no voter would have preferred that more members were added to it. Indeed, if all the voters vote for a set of candidates S, a deviation can only increase the set of new members.

All strategies in this section are pure and we shall rarely repeat this fact. To avoid trivialities we assume henceforth that $C^0 \neq \emptyset$.

PROPOSITION 4.2. *Let σ be a pure-strategy equilibrium profile for a* 1-*stage game* Γ. *The strategy profile $\bar{\sigma}$, generated from σ by common voting, is a quasi-strong equilibrium profile for* Γ.

Proof. If $|F^0| = 1$, then $\bar{\sigma} = \sigma$. It is an equilibrium point and vacuously a quasi-strong one. Let $|F^0| > 1$. The set S of players that was elected under $\bar{\sigma}$ is the same set that was elected under σ; therefore, it yields the same payments. Any deviation from $\bar{\sigma}$, made by a nonempty *proper* subset of the founders, can only yield a set that contains S, because the remaining founders still vote for S. Therefore, if such a deviation from $\bar{\sigma}$ resulted with some members gaining, then, in σ each of them could have forced the same better payment, by alone adding the same additional candidates, contrary to the fact that σ is an equilibrium profile for Γ. ■

One should be careful when one tries to generalize Proposition 4.2 to multi-stage games: At future stages "new" players may enter the game and one is inclined to take into account possible agreements involving them, as a condition to be elected. Consider the following:

EXAMPLE 4.3. Let $F^0 = \{1, 2\}$, $C^0 = \{a, b\}$. Under pure friendship and enmity (Assumption 8a), agents 1 and 2 like agent a. Agent a likes agent b. For all other pairs (i, j), j is an enemy of i. $k = 2$. Assume also that agent a prefers to be in the society to not being there, no matter who else is with him.

The following strategy profile is subgame-perfect,

$$\sigma_1^1 = \emptyset, \qquad \sigma_1^2 = \{a\}, \qquad \sigma_2^1 = \emptyset, \qquad \sigma_2^2 = \{a\}, \qquad \sigma_a^2 = \{b\},$$

where these actions are taken on and off the equilibrium path.

This profile is already in common voting for the original founders, who vote the same way throughout the play, on and off the equilibrium path. Nevertheless, this profile is not immune to deviation involving a proper subset of the founders: Agents 1 and a can deviate by 1 voting for a already in the first stage and a promising to vote for $\emptyset$ at the second stage. By this deviation agent 1 gains and agent a also gains, because he becomes elected.[27]

The point is that the strategy profile above is *not* in common voting with agents that may later be admitted to the society!

Indeed, augmenting the above example and requesting that both 1 and 2 vote also for b at the second stage, if a is elected in the first stage, then no profitable deviation can take place by a proper subset of the founders. For example, it will do no good that a will refrain from voting for b, because founder 2 will still vote for b.

With this understanding we can generalize Proposition 4.2 as follows:

THEOREM 4.4. *Let Γ be a game representing a voting scheme obeying general stream dependence* (*Assumption* 8b). *Let σ be a Nash equilibrium of Γ. Let $\bar{\sigma}$ be the profile derived from σ by common voting at each stage*, *on and off the equilibrium path.*[28] *Then*, *$\bar{\sigma}$ is a quasi-strong equilibrium of Γ, giving the same stream as σ. If σ is a subgame-perfect profile then $\bar{\sigma}$ is also subgame-perfect.*

Proof. Since actions in σ may depend only on the history of membership (and not on who voted for whom), common voting preserves the set of candidates voted into the society at every stage, both on and off the

[27] One can question how safe is this agreement between 1 and a. Obviously, a will desire not to honor the agreement. This, however, is irrelevant to the claim that 1 and a can both gain if they follow this agreement.

[28] The requirement of common voting refers only to the profile σ. Of course, if a player, or several players, decide to deviate, they need not adhere to the common voting stipulation.

equilibrium path. Therefore, the outcome stream $\mathcal{F} := \{F^0, F^1, \dots, F^k\}$ of σ coincides with that of $\bar{\sigma}$.

The profile $\bar{\sigma}$ is also an equilibrium point. Indeed, if a deviation of an agent i from $\bar{\sigma}$ profits him, then he could profit the same way by deviating alone from σ, voting at each stage for those members who were elected due to $\bar{\sigma}_{-i}$ together with those elected by him in his deviation, a contradiction. By the same token, $\bar{\sigma}$ is subgame-perfect if σ was subgame-perfect.

Suppose that $\bar{\sigma}$ is not quasi-strong. Then there exists a profile τ that coincides with $\bar{\sigma}$ up to a certain stage t^* and deviates from that stage, where the deviation is done by a proper subset of F^{t^*-1}, together with agents who are admitted later to the society. Let $\mathcal{G} := (F^0, F^1, \dots, F^{t^*-1}, G^{t^*}, \dots, G^k)$ be the stream of members that result from τ. Suppose that a member i of F^{t^*-1}, who is not necessarily a deviator, prefers $\mathcal{G}$ to $\mathcal{F}$. We claim that he alone could generate $\mathcal{G}$, if all other agents obey $\bar{\sigma}$. Indeed, consider an arbitrary subgame starting at an arbitrary stage t, $t \geq t^*$, on, or off the equilibrium path of $\bar{\sigma}$. Let H^t be the set of members elected at this stage due to $\bar{\sigma}$. Then it is contained in the set of members elected at this stage due to τ, because there are members of F^{t^*-1} who still vote as in $\bar{\sigma}$. This very same set can be voted into the society by agent i alone. It follows that i can benefit by a deviation from $\bar{\sigma}$, contrary to the fact that $\bar{\sigma}$ is an equilibrium profile. ■

We have shown that all pure-strategy equilibrium outcomes can be generated by common voting. The natural question that now comes to mind is how to characterize all streams that constitute such outcomes. Proposition 4.5, Theorem 4.6, and Corollary 4.7 provide an answer.

PROPOSITION 4.5. *Assume that there are at least two founders in a* 1-*stage game* Γ. *A set* S *of candidates chosen can result from a pure-strategy equilibrium profile iff* S *has the property that no founder would prefer to add members to* S.[29]

Proof. The "only if" part is explained in Remark 4.1. Conversely, suppose S has this property and is voted, say, by common voting. Then no player can benefit by deviating alone: He cannot delete members from S and he does not want to add members to S. ■

Thus, for a multi-person set of founders, to generate all equilibrium outcomes for a 1-stage game one has to examine all subsets S of C^0 and select those that have the property that no founder would like to augment them. This task is manageable by a computer if $|N|$ is reasonably small and $k = 1$. It becomes less so when the number of stages increases.

[29] Such an S always exists, for example, $S = C^0$.

To extend Proposition 4.5 to a k-stage game, for $k > 1$, we employ a process that we call *collation*, which we now explain.

Consider a tree game Γ, representing a k-stage voting scheme. Consider an arbitrary subgame starting at the last stage. This is a tree form for a 1-stage voting scheme, connected to the root of Γ by a unique path. Its endpoints represent the payoff vectors that would be obtained if the players proceeded along this path and continued along the subgame. Thus, if we fix an action[30] for each voter of the subgame (and remember it), we can delete the subgame and connect the resulting payoff vector to the new endpoint, at the root of the subgame. Fixing actions at each last-stage subgame allows us to delete them, thus converting the tree to a $(k-1)$-stage game. By this collation we can obtain backward induction results, by considering only 1-stage games, even though Γ is not a perfect-information game. Note that strategies constructed in this way usually do not depend only on the streams. To force a strategy that depends only on the stream, we have to require that actions taken at paths that correspond to the same stream are the same actions. Since this paper allows only stream-dependent strategies, we assume in this paper that this requirement is imposed during collation.

Note that two 1-stage games, belonging to the same stage, may have the same set of voters and yet differ in the resulting payoff vectors. This can happen because the streams leading to these voters differ. However, if the voting scheme obeys additivity across stages (Assumption 8c), two such 1-stage games are strategically equivalent—their payoff vectors differ by a constant, which is the difference between the payoff vectors accumulated until these stage games were reached. This results in a great saving when attempting to construct equilibrium strategies which are *markovian*, namely, depend at each stage only on the set of voters and on the number of stages left, and not on the paths reaching these 1-stage games.

To sum up: *collation is a protocol, during which one assigns fixed moves*[31] *to all last stage games and then truncates these games, assigning the appropriate payoff vectors to the new endpoints. This is continued until the root is assigned a payoff vector. If one takes care to use the same moves at vertices corresponding to the same stream, up to that stage, then the resulting strategy profile will be only a function of the stream.*

THEOREM 4.6. *Let Γ be a game representing a k-stage voting scheme obeying general stream dependence (Assumption* 8b). *If, during collation, we always choose an equilibrium profile for each* 1-*stage game, the resulting profile is a subgame-perfect equilibrium profile for* Γ. *Conversely, every*

[30] This action can also be a mixed strategies at each stage, which together form a behavioral strategy. We shall use mixed strategy collation in Section 5.

[31] Pure, or mixed.

subgame-perfect equilibrium profile can result in this fashion. If the 1*-stage profiles are quasi-strong, then the resulting profile is quasi-strong* (*Definition* 2.2).

Proof. (A) Suppose that during collation, we always choose an equilibrium profile for a one-stage game. Let σ be the resulting profile for Γ. Let (σ_{-i}, τ_i) be an arbitrary profile resulting from a deviation by player i. We show that this deviation does not yield this player any benefit. Indeed, switching to σ_i at all last-stage subgames does not decrease his payoff, because σ_i is a best reply to σ_{-i} at all stage k games. After the switches, collate on the last-stage games and continue in the same fashion. Performing this procedure k times, we observe that player i's payment never decreases. Finally, we arrive at his original payment due to σ_i.

(B) Suppose that at each one-stage game the chosen profile was a quasi-strong equilibrium. Suppose now that a deviation τ occurred, subject to the restriction that at the start of every subgame, there was at least one voter who adhered to σ for that stage. Let i be an arbitrary agent, who was a member of the society when the deviation started, and we show that he cannot benefit from τ. Indeed, consider all k-stage subgames and instruct all the players to revert to σ. This will not harm agent i, because the profiles restricted to the last stage were quasi-strong. Collate on this stage and continue in the same fashion $k - 1$ times. One winds up with player i not harmed and getting the payment as in σ. Thus, τ does not benefit agent i.

(C) Let σ be a subgame-perfect profile. Its restriction to any last-stage subgame is an equilibrium profile. Collate on all the subgames of this stage and look at the games of stage $k - 1$. Again, σ restricted to this subgame (after collation) is an equilibrium, because σ was subgame-perfect. Continuing in this fashion, we see that σ was indeed obtained by the process of collation. ■

COROLLARY 4.7. *The following collation protocols yield all possible pure-strategy subgame-perfect equilibrium streams.*

Starting with the last stage and continuing backwards, as long as there are at least two voters, select a set of candidates that has the property that, if elected, no agent prefers to add candidates to this stage. If there is one voter, select for him a move that maximizes his payment. Having done that for a stage, perform collation and continue in the same fashion until all stages are exhausted.

Proof. Proposition 4.5 and Theorem 4.6. ■

Interestingly, subgame-perfect outcomes do not yield strict refinements to Nash equilibrium outcomes, as the following theorem shows.

THEOREM 4.8. *Let σ be a Nash equilibrium profile for a voting game Γ. There exists a subgame-perfect equilibrium profile $\hat{\sigma}$ yielding the same stream of members.*

Proof. Denote by τ the universal equilibrium profile for Γ as defined at the beginning of Section 3. Note that τ is a subgame-perfect profile. Let $\hat{\sigma}$ be equal to σ along the equilibrium path of σ and equal to τ otherwise. Both σ and $\hat{\sigma}$ have the same equilibrium path. Off the equilibrium path, $\hat{\sigma}$ is subgame-perfect, due to τ. If an agent has a profitable deviation from $\hat{\sigma}$ starting on the equilibrium path, he could have achieved it alone against σ_{-i} by switching to τ_i after the starting point of the deviation. ■

Consider again a one-stage multi-founder game. It may well happen that several sets S have the property that no founder would have preferred to add more candidates, given that they were elected. If such a set S_1 is contained in another such a set S_2, then the payment to each of the founders under S_2 is not greater than the payment under S_1, since otherwise a founder who would have preferred to vote for S_2 rather than for S_1 could have forced this outcome. Consequently, all sequentially Pareto-undominated equilibrium outcomes in a one stage game can be found through the common-voting procedure described in Proposition 4.5 but choosing only sets S that are minimal under inclusion. Similarly, we can obtain all subgame-perfect sequentially Pareto-undominated equilibrium payoffs in a multi-stage game by performing the construction of Theorem 4.6, but restricting ourselves at each stage to sets S that are minimal under inclusion. (Of course some equilibria reached by this construction may not be sequentially Pareto-undominated.)

We have characterized all equilibrium outcomes and narrow down the outcomes that should be searched when we want to achieve Pareto undominated outcomes. Nevertheless, the common voting strategies that we employed may be critized as being "unnatural." Consider for example the one-stage voting scheme obeying pure friendship and enmity (Assumption 8a), where there are two founders, agents 1 and 2, and two candidates, agents a and b. Suppose that a is a friend of 1 and b is a friend of 2. From what we said so far, the unique equilibrium outcome is $\{a, b\}$, where both voters vote for both candidates. But why would agent 1 support candidate b that he does not like and agent 2 vote for a? It is much more natural for each agent to vote only for his friend, which happens to also be an equilibrium profile. This, and more complicated examples, point to the need of obtaining equilibrium profiles that do not employ common voting. We present here a class of such profiles, of which the above profile is a particular case. This class employs common voting for a subset of the new members, which sometimes may even be empty. In view of Theorem 4.6, it is sufficient to consider 1-stage voting games.

PROPOSITION 4.9. *Let* Γ *be a* 1-*stage voting game having at least two founders. Let* S *be a set of candidates from* C^0, *having the property that, if elected, no original founder will prefer to add players to* S. *For each founder* i, *choose a set* P_i, *contained in* S, *that is a best response to*[32] $S \setminus P_i$. *Let* $C = S \setminus \bigcup_{j \in F^0} P_j$. *Finally, let* $V_i = P_i \cup C$. *Under these conditions,* $\{V_i: i \in F^0\}$ *is an equilibrium profile for* Γ.

The proof requires two lemmas:

LEMMA 4.10. *Let* P_i *be a best response of founder* i *against* $S \setminus P_i$, *where* S *is an arbitrary given set of candidates from* C^0 *containing* P_i. *If* $Q \subseteq S \setminus P_i$, *then* $P_i \cup Q$ *is also a best response of* i *to* $S \setminus P_i$.

Proof. Q is covered anyhow by $S \setminus P_i$, so it makes no difference whether i includes Q in his vote, or note. ■

LEMMA 4.11. *Let* P_i *be a best response of founder* i *against* $S \setminus P_i$, *where* S *is an arbitrary set of candidates containing* P_i. *If* $R \subseteq P_i$ *then* $P_i \setminus R$ *is a best response of* i *to* $(S \setminus P_i) \cup R$.

Proof. Voting $P_i \setminus R$ against $(S \setminus P_i) \cup R$ would yield player i the utility gained from S being elected. If voting for another set, Q, would yield him a higher utility, then voting $Q \cup R$ would be a better response to $S \setminus P_i$ then voting P_i, because $(Q \cup R) \cup (S \setminus P_i) = Q \cup (R \cup (S \setminus P_i))$. ■

Proof of Proposition 4.9. P_i is a best response of i against $S \setminus P_i$; therefore, V_i is a best response of i against $S \setminus P_i = (S \setminus P_i) \cap (\bigcup_{j \in F^0 \setminus \{i\}} V_j)$ (Lemma 4.10). By Lemma 4.11, $(C \cup P_i) \setminus \bigcup_{j \in F \setminus \{i\}} P_j$ is a best response of i against $\bigcup_{j \in F^0 \setminus \{i\}} V_j$. Invoking Lemma 4.10 once more, we find that V_i is a best response of i against $\bigcup_{j \in F^0 \setminus \{i\}} V_j$. ■

5. PERFECT EQUILIBRIA IN PURE STRATEGIES

Common-voting equilibria are usually not perfect. A voter may be tempted to deviate, figuring that the others will continue to vote in the same way with high probability, in order to extract some profit in case of "trembles." In this section we provide a sufficient condition for the existence of perfect equilibria in pure strategies and show how one can construct them (Proposition 5.2 and Theorem 5.12). We then show by examples that this condition is not necessary, as there are other cases in which pure-strategy perfect equilibria exist. Nevertheless, we show that for 2-stage games *with additive preferences across stages and within a stage*

[32] Such a set always exists, for example, $\varnothing$.

(*Assumption* 8c), pure-strategy perfect equilibria always exist (Theorem 5.7). Whether this result can be extended to games with more stages is still an open problem.

We are able to prove the main theorems of this section under the assumption that the voting scheme is *generic*, in the sense that *different streams yield different utilities for each player*. Example 5.6 shows that this assumption is necessary for the results.

DEFINITION. For a set $S \subseteq C^0$ we say that *i supports x* with respect to S if $S \succ_i S \setminus \{x\}$. Here, $\succ_i$ means "preferred by i."

The following lemma is easily proved by induction.

LEMMA 5.1. *For all* $n \geq 1$, *for all* $0 < \varepsilon < 1$, *it is true that* $1 - (1 - \varepsilon)^n \leq n\varepsilon$.

PROPOSITION 5.2. *Let* Γ *be a generic* 1-*stage multi-founder voting game. If S is a set of candidates,* V_i *is the set of candidates supported by founder i in S, and the strategy profile* $\{V_i\}_{i \in F^0}$ *is a Nash equilibrium with* $S = \bigcup_{i \in F^0} V_i$, *then* $\{V_i\}_{i \in F^0}$ *is a perfect equilibrium of* Γ.

Proof. Denote $c = |C^0|$, $f = |F^0|$. Denote by d the minimum payoff difference for any two sets of candidates and any founder. Similarly, denote by M the maximal payoff difference for any two sets of candidates and any founder, i.e.,

$$d = \min_{\substack{i \in F^0 \\ T_1, T_2 \subseteq C^0}} |u_i(T_1) - u_i(T_2)|, \qquad M = \max_{\substack{i \in F^0 \\ T_1, T_2 \subseteq C^0}} |u_i(T_1) - u_i(T_2)|. \tag{5.1}$$

The voting scheme is generic, and $c \geq 1$, therefore $d > 0$ and $M > 0$.

Assume fixed positive ε_1 and ε_2. Assume initially that they are each less than $\frac{1}{4c}$ and that $\varepsilon_2 \leq \varepsilon_1$. Additional conditions will be provided later.

Define the following completely mixed strategy for each founder i:

(1) For each $x \in V_i$, vote for $V_i \setminus \{x\}$ with probability ε_1.

(2) For any other set of candidates, except V_i, vote for this set with probability $\varepsilon_2/2^c$.

(3) Vote for V_i with the residual probability. This probability is greater than $1 - c\varepsilon_1 - \varepsilon_2$ as $|V_i| \leq c$, and from the restrictions already imposed on the epsilons it is greater than $\frac{1}{2}$.

As ε_1 and ε_2 tend to zero, this completely mixed strategy tends to V_i for every founder i. Let i be an arbitrary fixed founder. The proof will conclude if we show that V_i is his best reply against the others using these strategies, provided the epsilons are small enough. Consider two possible types of deviation by agent i. The first is a deviation that makes a difference when all others vote for the candidates they support, and the second is a deviation that makes no difference when all others vote for the candidates they support.

The first type of deviation causes a loss of at least d whenever all others vote V_j, and a gain of at most M in other cases. The loss occurs with a probability of at least $1/2^{f-1}$ (as this number is less than the probability of every other founder j voting V_j) and the gain can occur with a probability of at most

$$1 - (1 - c\varepsilon_1 - \varepsilon_2)^{f-1} \leq (f-1)(c\varepsilon_1 + \varepsilon_2) \leq cf(\varepsilon_1 + \varepsilon_2), \quad (5.2)$$

as at least one founder $j \neq i$ must vote for a set different from V_j. The first inequality is implied by Lemma 5.1.

A sufficient condition for the expected loss from such a deviation to exceed the expected gain is therefore

$$\frac{d}{2^{f-1}} \geq Mcf(\varepsilon_1 + \varepsilon_2), \quad (5.3)$$

and this always holds if $\varepsilon_1 < d/Mcf2^f$ as $\varepsilon_2 \leq \varepsilon_1$.

We now investigate the other type of deviations and find restrictions on the epsilons to ensure that it too will not be profitable.

Consider a deviation by agent i to $(V_i \setminus R) \cup A$, where $R \cup A \subseteq V_{-i}$, $R \subseteq V_i$, and $A \cap V_i = \varnothing$. Thus, player i removes members of R from his bid and adds members of A, and each of these candidates is supported by at least one other founder. Denote $Q = A \cup R \neq \varnothing$ and $V_i' = (V_i \setminus R) \cup A$.

There are three cases of bids of the other founders we now consider. The first, where V_i' gives a sure loss of at least d relative to V_i, the second, where a gain of up to M is possible, and the third, where the payoff to i from V_i and V_i' is the same.

The first case is when the others vote for $V_{-i} \setminus \{x\}$ for some $x \in Q$. Regardless of whether i supports x and does not vote for him ($x \in R$), or whether i does not support x and does vote for him ($x \in A$), the deviation to V_i' gives a loss of at least d compared to voting V_i. For each $x \in Q$ denote the probability of this subcase by $\eta_1(x)$.

The second case (possibility of gain) is when the vote of the others does not contain x for some $x \in Q$, but it is not $V_{-i} \setminus \{x\}$. Denote the probability of these subcases by $\eta_2(x)$ for each $x \in Q$. Note that the $|Q|$ such possibilities are not mutually exclusive. Note also that these two cases cover all situations where any member of Q is missing from V_{-i}.

Any other set voted for by the others (such a set must contain Q) gives the same payoff to i from both V_i and V_i'.

A sufficient condition for V_i' not to be a profitable deviation is that the expected loss is greater than the expected gain. A sufficient condition for this is

$$d \sum_{x \in Q} \eta_1(x) > M \sum_{x \in Q} \eta_2(x). \tag{5.4}$$

Let $m(x)$ be the number of supporters of x with respect to S, not including agent i. For all $x \in Q$ it is true that $m(x) \geq 1$.

The following bounds hold, as we explain:

$$\eta_1(x) \geq \varepsilon_1^{m(x)}(1 - c\varepsilon_1 - \varepsilon_2)^{f-m(x)-1} \geq \frac{\varepsilon_1^{m(x)}}{2^{f-m(x)-1}} \geq \frac{\varepsilon_1^{m(x)}}{2^f}. \tag{5.5}$$

The first inequality holds, as the event $\eta_1(x)$ includes the event that each supporter of x votes $V_j \setminus \{x\}$ and all others vote V_j. The second inequality is implied by $1 - c\varepsilon_1 - \varepsilon_2 > \frac{1}{2}$,

$$\begin{aligned}\eta_2(x) &\leq 1 - (1 - \varepsilon_2)^{f-1} + \varepsilon_1^{m(x)}\left(1 - (1 - c\varepsilon_1 - \varepsilon_2)^{f-m(x)-1}\right) \\ &\leq (f-1)\varepsilon_2 + \varepsilon_1^{m(x)}(f - m(x) - 1)(c\varepsilon_1 + \varepsilon_2).\end{aligned} \tag{5.6}$$

The first inequality holds, as for this case to occur, [at least one of the events at least one founder j votes for neither V_j nor $V_j \setminus \{y\}$ for any candidate y] which has probability no greater than $1 - (1 - \varepsilon_2)^{f-1}$, or [all the supporters j of x vote for $V_j \setminus \{x\}$ and at least one of the other founders j' of x votes for a set different from $V_{j'}$] must occur. The second inequality holds from two applications of Lemma 5.1.

If we now assume that $\varepsilon_2 \leq \varepsilon_1^{m(x)+1}/(f-1)$ then (5.6) implies

$$\eta_2(x) \leq \varepsilon_1^{m(x)+1} + \varepsilon_1^{m(x)}(f-1)(c+1)\varepsilon_1 \leq 2cf\varepsilon_1^{m(x)+1}. \tag{5.7}$$

Inequalities (5.5) and (5.7) together imply

$$\eta_2(x) \leq \eta_1(x)\frac{2cf\varepsilon_1^{m(x)+1}2^f}{\varepsilon_1^{m(x)}} = \eta_1(x)2^{f+1}cf\varepsilon_1. \tag{5.8}$$

Now, using inequality (5.8), inequality (5.4) is implied by

$$d \sum_{x \in Q} \eta_1(x) > M2^{f+1}cf\varepsilon_1 \sum_{x \in Q} \eta_1(x). \tag{5.9}$$

This is equivalent (since $\sum_{x \in Q} \eta_1(x) > 0$), to

$$\varepsilon_1 < \frac{d}{Mcf2^{f+1}}. \tag{5.10}$$

Taking all the restrictions together, and using the fact that $m(x) \leq f - 1$, for all $x \in Q$, we have that

$$\varepsilon_1 < \frac{d}{Mcf2^{f+1}}, \qquad \varepsilon_2 \leq \frac{\varepsilon_1^f}{f-1}, \tag{5.11}$$

imply that V_i is a best response to the mixed strategies of the others.

Since we can take a sequence of epsilons that tend to zero while keeping all the restrictions, the proof is complete, as we have a sequence of completely mixed strategy equilibrium profiles tending to $\{V_i\}_{i \in F^0}$. ■

Guessing a set of candidates S, for a single game Γ, that generates $\{V_i\}_{i \in F^0}$ that covers S, the furthermore, constitutes an equilibrium profile, might be a difficult task. Searching for all such S's makes it even more difficult. Even then, as we shall see (Example 5.5), we may not construct all pure-strategy perfect-equilibrium profiles. Sometimes, we are not interested in a specific voting game, but rather in a large class of games, and we wish to prove that every game in this class has a pure-strategy perfect-equilibrium profile. We may even want to characterize such a profile. In such cases, the following two corollaries might be useful. In fact, one of them will be employed subsequently.

COROLLARY 5.3. *Let Γ be a generic one-stage voting game. If there exists a set of votes $\mathscr{P} = \{P_i\}_{i \in F^0}$ where $P_i \subseteq C^0$, satisfying*

(1) *$\mathscr{P}$ is an equilibrium profile for Γ,*

(2) *$P_i \cap P_j = \emptyset$, whenever $i \neq j$,*

then Γ has a pure-strategy perfect-equilibrium profile.

Terminology. Profile $\mathscr{P}$, satisfying (1) and (2) above will henceforth be called *a generalized-partition equilibrium profile*.

Proof. Denote by S the union $\bigcup_{i \in F^0} P_i$. Let $V_i := \{x \in S$: x is supported by i with respect to $S\}$. It follows that $V_i \supseteq P_i$ because $\mathscr{P}$ is an equilibrium profile. For the same reason, $\{V_i\}_{i \in F^0}$ is an equilibrium profile. Therefore, $\{V_i\}_{i \in F^0}$ satisfies the conditions of Theorem 5.2 and constitutes a pure-strategy perfect-equilibrium profile. ■

This corollary is a special case of the following:

COROLLARY 5.4. *Let Γ be a generic one-stage voting game. Consider an arbitrary equilibrium profile $\{P_i \cup C\}_{i \in F^0}$, employing partial common voting* *as in Proposition* 4.9. *If* $P_i \cap P_j = \emptyset$, *whenever* $i \neq j$, *and every agent in* C *is supported by every voter, with respect to* $S = C \cup (\bigcup_{j \in F^0} P_j)$ *then* Γ *has a pure-strategy perfect-equilibrium profile.*

The proof is similar to the previous one and will be omitted.

Proposition 5.2 raises the question whether the conditions are also necessary for the existence of pure-strategy perfect equilibrium. We answer the question negatively, by the following example:

EXAMPLE 5.5. The population consists of

$$F^0 = \{1,2\}, \qquad C^0 = \{a,b\}.$$

There is only one period, $k = 1$. The utilities of the founders are

$$u_1(\emptyset) = 2, \quad u_1(\{a\}) = 3, \quad u_1(\{b\}) = 4, \quad u_1(\{a,b\}) = 1,$$
$$u_2(\emptyset) = 4, \quad u_2(\{a\}) = 2, \quad u_2(\{b\}) = 1, \quad u_2(\{a,b\}) = 3.$$

The payoff matrix is given by[33]

	∅	*a*	*b*	ab
∅	2, 4	3, 2	4, 1	1, 3
a	3, 2	3, 2	1, 3	1, 3
b	4, 1	1, 3	4, 1	1, 3
ab	1, 3	1, 3	1, 3	1, 3

In this example the pure equilibrium profiles are $(\{a\}, \{a,b\})$, $(\{b\}, \{a,b\})$ and $(\{a,b\}, \{a,b\})$. None of them satisfies the conditions of Proposition 5.2. Nevertheless, $(\{a\}, \{a,b\})$ and $(\{b\}, \{a,b\})$ are perfect equilibrium profiles.[34] This shows that Proposition 5.2 does not yield necessary conditions. On the other hand, Example 3.1 shows that voting schemes exist that do not have any pure-strategy perfect equilibrium. Providing a necessary and sufficient condition for the existence of pure-strategy perfect equilibrium in a 1-stage game remains an open question.

[33] For simplicity we omit the curly brackets that denote sets.

[34] Note that $(\{a\},\{a,b\})$ can be eliminated by successive weak domination.

The next example will show that the requirement that the game is generic is needed for Proposition 5.2 to hold.

EXAMPLE 5.6. The population is

$$F^0 = \{1,2\}, \qquad C^0 = \{a,b\}.$$

There is only one period, $k = 1$. The utilities of the founders are

$$u_1(\emptyset) = 2; \qquad u_1(\{a\}) = 1; \qquad u_1(\{b\}) = 3; \qquad u_1(\{a,b\}) = 1,$$

and

$$u_2(\emptyset) = 0; \qquad u_2(\{a\}) = 1; \qquad u_2(\{b\}) = 1; \qquad u_2(\{a,b\}) = 1.$$

The game is not generic, as, for example, $u_1(\{a\}) = u_1(\{a, b\})$. For $S = \{a\}$, founder 1 supports the empty set and founder 2 supports $\{a\}$. This voting profile is a Nash equilibrium. However, it is not a perfect equilibrium, as founder 1's strategy of $\emptyset$ is weakly dominated by voting $\{b\}$. This shows that requiring genericity is needed in Proposition 5.2. Note that (b, b) is a perfect Nash equilibrium which does support the conditions of Proposition 5.2.

An interesting application of Corollary 5.3 is the following:

THEOREM 5.7. *Let* Γ *be a game representing a* 2-*stage generic voting scheme*, *whose utilities obey additivity across stages and additivity within each stage* (*Assumption* 8c). *Under these conditions*, Γ *has a perfect equilibrium in pure strategies*.

Proof. Any perfect equilibrium profile for Γ must specify for each subgame of the second stage a profile under which each voter votes precisely for the set of his friends (who are not already in the society). This is a perfect equilibrium of the subgame (Section 2, case $k = 1$) and unique, by genericity. With this understanding, we can construct a 1-stage game Γ^1 by collation. The proof will be concluded if we show that Γ^1 has a pure-strategy perfect equilibrium, as the combination of this strategy with the continuation is a perfect strategy[35] for Γ. To achieve that, it is sufficient, by Corollary 5.3, to exhibit a generalized-partition equilibrium profile for Γ^1. This we are about to do by a construction under which voters add candidates to the society piecewise: There will be a variable set of candidates, called a *current set*, that grows, or stays put, as the voters add to it during the construction, until it eventually becomes the outcome for Stage 1, as well as an outcome of Γ^1. We introduce the following terminology: Let A be a current set of candidates. We say that a, possibly empty, set of candidates taken from $C^0 \setminus A$ is *optimal for voter* i w.r.t. A,

[35] A proof for any game representing a k-stage voting scheme is given in Theorem 5.12.

and denoted $X_i(A)$, if it is the best set of candidates that i could add to A, so as to increase his utility from the two stages. Note that $X_i(A)$ cannot contain enemies of i, since such candidates are enemies and can only contribute more enemies at Stage 2. (The friends of i will be brought in anyhow by i at Stage 2). In symbols, $X_i(A)$ is characterized by

$$w_i(A \cup X_i(A)) + w_i\big(\mathrm{en}_i\big(F^0 \cup A \cup \mathrm{fr}\big(F^0 \cup A \cup X_i(A)\big)\big)\big)$$
$$\geq w_i(A \cup B) + w_i\big(\mathrm{en}_i\big(F^0 \cup A \cup \mathrm{fr}(F^0 \cup A \cup B)\big)\big),$$
$$\text{all } B \subseteq \mathrm{fr}_i\big(C^0 \setminus A\big). \quad (5.12)$$

(In this calculation friends of i at the second stage are omitted from both sides of each inequality). Here, $w_i(T) := \sum_{t \in T} w_i(t)$, $\mathrm{fr}_i(S) := \{j\colon\ j \in \mathrm{fr}(i) \cap S\}$, $\mathrm{en}_i(S) := \{j : j \in \mathrm{en}(i) \cap S\}$ and $\mathrm{fr}(B) := \{l\colon l \in \mathrm{fr}(j)$ for some j in $B\}$. Sums over the empty set are considered equal to zero. By genericity, the set $X_i(A)$ is unique.

The Construction. Starting with a current set $A = \emptyset$, a referee approaches the voters repeatedly, one by one, and suggests to them to add candidates to the current set. Each approached voter i adds $X_i(A)$ and the set $A \cup X_i(A)$ becomes a new "current set" A. The referee continues to approach the voters, perhaps approaching a voter several times, taking care not to ignore voters whose optimal addition is not empty. This assures that after a finite number of approaches, there comes a situation when all optimal sets w.r.t. the current A are empty for all voters. At this the construction ends. This determines a pure-strategy profile $\{P_j\}_{j \in F^0}$, where P_j is the set consisting of all the members that voter j added along the construction.

It follows from the construction that $\{P_j\}_{j \in F^0}$ is a generalized partition of $S := \bigcup_{j \in F^0} P_j$. It remains to show that it is an equilibrium profile for Γ^1. To this end we require a lemma, which unfortunately is not true if $k > 2$:

LEMMA 5.8. *Assume the conditions and notations of Theorem* 5.7. *Let A and B be two sets of candidates, $A \subseteq B$. Let C be a set of friends of a voter i satisfying $C \cap B = \emptyset$. If $A \cup C \succ_i A$ then $B \cup C \succ_i B$.*

Proof. From the data it follows that the total weight of i from C exceeds the absolute value of the total weight of the new enemies that C brings at Stage 2.[36] When C is added to B he brings the same number of friends, namely $|C|$, and no new enemies. Perhaps even less—the previous ones that happen to be in $B \setminus A$. ■

[36] Namely, $w_i(C) + w_i(\mathrm{en}_i(\mathrm{fr}(C \setminus \mathrm{fr}(F^0 \cup A)))) > 0$.

Continuation of the Proof of Theorem 5.7. If $(P_j)_{j \in F^0}$ is not an equilibrium profile, then a voter i can benefit from a deviation. A deviation means that he deletes a set T of candidates from his vote P_i and adds a set Q of candidates not in S.[37] At least one of these sets is not empty. The set T, if not empty, is a union of nonempty sets $T_1, T_2, \ldots, T_r$, which are, respectively, subsets of his votes $P_i^1, P_i^2, \ldots, P_i^r$ taken when i was approached at times that we enumerate chronologically $1, 2, \ldots, r$. Denote by $S_1, S_2, \ldots, S_r$ the current sets at these times after his addition. Consider a hypothetical sequence when all founders vote as in the construction except that agent i votes $P_i^1 \setminus T_1$ at time 1, $P_i^2 \setminus (T_1 \cup T_2)$ at time $2, \ldots, P_i^r \setminus T$ at time r, and at the first time he also adds the candidates of Q. The end of this sequence is the deviation, which, as we assumed, benefited player i. We now modify this sequence in such a way that player i will continue to benefit and at least as much. To this end, add T_1 to the hypothetical vote of voter i at all times, starting from time 1. This will benefit him at time 1. Indeed, he would benefit if the current set were $S_1 \setminus T_1$ because $X_i(S_1 \setminus P_i^1) = S_1$ is the unique optimal response and so, by Lemma 5.8, he would benefit by adding T_1 to $(S_1 \setminus T_1) \cup Q$. For the same reason i would benefit by adding T_1 at every part of the hypothetical sequence, since $S_1 \setminus T_1 \subseteq Q \cup (S_t \setminus (T_1 \cup T_2 \cdots \cup T_t))$ and $T_1 \cap (Q \cup (S_t \setminus (T_1 \cup T_2 \cdots \cup T_t)) = \emptyset$, $t \in \{1, 2, \ldots, r\}$. After adding T_1 we are in an improved deviation that starts at time 2. We make a similar modification and continue for r times. Eventually, we arrive at an improved deviation at which only Q is added. But this is impossible, since the original construction ended when no voter could beneficially add members outside the current set. The contradiction shows that we are indeed at equilibrium. ■

The construction in the above proof is not specific about the order in which the referee approaches the voters. We are going to show that although different orders yield different equilibrium profiles, the outcome S remains the same. Therefore, the *perfect equilibrium profile* that is generated as described in Proposition 5.2 is the same, regardless of the order of approach.

LEMMA 5.9. *If* $A \subseteq B \subseteq C^0$, *then* $A \cup X_i(A) \subseteq B \cup X_i(B)$ *for every agent i in* F^0.

Proof. Assume negatively, that for some i in F^0, $D := (A \cup X_i(A)) \setminus (B \cup X_i(B)) \neq \emptyset$. By optimality of $X_i(A)$ and genericity of Γ, it follows

[37] It is irrelevant if he also votes for agents in $S \setminus P_i$, so we assume that he does not.

from (5.12), replacing B by $X_i(A) \setminus D$, and noting that $D \cap A = \varnothing$, that

$$w_i(D) + w_i\big(\mathrm{en}_i\big(F^0 \cup A \cup \mathrm{fr}\big(F^0 \cup A \cup X_i(A)\big)\big)\big)$$
$$- w_i\big(\mathrm{en}_i\big(F^0 \cup A \cup \mathrm{fr}\big(F^0 \cup A \cup (X_i(A) \setminus D)\big)\big)\big)$$
$$= w_i(D) + w_i\big(\mathrm{en}_i\big(\mathrm{fr}\big(F^0 \cup A \cup (X_i(A) \setminus D)\big)\big)\big) > 0. \quad (5.13)$$

Using (5.12) once more, replacing A, $X_i(A)$, B by B, $X_i(B)$, $X_i(B) \cup D$, respectively, we obtain

$$w_i(D) + w_i\big(\mathrm{en}_i\big(\mathrm{fr}(D) \setminus \mathrm{fr}\big(F^0 \cup B \cup X_i(B)\big)\big)\big) < 0. \quad (5.14)$$

However, $(A \cup X_i(A)) \setminus D \subseteq B \cup X_i(B)$, and enemies of i carry negative utilities; therefore, the left side of (5.14) is not smaller than the left side of (5.13)—a contradiction. ■

COROLLARY 5.10. *Changing the order of the referee's approaches leads to the same final set S, although the actual votes of the players may be different.*

Proof. Let $\varnothing = T^0, T^1, \dots, T^r = T$ be the sequence of "current sets" generated by a different order of approaches. We shall show that $T^m \subseteq S$ for every m and therefore $T \subseteq S$. Reversing the roles of S and T one gets $S \subseteq T$ and this concludes the proof. Proceed by induction: certainly $T^0 \subseteq S$. Suppose $T^{m-1} \subseteq S$ and $T^m \nsubseteq S$. Then some i in F^0 has $X_i(T^{m-1}) \nsubseteq S$. Thus, a candidate a exists in $X_i(T^{m-1})$, $a \notin S$. From Lemma 5.5, $a \in X_i(T^{m-1}) \subseteq X_i(S)$, which contradicts the fact that the construction terminates when $X_i(S) = \varnothing$ for all i. ■

One may now ask whether a perfect equilibrium profile is always unique under the conditions of Theorem 5.7. The following example settles this question negatively.

EXAMPLE 5.11. The set of founders is $F^0 = \{1, 2\}$. The set of candidates is $C^0 = \{a, b, c\}$. $k = 2$ and we assume pure friendship and enmity (Assumption 8a). fr(1) = $\{a\}$, fr(2) = $\{b\}$, fr(a) = fr(b) = $\{c\}$. The construction if Theorem 5.7 leads to $S = \varnothing$. However, it can be checked that 1 and 2 voting for their friends at all stages and a and b voting for their friend at Stage 2 is also a perfect equilibrium profile.

We conclude this section by extending Proposition 5.2 to several-stage voting schemes.

THEOREM 5.12. *Let Γ be a game representing a k-stage generic voting scheme, obeying general stream dependence (Assumption* 8b). *If, during collation, we always manage to choose a pure-strategy perfect-equilibrium* *profile at each one-stage game,*[38] *the resulting strategy is perfect for* Γ. *Conversely, every pure-strategy perfect profile for* Γ *can be obtained by collation in this fashion.*

Proof. The game Γ is a game of perfect recall, therefore, by Kuhn's (1953) theorem (see also Selten, 1975), we can work only with behavioral strategies.

We regard Γ as given in extensive form. Denote by $T_{t,r}$ the 1-stage tree that corresponds to the rth tree[39] of stage t. Denote by $\hat{\Gamma}_{t,r}$ the subgame of Γ that starts with $\Gamma_{t,r}$.

Collation with respect to a strategy τ converts the 1-stage tree $\Gamma_{t,r}$ to a 1-stage game $\Gamma_{t,r}(\tau)$, where, even if τ happens to be defined on all of Γ, we mean here its restriction *to the subgame* $\hat{\Gamma}_{t,r}$ *excluding the first stage of this subgame*. $\Gamma_{t,r}(\tau)$ is a 1-stage game, so, if its voters employ a strategy profile[40] $(\rho)^1$, we denote by $h_{t,r}(\tau)(\rho)^1$ the payoff vector that results.

Note that

$$h_{t,r}(\tau)(\rho)^1 = \hat{h}_{t,r}\big((\rho)^1(\tau)\big), \tag{5.15}$$

where the right side is the payoff vector that results when the agents play $\hat{\Gamma}_{t,r}$, using ρ at the first stage and then continue with τ. Note that for $(t,r) = (1,1)$, $h_{1,1}(\tau)(\rho)^1$ is the expected payoff in Γ if $(\rho)^1$, followed by τ, is played.

(A) Assume that σ is constructed backwards, by collation, such that at each 1-stage game, a pure-strategy perfect profile is chosen. We require that identical 1-stage moves are chosen at $\Gamma_{t,r}$'s with identical histories, i.e., identical streams of members until stage t (see Assumptions 5 and 6).[41] Then, for each (t,r), there exists a test sequence $(\sigma_{t,r}^m)_{m=1}^{\infty}$ of completely mixed 1-stage strategy profiles converging to the restriction $\sigma_{t,r}$ of σ to $\Gamma_{t,r}$, such that for every agent i,

$$h_{t,r;i}(\sigma)\left(\sigma_{t,r;-i}^m, \sigma_{t,r;i}\right)^1 \geq h_{t,r;i}(\sigma)\left(\sigma_{t,r;-i}^m, \sigma_i'\right)^1, \tag{5.16}$$

whenever σ_i' is a pure 1-stage move different from $\sigma_{t,r;i}$. Again, to ensure that eventually strategies depend only on histories, identical $\sigma_{t,r}^m$ should be chosen at $\Gamma_{t,r}$'s that result from the same stream up to stage t. This is always possible, because of the way σ was constructed.

[38] Proposition 5.2 and Corollaries 5.3 and 5.4 might be useful here.

[39] Counted, e.g., from left to right.

[40] The superscript 1 comes merely to remind us that ρ is a 1-stage profile, or a restriction to a 1-stage profile, if ρ happens to be defined in a larger game.

[41] At the expense of somewhat more technicality, a similar theorem can be proved even if the strategies are more complicated than those used in this paper.

Let us examine the payments at endpoints of $\Gamma_{t,r}(\tau)$, where τ is an arbitrary pure strategy that depends only on histories. Since Γ is assumed to be generic, payments to an agent i at two endpoints of $\Gamma_{t,r}(\tau)$ are different unless, and only unless, one of the following cases occurs:

(i) Agent i is not a voter in $\Gamma_{t,r}$. Observe that in this case, whatever agent i "does" against whatever the other agents are doing in this game, or in a different continuation, is always a best reply. Or

(ii) The two endpoints result from the same 1-stage "stream" in $\Gamma_{t,r}$. In this case τ, and any pure, or mixed (behavioral) strategy, that depends only on histories, specify the same continuation following these points. Indeed, to every path that follows one endpoint, there corresponds a path following the other endpoint that specifies the same stream of members.

It follows that $\Gamma_{t,r}(\sigma)$ has the same payment to agent i at two endpoints then $\Gamma_{t,r}(\sigma^m_{-i}, \sigma_i)$ has the same payments to agent i, for every $m \in \{1, 2 \dots\}$, where σ^m denotes the aggregate of the $\sigma^m_{t,r}$'s.

Now, as m gets larger, the payoff vectors at the endpoints of $\Gamma_{t,r}(\sigma^m_{-i}, \sigma_i)$ approach those of $\Gamma_{t,r}(\sigma_{-i}, \sigma_i)$, and are equal whenever the latter are equal. Therefore, if $(\sigma_{t,r;i})^1$ is a best response to $(\sigma_{t,r;-i})^1$ in $\Gamma_{t,r}(\sigma_{-i}, \sigma_i)$, then it remains a best response in $\Gamma_{t,r}(\sigma^m_{-i}, \sigma_i)$, if m is large enough. Thus,

$$h_{t,r;i}(\sigma^m_{-i}, \sigma_i)(\sigma^m_{t,r;-i}, \sigma_{t,r;i})^1 \geq h_{t,r;i}(\sigma^m_{-i}, \sigma_i)(\sigma^m_{t,r;-i}, \sigma'_i)^1, \quad (5.17)$$

for all strategies σ'_i.

Now, $(\sigma^m)^\infty_{m=1}$ is a sequence of completely mixed (behavioral) strategies. Our proof will conclude if we prove that for every agent i and every large enough m, σ_i is a best reply to σ^m_{-i}. Indeed, let σ'_i be an arbitrary pure strategy for agent i in Γ. By (5.15), it yields agent i the payoff $h_{1,1;i}(\sigma^m_{-i}, \sigma'_i)(\sigma^m_{1,1;-i}, \sigma'_{1,1;i})^1$ against σ^m_{-i}.

Working backwards, we instruct agent i to switch to $\sigma_{t,r;i}$ at each stage. By (5.17), this will not decrease his payment at each 1-stage game, that sequentially becomes $\Gamma_{t,r}(\sigma^m_{-i}, \sigma_i)$. The final payoff to agent i eventually becomes $h_{1,1;i}(\sigma^m_{-i}, \sigma_i)(\sigma^m_{1,1;-i}, \sigma_{1,1;i})^1$. We have proved that

$$h_{1,1;i}(\sigma^m_{-i}, \sigma_i)(\sigma^m_{1,1;-i}, \sigma_{1,1;i})^1 \geq h_{1,1;i}(\sigma^m_{-i}, \sigma'_i)(\sigma^m_{1,1;-i}, \sigma'_{1,1;i})^1. \quad (5.18)$$

So, by (5.15), σ_i is indeed a best reply in Γ, to σ^m_{-i}, whenever m is large enough, and σ is a pure-strategy perfect profile in Γ.

(B) Conversely, suppose that σ is a pure-strategy perfect profile for Γ. Then there exists a sequence $(\sigma^m)^\infty_{m=1}$, of completely mixed behavioral strategies, converging to σ, such that for each agent i and for each m, σ_i

is a best reply to σ^m_{-i}. This is true also in every subgame $\hat{\Gamma}_{t,r}$. Invoking (5.15), we find that

$$h_{t,r;i}(\sigma^m_{-i}, \sigma_i)(\sigma^m_{t,r;-i}, \sigma_{t,r;i})^1 \geq h_{t,r;i}(\sigma^m_{-i}, \sigma'_i)(\sigma^m_{t,r;-i}, \sigma'_{t,r;i})^1, \quad (5.19)$$

for every pure-strategy σ'_i of agent i, where, $\sigma^m_{t,r}$ is the restriction of σ^m to the tree $\Gamma_{t,r}$. In particular, this is true if σ'_i differs from σ_i only at $\Gamma_{t,r}$. Thus,

$$h_{t,r;i}(\sigma^m_{-i}, \sigma_i)(\sigma^m_{t,r;-i}, \sigma_{t,r;i})^1 \geq h_{t,r;i}(\sigma^m_{-i}, \sigma_i)(\sigma^m_{t,r;-i}, \sigma'_{t,r;i})^1, \quad (5.20)$$

for every pure-strategy 1-stage deviation $\sigma'_{t,r;i}$.

Now, the voting scheme is generic, so the payments corresponding to distinct endpoints of $\Gamma_{t,r}(\sigma)$ are different for any agent i, who is a voter, whenever these endpoints represent different 1-stage streams. As explained in the previous part, those endpoints that represent the same 1-stage stream have the same payoff vectors both in $\Gamma_{t,r}(\sigma)$ and in the games $\Gamma_{t,r}(\sigma^m_{-i}, \sigma_i)$, $m = 1, 2, \ldots,$ because all strategies depend only on streams. Thus, there exists a d, such that for all $m \geq d$, the relative order among the payments to each voter at $\Gamma_{t,r}(\sigma)$ and at $\Gamma_{t,r}(\sigma^m_{-i}, \sigma_i)$, are identical. We conclude that $\sigma_{t,r;i}$ is a best reply against $\sigma^m_{t,r;-i}$ in both games. In particular,

$$h_{t,r;i}(\sigma)(\sigma^m_{t,r;-i}, \sigma_{t,r;i})^1 \geq h_{t,r;i}(\sigma)(\sigma^m_{t,r;-i}, \sigma'_i)^1, \quad (5.21)$$

for every 1-stage deviation σ'_i of agent i. This means that the restriction of $(\sigma^m)^\infty_{m=d}$ to $\Gamma_{t,r}(\sigma)$, is an appropriate test sequence and $\sigma_{t,r}$ is therefore perfect for the 1-stage game $\Gamma_{t,r}(\sigma)$, obtained by collation with respect to σ. ■

REFERENCES

Aumann, R. J. (1959). "Acceptable Points in General Cooperative n-Person Games," in *Contribution to the Theory of Games IV*, Annals of Mathematics, Vol. 40, pp. 287–324. Princeton: Princeton Univ. Press.

Aumann, R.J., and Myerson, R. B. (1988). "Endogenous Formation of Links Between Players and of Coalitions: An Application of the Shapley Value," in *The Shapley Value. Essays in Honor of Lloyd S. Shapley* (A. Roth, Ed.), pp. 175–191. Cambridge: Cambridge Univ. Press.

Banerjee, S., Konishi, H., and Sönmez, T. (2001). "Core in a Simple Coalition Formation Game," *Social Choice and Welfare*, to appear.

Barberà, S., Sonnenschein, H., and Zhou, L. (1991). "Voting by Committees," *Econometrica* **59**, 595–609.

Bogomolnaia, A., and Jackson, M. O. (2000). The stability of hedonic coalition structures, mimeo, Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California.

Chwe, M. S.-Y. (1994). "Farsighted Coalitional Stability," *J. Econ. Theory* **63**, 299–325.

Drèze, J., and Greenberg, J. (1980). "Hedonic Coalitions: Optimality and Stability," *Econometrica* **48**, 987–1003.

Jéhiel, P., and Scotchmer, S. (1997). "Free Mobility and the Optimal Number of Jurisdictions," *Ann. Econ. Statist.* **45**, 219–231.

Kohlberg, E., and Mertens, J. F. (1986). "On the Stability of Equilibria," *Econometrica* **54**, 1003–1039.

Kuhn, H. W. (1953). "Extensive Games and the Problem of Information," in *Contributions to Game Theory II* (H. W. Kuhn and W. Tucker, Eds.), Annals of Mathematics Studies, Vol. 28, pp. 193–216. Princeton: Princeton Univ. Press.

Milchtaich, I., and Winter, E. (2001). "Stability and Segregation in Group Formation," *Games and Economic Behavior*, to appear.

Roberts, K. (1999). Dynamic voting in clubs, mimeo, London School of Economics.

Selten, R. (1975). "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *Int. J. Game Theory* **4**, 25–55.

Sobel, J. (1998). "On the Dynamics of Standards," Discussion Paper 97-15R, University of California at San Diego.

Sobel, J. (2000). "A Model of Declining Standards," *International Econ. Review* **41**, 295–303.