Int J Game Theory (2002) 31:155–202

International Journal of Game Theory

© Springer Verlag 2002

Voting for voters: the unanimity case*

Daniel Granot[†], **Michael Maschler**[††], **Jonathan Shalev**[†††]

[†] Faculty of Commerce and Business Administration, University of British Columbia Vancouver, B. C., Canada V6T 1Y2 (e-mail: daniel.granot@commerce.ubc.ca)
[††] Department of Mathematics and Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem Jerusalem 91904, Israel (e-mail: maschler@vms.huji.ac.il)
[†††] 10 Gonen Street Ra'anana 43399, Israel (e-mail: shalev@witty.com)

Received: December 2001

Abstract. We present a simplified model of the evolution of a society which is regulated by a formal unanimity voting procedure. We examine several protocols, which depend on whether admission or expulsion are permissible, and on the order with which they are implemented. Conditions which ensure the existence of pure-strategy perfect equilibrium profiles for some voting protocols, and counter examples for the existence of such profiles in other protocols are presented. Finally, we prove that, if the agents insist on perfect equilibrium strategy profiles in a one-stage play, the original founders would prefer a protocol in which expulsion precedes admission to protocols in which either admission precedes expulsion, or the two are treated simultaneously. The paper concludes with an overview and a discussion on the results and suggestions for further research.

Journal of Economic Literature Classification: C7, D7, D71.
Mathematical Review Classification: 90D06, 90D10, 90D20, 90D40, 90D50, 90D55, 90D80.

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

1. Introduction

Noncooperative game theory has proved successful in analyzing the evolution of animal and plant species. This is due to Darwin's principle of the survival

* The authors thank Salvador Barberà for his critical reading and his many suggestions. The authors are also grateful to the anonymous referee for many very useful suggestions.
Support by a Natural Sciences and Engineering Research Council of Canada grant and by the Center for Rationality and Interactive Decision Theory of the Hebrew University of Jerusalem is gratefully acknowledged.

of the fittest which allows one to assume that species are "programmed" to act rationally, otherwise they will become extinct.

Human societies[1] show similar "evolution characteristics". Sometimes they flourish, other times they barely exist, some of them simply disappear or lose their identity. Thus, one hopes that noncooperative game theory might be useful to study aspects of the "evolution" of human societies.

There is, however, a slight difference between the purpose of studying the evolution of species and that of human groups. The study of plant and animal "societies" is mainly descriptive – one likes to explain why it evolved along one path and not another. Sometimes though, there are normative issues, such as when one considers preservation of nearly extinct species, or when one is interested in improving a stock. In human societies the normative aspect is as important as the descriptive part. One may be interested in recommendations that lead to the success of one company, or in the explanation why another company went bankrupt.

When one observes the evolution of a certain society and realizes that it disintegrated because of internal frictions, one may be inclined to think that this came about because the members of the society were not playing an equilibrium profile. Thus, a useful recommendation may be to examine equilibrium selections. It is here that noncooperative game theory may play a useful role. One then asks how "rational" players could, or should, have acted for their own benefit, given that this is also good for the entire society.

Animals and plants "enter" their societies via sexual or vegetative reproduction. They leave their society by death. Humans enter or leave a society in many other ways. One of them is via formal voting procedures. It is this method that is studied here; hence the title: "voting for voters". The first paper on this subject was written by Barberà et al. [1], where only admission was considered, which was subject to the quota-1 rule.[2] In this paper we study another model, where both admission and expulsion are governed by a unanimity rule. Namely, a rule which specifies that in order to admit a candidate, all voters must agree and in order to expel a member, all but that member must agree. The reason for this choice is explained below.

There are two main approaches to study societies which are regulated by formal voting procedures: Under one approach, one attempts to model a given real society as much as possible while retaining technical feasibility, then study the model, draw predictions and formulate recommendations. We are not following this approach, but we certainly recognize its importance. Another approach aims at studying general issues and their impact on the evolution of the society. When one takes such an approach, one may construct a very simple model, artificial perhaps, that isolates as much as possible a single issue and studies its influence on the society's evolution. This is the approach taken in this paper. We allow many simplifying assumptions in order to isolate a main issue. Thus, we assume that the society is fixed, composed only of original members and original candidates. Nobody is born or dies. We also do not permit resignations. We consider only a finite horizon model under which votes take place at a finite number of stages. Thus, our model assumes that the

[1] By *society* we mean any group of people who have some common interests, or traits. A society can be a small company, a conglomerate, a nation, a religious sect, a group of nations, etc.

[2] Namely a rule under which each voter can bring into the society at any stage any group of candidates that he wishes.

voting scheme consists of a known number of stages, after which the society disintegrates, having completed the goals for which it was created. The single issue of concern in our model is whom to admit and/or whom to expel at each stage. We are looking for pure-strategy Nash equilibrium profiles and their refinements.

When one votes to admit/expel an agent, one should take into consideration several factors:

(1) The contribution of the agent to the society if he becomes a member. This can be measured by the (positive, or negative) utility he adds to the group, throughout the periods that he is in the society.
(2) His position on admission/expulsion of other agents.
(3) The influence of his presence on other voters.
(4) The likelihood of him helping the voters recruit/expel other agents.[3,4]

In this paper we want to concentrate precisely on the last item. We want to study the rational behavior of a voter if he knows that any candidate that is admitted can veto his decisions. For this reason we impose a unanimity rule. This puts the issue into focus: Asking for less, for instance that decisions are adopted by a majority rule, would allow a member to admit a candidate who may vote against him in the future, figuring that he may still have enough members to produce favorable decisions for him.

There are several real-life cases where decisions are adopted by a unanimity rule. New membership into NATO and the European Union (EU) require unanimous consent by all member states. Any changes to existing treaties in the EU require unanimous consent and decisions concerning, e.g., foreign policy and defense issues (the so-called second and third pillar topics) require unanimous consent by all member states in the European council. No member state can opt out of the EU (which resembles our no resignation requirement). Further, Articles 6 and 7 in the EU treaty consider the situation where a member state consistently, and over an extended period, violates basic democratic and human-rights principles. In such cases, the rights of a violating member state, but not its obligations, can be suspended. Such an action requires, among other measures, the unanimous consent of all member states in the EU council, with the exception, of course, of the consent of the violating member state. Jury decisions in criminal trials must be unanimous[5] and admissions to some clubs and sororities require unanimous consent. Sometimes, votes taken in this case were performed by a blackball method. Finally, trustees' decisions must be taken by unanimous consent and should abide by the trust document. A trust document sometimes allows for the admission of a

[3] E.g., the agent may not desire to admit/expel another agent, which is a subject of consideration in (2) above, but nevertheless he will not vote against admission/expulsion in order to help other voters.

[4] There are many other very important issues that we do not discuss in this paper, such as: What will a candidate do if he is not in the society. Will he start a competing society? Are several members going to decide to resign and form their own society?

[5] A criminal trial can be regarded a 'voting for voters': If e.g., a jury sends a person to several years of prison, one can look at it as expelling him from the society, say, composed of the persons in the country and admit him at a later period. Of course it differs from our model in several respects. For example, the set of voters does not coincide with the set of all members of the society.

new trustee or the expulsion of an existing trustee and both these decisions must be reached by a unanimous consent. Naturally, unanimity has many drawbacks. It is sometimes considered undemocratic and it may lead to undesirable stalemates. We note, however, that decision making by consensus, or the so-called "participative approach to decision making", wherein a decision is reached only after *everyone* who may feel its impact has given his approval, is at the heart of the Japanese style of management (see Ouchi [5]).

There is some duality between the quota-1 rule and the unanimity rule: Voting by a unanimity rule means that a single member can veto an admission/expulsion. So, it is in a sense a 1-quota for vetoing. This analogy, however, is limited to one stage. When one votes for a candidate under quota-1 rule, this candidate becomes a member of the society and he then participates in future votings. When one vetoes the admission of a candidate, this candidate will not become a voter. Another kind of duality is also interesting: Voting for a candidate under the quota-1 rule is compared to vetoing an expulsion of a member. In this case the agent is a voter in the next stage. We were able to use these analogies in order to extend several results in Barberà et al. [1].

When one considers both admission and expulsion, the order of votes may be important. Accordingly, we study several protocols: Those in which only admissions exist, others where only expulsions exist, protocols where both admissions and expulsions take place simultaneously and protocols in which each stage consists of two periods: admission in the first period and expulsion in the second period, or the other way around. Some comparisons between the various protocols are carried out in this paper.

The paper is organized in the following fashion: We describe the model and the various protocols in Section 2. In Section 3 we present a few simple examples that show that sometimes one may wish to admit an undesirable candidate and to expel a desirable one. We also show that some equilibrium profiles are not plausible, pointing to the need of careful selections. It is proved in this section that except for some degenerate cases involving few voters, pure-strategy equilibrium profiles always exist in one stage protocols. In Section 4 we characterize all outcomes in a single stage voting schemes that can be obtained by pure-strategy Nash equilibrium profiles, and in Section 5 we explain how to obtain all subgame perfect equilibrium profiles in a multi-stage voting scheme.

The rest of the paper concentrates on pure-strategy perfect equilibrium profiles. Often such profiles do not exist, but we are able to provide in Section 6 a general sufficient condition for the existence of such profiles. Section 7 shows how to perform a kind of backward induction to obtain a pure-strategy perfect equilibrium profile in multi-stage games. The results in Sections 6 and 7 enable us to prove in Section 8 the following results concerning voting schemes in which utilities are separable and additive:[6] There always exists a pure-strategy perfect equilibrium profile if only admission of candidates is allowed. It consists of sincere voting.[7] It is not unique and this fact can lead to the existence of pure-strategy perfect equilibrium profiles which incorporate "punishments". Sincere voting in a one-stage scheme is always a perfect equi-

[6] See definition in Section 2.

[7] "Sincere voting" means that one votes at each stage to admit precisely all candidates whose presence adds to his utility at that stage and, in other protocols, also to expel precisely those that diminish his utility at that stage. The term is relevant only when utilities are additive across stages.

librium profile in the other protocols that consist of one period at each stage, but not if the number of stages exceeds one. Concerning the two-period protocols and assuming that utilities are determined by weights,[8] interestingly, even though there always exists a pure-strategy perfect equilibrium profile in a one stage voting scheme in which agents first vote for expulsion and then for admission, this is not the case if the order is reversed. In Section 8 we also present a voting scheme that has a unique perfect equilibrium profile which is Pareto undominated, however this profile does not make intuitive sense. This surprising example suggests that refinements together with Pareto undomination do not always provide a reasonable equilibrium selection. Thus, in order to make a reasonable recommendation in each particular case, one has to examine and choose from the set of *all* equilibrium outcomes.

Sometimes the founders of a society have the option of deciding which protocol to follow. It is then interesting to know who stands to gain under the various protocols. In Section 9 we are able to show that in a one-stage play, if perfect equilibrium profiles are considered, founders who survive the first-expel then-admit protocol always survive both the converse order protocol and the simultaneous protocol and are weakly better off in the first protocol.

The main purpose of this paper is to present the dynamic models and show how various features and properties can be deduced using game-theoretic concepts and techniques. The study covers several topics with mathematical sophistication ranging from simple perhaps tedious proofs to more involved and interesting ones. Section 10 provides a general overview of the various results and discusses the limitations that we imposed. It also offers suggestions for further research. Some readers may prefer to look at this section before concentrating on details.

2. Finite horizon voting schemes requiring unanimity

In this section we describe several *finite horizon voting schemes* and interpret them as games. These games are then analyzed in subsequent sections. For additional motivation and more detailed discussion, the reader is referred to Barberà, Maschler and Shalev [1], where a related model is treated.

We consider a fixed set N of *agents* which is originally partitioned into two subsets: F^0 – the set of *original founders* of a *society* and a set C^0 – the set of *original candidates* for admission into the society. In general, we refer to agents belonging to the society as *members* or *voters*, and agents outside the society as *candidates*. The voting process consists of k *stages*, where, at each stage, members vote simultaneously to *admit* candidates and/or to *expel* members. After k stages the society dissolves, having concluded the tasks for which it was created. The voting procedure is subject to the following rules:

(1) Fixed population. No agent is added or deleted during the play.
(2) Unanimous consent for admission and compulsory membership. A candidate at stage t is admitted to the society if, and only if, all members at that stage, who are allowed to vote, vote for his admission. An agent who is offered admission always agrees/must become a member.

[8] See definition in Section 2.

(3) Unanimous consent for expulsion. A member of the society at stage t, who is not alone, is expelled if and only if all other members at that stage, who are allowed to vote, vote to expel him.
(4) No resignation. Resignation is not allowed in this model. Resignation would complicate the analysis considerably. We require this restrictive assumption for simplification and since our objective is to concentrate only on the effects of admission and expulsion.

We consider several protocols.

(5a) Admission only (A). Only admissions take place. No members, including those admitted, can be expelled. Votes are conducted simultaneously at the beginning of each stage. Admitted agents can vote only starting at the next stage.
(5b) Expulsion only (E). Only expulsions take place. Thus, an expelled person cannot become a member again. Votes are conducted simultaneously at the beginning of each stage. Members expelled cease to vote as of the next stage.
(5c) Simultaneous admission and expulsion (SIM). At each stage, the members vote simultaneously whom to expel from among themselves, and whom to admit from the candidates not already in the society. Voting for expulsion and admission takes place simultaneously. The newly admitted members can vote only starting at the next stage and expelled members cease to vote as of the next stage.
(5d) Admission, then expulsion (AE). Each stage consists of two periods – the A-period followed by the E-period. At the A-period the members vote simultaneously whom to admit. Then, together with the new members, they vote simultaneously in the E-period whom to expel. Expelling members who were just admitted at this stage is forbidden.[9]
(5e) Expulsion, then admission (EA). Each stage consists of two periods – the E-period and the A-period. At the E-period the members vote simultaneously whom to expel. Those that remain, then vote simultaneously in the A-period whom to admit. Admitting a person who was just expelled at this stage is forbidden.[10]

One can think of several other protocols, but we limit ourselves only to those described above in order not to raise additional issues.[11]

Part of the outcome up to stage t is a *stream of members*, namely $h^t := (F^0, F^1, \ldots, F^{t-1})$. These will be referred to as *stream histories*. In general, however, histories contain more information; namely, who voted to admit or to expel whom at previous stages. In the two-period protocols (AE) and (EA)

[9] We impose this restriction in order to rule out an admission of a highly non-desirable agent merely because he may prevent some members from being expelled at the same stage. Of course, they can be expelled at later stages.
[10] We impose this restriction in order to rule out an expulsion of a highly desirable agent for the sole purpose of prohibiting him to prevent another desirable person from being admitted at the same stage. Of course, an expelled agent can be admitted at later stages.
[11] For example, one can think of a protocol in which voting whom to admit is taken sequentially for each candidate. This raises immediately the issue of deciding the order in which the votes should take place.

it contains also the results of the votes at each period. Such histories will be called *full histories*.[12]

We continue with the description of the model.

(6) Known histories. We assume that the full histories are known to every agent as soon as they take place, and are, in fact, common knowledge.
(7) Strategies are functions of the full histories. In some cases, it may be sufficient to employ strategies that are only functions of the stream histories. In other cases, especially when "punishments" are involved, we must employ strategies that make use of the full histories.

We have specified all the ingredients needed to construct a game form for our voting schemes. In order to convert it into a game, we have to specify utilities.

(8) Utilities. We assume that the priorities of agent i over the outcomes are given by a complete and transitive binary relation, and therefore they can be represented by a utility function u_i. Later, when we deal with mixed strategies, we assume that these utilities are in fact von Neumann – Morgenstern utilities.

    We assume that agents care[13] only about the stream of members $\mathscr{F} := (F^0, F^1, \ldots, F^k)$. We allow an agent to have a preference over streams, even if he himself is not a member of the stream. This is often the case in real life: One may care who enters the government and still not have any desire (nor hope) to become a member thereof.

In this paper, we distinguish among several types of utilities.

(9a) General stream dependence. The utility of an agent is an arbitrary function of the stream of members that occurred:

$$u_i(\mathscr{F}) = u_i(F^0, F^1, \ldots, F^k). \tag{2.1}$$

This is the most general type of utilities that we allow in this paper. Other utilities we consider are special cases thereof.

(9b) Additivity across stages. In this variant we assume that at each stage, utilities for agent i are given by *utility-per-stage* functions $v_i(t, F^t)$, where t is the stage number and F^t is the set of members in the society at that stage. These utilities add up to yield the utility of the stream, so that

$$u_i(\mathscr{F}) = \sum_{t=1}^{k} v_i(t, F^t). \tag{2.2}$$

For the next type of utilities that we consider, we introduce the following definitions:

[12] Sometimes a protocol specifies that only outcomes of votes are known but the ballots are secret. In such cases full histories are equal to the stream histories in the single-period protocols, but not in the 2-period protocol. We do not discuss these cases here as they may involve complications when a member can guess who voted what, knowing his own vote. However, most of the results of this paper cover also the cases of secret ballots.

[13] We do not want to complicate the model by allowing preferences that depend on who voted for the admission or for the expulsion of whom.

Definition 2.1. An agent j is called a *friend* of agent i, in the case where additivity across stages prevails, if for every stage t and every set of members F not containing j,

$$v_i(t, F) \leq v_i(t, F \cup \{j\}). \tag{2.3}$$

Agent j is called an *enemy* of agent i, if for every stage t and every set of members F not containing j,

$$v_i(t, F) \geq v_i(t, F \cup \{j\}). \tag{2.4}$$

Note that (2.3) and (2.4) are required to hold even if agent i is not a member of the society.

Agent j is called *neutral* for agent i, if for every stage t and every set of members F not containing j,

$$v_i(t, F) = v_i(t, F \cup \{j\}). \tag{2.5}$$

Definition 2.2. Utilities are called *separable* if, for every pair of agents i and j, either j is a friend of i or j is an enemy of i.

We denote by $\mathrm{fr}(i)$ [$\mathrm{en}(i)$] the set of friends [enemies] of agent i.

(9c) Separable utilities and additivity across stages. In this variant, the utilities are separable and additivity across stages prevails.

(9d) Weighted agents. In this variant, every agent derives a fixed utility A_i per stage (usually a negative number, large in absolute value) when he is not a member of the society, regardless of who is in the society. When he is in the society, he assigns every other member *who serves with him in the society a weight* $w_i(j)$, $j \neq i$. His utility for a stream is given by

$$u_i(\mathscr{F}) = \sum_{t:\ i \notin F^t} A_i + \sum_{t:\ i \in F^t} \sum_{j \in F^t \setminus \{i\}} w_i(j). \tag{2.6}$$

Thus, j is a friend of i if $w_i(j) > 0$ and an enemy of i, if $w_i(j) < 0$.[14]

Observe that in this variant, the utility-per-stage of agent i of being alone in the society is normalized to be zero, and the model expresses the desire of each agent to first and foremost belong to the society. Once in the society, he would like to spend as much time as possible with friends and as little time as possible with enemies.

(9e) Pure friendship and enmity. This is a special case of the previous one, in which raw friendship and enmity are analyzed, without paying attention to the "intensity" of these relationships. At this level, the weight of a friend is 1 and the weight of an enemy is -1, if both are in the society. We treat in this variant only cases in which players are not neutral to each other. This allows us to specify only the list of friends of i if we so wish.

(10) Common knowledge. All the description above, including the utility functions of each agent, is common knowledge.

[14] This does not conflict with the fact that j is neutral to i, if j is not in the society.

3. Some examples of Nash equilibrium profiles

To gain some insight into our model, we analyze in this section several relatively simple examples. These examples reveal some of the intricacies that may arise, and describe some interesting equilibrium profiles.

First, we introduce the following convention, which will be used throughout this paper: *'voting by a member for a set S at a certain stage' means 'voting to admit agents in S, who are not members of the society when the vote takes place, and to expel agents in S who are in the society at the time the vote takes place'*.[15] With this convention we do not have to specify each time for whom the vote is to admit and for whom it is to expel. Of course, such an S contains no candidates in the (E) protocol and contains no members in the (A) protocol.

Similarly, we say that *the outcome* of a vote at a certain stage t is S, if, as a result of the vote, the members of $F^{t-1} \cap S$ were expelled and the members of $C^{t-1} \cap S$ were admitted at that stage. In the (AE) and (EA) protocols we mean by the above that the candidates in S are admitted in the admission period of the stage and members in S are expelled in the expulsion period.[16] Note that in the (AE) and (EA) protocols, S is usually an outcome of a stage strategy profile. The strategies themselves are more involved, as one has to specify in the second period how to move, as a function of what happened in the first period.

The following example shows that a pure-strategy equilibrium profile always exists, except for cases that we call 'degenerate' (see Example 3.2). Often, however, the trivial equilibrium described below is unreasonable.

Example 3.1. The trivial equilibrium. In all protocols and arbitrary number of stages k, if $|F^0| \geq 3$, and in the (A) protocol even if $|F^0| = 2$, every member voting[17] $\varnothing$ at every stage, regardless of the history, is always an equilibrium profile. Indeed, this strategy prevents any single deviator from changing the set of members.

If $|F^0| = 2$, in the (E), (EA), (AE) and (SIM) protocols, each member voting to admit no one and to expel the other member is often an equilibrium profile. However, there may be 'degenerate situations', in which there does not exist a pure-strategy equilibrium profile. These are similar to the following:

Example 3.2. The population is $F^0 = \{1, 2\}$, $C^0 = \varnothing$, $k = 1$ and the protocol is (E). Agent 1 prefers that if he is in the society, agent 2 is also in the society, but if he is not in the society then nobody should be in the society. Agent 2 prefers that if he is in the society, agent 1 should be out[18] and if he is not in the society, agent 1 should be in. Taking the utilities to be 0 or 1, [or other

[15] For example, if $F^0 = \{1, 2, 3\}$ and $C^0 = \{a, b, c\}$, and agent 1 votes $\{3, a\}$, that means that he votes to expel agent 3 and to admit agent a.

[16] Sometimes, when there can be no confusion, we allow for the sake of brevity 'impossible votes'; namely, that non-members vote or that members expel non-members and/or admit members. Of course such votes have no effect on the outcome.

[17] Namely, vetoing both the expulsion and the admission of any agent.

[18] In this example, and elsewhere, being alone in the society could mean, e.g., being in the sole position to make decisions for a company.

numbers that are compatible with the above preferences] we obtain a payoff matrix [similar to]:

	$\varnothing$	$\{1\}$
$\varnothing$	1, 0	0, 1
$\{2\}$	0, 1	1, 0

Clearly such a game does not have a pure-strategy equilibrium profile.

In general, if $|F^0| = 1$, an equilibrium is reached by the single member voting to maximize his utility, given that as soon as there are more members, the profile will continue as in Example 3.1. For example, in all protocols, but (E), if utilities are separable and additivity across stages prevails (Assumption 9c) and the founder has at least two friends, he invites them all in the first stage and from that stage everyone votes $\varnothing$.

Note that degenerate cases can occur also when $|F^0| = 1$, in situations leading to those similar to that of Example 3.2. For example, suppose that in the (AE) protocol, $F^0 = \{1\}$, $C^0 = \{2\}$ and $k = 1$. Suppose, also, that agent 1 would like that agent 2 will be in the society to such an extent that he prefers to invite him even though he himself may be expelled in the next period. If, otherwise, the utilities are as described in Example 3.2, then no pure-strategy equilibrium exists. Indeed, during the A-period member 1 admits candidate 2 in any equilibrium profile and the E-period has no pure-strategy equilibrium continuation.

The rest of the examples in this section deal with the possibility and desirability to expel a common friend, i.e., a friend of every member, and/or to admit a common enemy, i.e., an enemy of every member.

Example 3.3. Expelling a common friend. This example is valid in the (SIM), (AE) and (EA)[19] protocols with pure friendship and enmity (Assumption 9e).

$$F^0 = \{1,2,3\}, \quad C^0 = \{a,b\}, \quad k = 2, \quad A_i = -100, \text{ all } i,$$

$$\text{fr}(1) = \{2,3,a,b\}, \quad \text{fr}(2) = \{1,3,a,b\}, \quad \text{fr}(3) = \{1,2\},$$

$$\text{fr}(a) = \text{fr}(b) = \{3\}.$$

Voter 3 is a friend of each of the other founders, but prevents them from admitting candidates a and b. It behooves 1 and 2 to expel him in the first stage and admit him back, together with their other common friends in the next stage. Thus, an equilibrium profile in the (SIM) protocol, is e.g., one in which in the first stage agents 1 and 2 vote $\{3,a,b\}$ and 3 votes $\varnothing$. All voters vote sincerely in the second stage. Minor modifications are required in the other protocols.

[19] In the (EA) protocol one can work out a similar example even with $k = 1$.

Why would one want to admit a common enemy? One reason is that an enemy might prevent one's expulsion. This is demonstrated in the next example, where we incorporate also an expulsion of a common enemy, to show how a common enemy can replace a more 'expensive' common enemy.

Example 3.4. Admitting a common enemy. This example is valid for the (SIM), (EA) and (AE) protocols.[20] We assume weighted agents (Assumption 9d).

$$F^0 = \{1, 2, 3\}, \quad C^0 = \{a\}, \quad k = 2, \quad A_i = -100, \text{ all } i,$$

$$w_1(2) = -1, \quad w_1(3) = -10, \quad w_1(a) = -1,$$

$$w_2(1) = -1, \quad w_2(3) = -10, \quad w_2(a) = -1,$$

$$w_3(1) = 1, \quad w_3(2) = 1, \quad w_3(a) = 1,$$

$$w_a(1) = 1, \quad w_a(2) = 1, \quad w_a(3) = -1.$$

In the (SIM) protocol, an equilibrium profile is reached if agents 1, 2 vote $\{3, a\}$ and 3 votes $\{a\}$ at the first stage, and in the second stage, every member votes to expel all his enemies among the members at the second stage and to invite all his friends among the remaining candidates. On the equilibrium path, agent 3 prevents 1 and 2 from expelling each other in the first stage. He is replaced by the less expensive enemy a in the first stage, who serves the same purpose for agents 1 and 2 in the second stage.

Although this is a Nash equilibrium, it is far from being convincing! If we take the view that the agents discuss in a pre-play communication which equilibrium profile should be selected, it is unlikely that they will agree to implement the above equilibrium. Indeed, agent 3 may rightfully claim that he should remain in the society for the first stage, threatening otherwise not to help the others recruit agent a. If a is not elected then, although 3 may be expelled, he says, 1 and 2 will ruin each other in the next stage. This example shows that a selection of an equilibrium profile is often needed and its choice is not a trivial matter.

In the (EA) protocol, the following is an equilibrium profile: In the first period of the first stage, 1 and 2 vote to expel 3 and 3 votes to expel nobody.[21] In the second period, if 3 is expelled, and 1 and 2 survive, they vote to admit agent a. In all other cases, the survivors vote to admit nobody. In the second stage, every member votes in the E-period to expel all his enemies from among the survivors of the first E-period and in the A-period he votes to admit all his friends among the remaining candidates at the end of the first stage.

In the (AE) protocol, an equilibrium profile is reached if every founder votes to admit nobody in the first period of the first stage and every member votes to expel 3 in the second period (agent 3's vote does not matter). In the second stage, if 1 and 2 survive and 3 is expelled, they invite agent a in the first period and expel 3 if he and a are still there (off the equilibrium path). In all

[20] In the (AE) protocol one can work out a similar example even with $k = 1$.

[21] Even if 3 is the only survivor, he should not invite a, in order not to be expelled in the second stage.

other cases every member votes in the A-period to admit his friends among the remaining candidates at the end of the first period and every member votes to expel his enemies among the survivors of the first stage.

Thus, on the equilibrium path in the (EA) and the (AE) protocols, founder 3 is also replaced by a, the common enemy of founders 1 and 2.

A similar example can be constructed in which a common enemy is not expelled.

Example 3.5. Keeping a common enemy. This example is valid for the (SIM), (AE), (EA) and (E) protocols, with friendship and enmity (Assumption 9e):

$$F^0 = \{1,2,3\}, \quad C^0 = \varnothing, \quad k = 2, \quad A_i = -100, \text{ all } i,$$

$$\text{fr}(1) = \text{fr}(2) = \varnothing, \quad \text{fr}(3) = \{1,2\}.$$

One equilibrium profile in the (SIM) (AE) and (EA) protocols requires members 1, 2 and 3 to vote $\varnothing$ at stage 1, and every survivor votes to expel his enemies and admit his friends in the second stage, except that in the (AE) protocol, if 3 is expelled in the first stage and 1 and 2 survive (off equilibrium path), 3 is admitted in the A-period of the second stage, again, in order to prevent 1 and 2 from expelling each other.

4. Nash pure-strategy equilibrium outcomes for 1-stage voting schemes

In this section we characterize all equilibrium outcomes and all subgame-perfect equilibrium outcomes that can be achieved in pure strategies for games representing 1-stage voting schemes. We also show that in the (A), (E) and (SIM) protocols, all the founders essentially prefer pure-strategy equilibrium outcomes which are maximal under set inclusion. Finally, we describe some interesting pure-strategy equilibrium profiles for the (A), (E) and (SIM) protocols.

We remind the reader that *'voting S' means voting to expel* $S \cap F^0$ *and to admit* $S \cap C^0$. *Saying that an outcome was S in the first stage means that* $F^1 = (F^0 \backslash S) \cup (S \cap C^0)$. Members of $F^1 \cap F^0$ are *the survivors* of the voting. We also use the phrase: voter i votes S, when we mean that he votes $S \backslash \{i\}$. This terminology is justified by the fact that whatever a member votes concerning himself makes no difference to the outcome.[22]

It turns out that there is a difference between the 1-period protocols (A), (E) and (SIM) on the one hand and the 2-period protocols (AE) and (EA) on the other, because there exist equilibrium profiles in the latter protocols which incorporate punishments. Accordingly, we first consider the (A), (E) and (SIM) protocols.

I. Equilibria with $k = 1$ in the (A), (E) and (SIM) protocols. The characterization of pure-strategy equilibrium outcomes is provided by the following proposition:

[22] This would *not* be the case if resignations were allowed.

Proposition 4.1. *Let Γ be a game representing a 1-stage voting scheme, obeying general (1-stage) stream dependence (Assumption 9a), in any of the (A), (E), or (SIM) protocols, and having at least three founders (at least two founders in the (A) protocol). A set of agents S is an outcome of a pure-strategy equilibrium profile, if and only if S has the property that no founder i strictly prefers that $S\backslash\{i\}$ is replaced by a proper subset thereof.*[23]

Proof: Let S be an outcome that results from a pure-strategy equilibrium profile. Then, no founder i prefers that $S\backslash\{i\}$ be replaced by a smaller set. Indeed, had he preferred a proper subset of $S\backslash\{i\}$, he could have realized it by simply, alone, voting for this subset, contrary to the fact that S is an equilibrium outcome.

Conversely, if S has the property stated in the proposition, then common voting for S is an equilibrium profile. Indeed, an expelled founder and an admitted candidate can do nothing alone to change their own status regarding membership in the society and no candidate can alone change the elected set S. Any member i, on his own, can only decrease $S\backslash\{i\}$, which he does not prefer. ■

In the next proposition we prove that the founders essentially prefer pure-strategy equilibrium outcomes which are maximal under set inclusion.

Proposition 4.2. *If S and T are two pure-strategy equilibrium outcomes for a game that satisfies the conditions of Proposition 4.1 (or its attached footnote) and $S \subset T$, then every founder that survives under both outcomes, and every founder that does not survive*[24] *under both of the two outcomes, prefers T to S, or is indifferent between the two outcomes. Every founder i that survives S, but does not survive T, still weakly prefers T to $S \cup \{i\}$.*

Proof: Indeed, a survivor of both S and T or a nonsurvivor of both S and T, who prefers S to T, could have achieved S by deviating alone and voting for it, contrary to the fact that T is a pure-strategy equilibrium outcome. Similarly, a survivor i of S, but not of T, could have achieved $S \cup \{i\}$ by voting S, had he preferred it to T. ■

We note, however, that Pareto undominated pure strategy equilibrium outcomes for the (SIM) protocol with respect to the set of *all players* cannot be characterized by set inclusion, as can be seen from the next example.

Example 4.3. The protocol is (SIM) with weighted agents (Assumption 9d). Let $C^0 = \{a\}$, $|F^0| \geq 3$, all members strictly like one another, even if they are not in the society, and all members, except for member i, strongly prefer candidate a over member i. Candidate a is an enemy of i but if i is outside of the society he is indifferent whether a is in the society or not. Being outside the society is the worst possible outcome for every player.

---

[23] If $|F^0| = 1$, then S is a pure-strategy equilibrium outcome, if and only if the single founder votes to maximize his utility. If $|F^0| = 2$, we have to take into consideration that either founder can expel the other founder. Accordingly, S is a pure-strategy equilibrium outcome in the (E) and (SIM) protocols, iff no founder i can gain by either adding the other founder j to S, or removing a non-empty subset of $S\backslash\{i\}$ from S, or by doing both simultaneously.

[24] This cannot happen in the (A) protocol.

By Example 3.1, $S = \emptyset$ is a pure-strategy equilibrium outcome. Similarly, $\{i, a\}$ is also a pure-strategy equilibrium outcome, because all members but i strongly prefer a over i. These members expel i in order to admit candidate a in equilibrium, and member i, since he is expelled, is indifferent whether a enters the society or not. Finally, neither $\emptyset \neq S_1 \subseteq F^0$ nor $S_2 \cup \{a\}$, where $S_2 \subseteq F^0$ contains a member other than i, are equilibrium outcomes. Thus, in the above example, $\emptyset$ and $\{i, a\}$ are the only pure-strategy equilibrium outcomes. However, neither is Pareto dominated by the other since, e.g., candidate a strictly prefers $\{i, a\}$ to $\emptyset$ and member i strictly prefers $\emptyset$.

Examples can be constructed which demonstrate that also in the (A) and (E) protocols, Pareto undominated equilibrium outcomes with respect to the set of all players cannot be characterized by set inclusion.

II. Equilibria for $k = 1$ in the (EA) protocol. In the (EA) protocol the voters can 'punish' deviators even in a 1-stage game. Thus, a characterization of an equilibrium outcome is somewhat more complicated.

Proposition 4.4. *Let Γ be a game representing a 1-stage voting scheme in the (EA) protocol, having at least three founders.*[25] *A set $S := S_E \cup S_A$, $S_E \neq F^0$, $S_E \subset F^0, S_A \subseteq C^0$, is a pure-strategy equilibrium outcome if and only if:*

(1) *S_A is a pure-strategy equilibrium outcome for the survivors $F^0 \backslash S_E$, as described in the previous subsection.*
(2) *For every i in $F^0 \backslash S_E$ (i.e., a survivor of the (E) period), and every proper subset X of S_E, there exists a subset Y of candidates (which may depend on i and X), such that for every subset Z of Y, agent i does not prefer $X \cup Z$ to S.*
(3) *For every founder i not in $F^0 \backslash S_E$ (i.e., a non-survivor of the (E)-period), and every proper subset X of S_E that contains i, there exists a set Y of candidates*[26] *such that agent i does not prefer $X \cup Y$ to S.*

Proof: Let $S := S_E \cup S_A$ be an outcome of a pure-strategy equilibrium profile σ. After the E-period, the survivors face an (A) protocol. Therefore, S_A must satisfy (1), as demonstrated in the previous subsection.

Any deviation in the E-period, caused by voter i, would result in a set X which is a proper subset of S_E. Let Y be the set of candidates voted for by the remaining founders at the A-period, using σ; any vote by i either diminishes Y or keeps it unchanged. Agent i can force any set Z, $Z \subseteq Y$. Since σ was an equilibrium profile, i does not prefer the outcome $X \cup Z$, which is condition (2).

Any deviation in the E-period by a non-survivor i results with a set X, containing i and properly contained in S_E. Let Y be the set chosen in the (A) period, then i cannot force a different set and the outcome is $X \cup Y$. This i does not prefer, because σ is an equilibrium profile, so condition (3) is also satisfied.

[25] Similar statements can be made if there are fewer founders, using reasoning similar to those of the previous subsection. We omit the details. Note that voting to expel every founder is an additional pure-strategy equilibrium profile, unless an expelled founder still prefers that some other founders remain in the society.
[26] We remind the reader that readmitting an expelled member at the same stage is prohibited.

Conversely, let S be a set satisfying the conditions of the proposition. Consider the profile where all founders vote for S_E and all survivors vote S_A, except when the following deviation by a single founder occurs: If he deviates (whether he is a survivor, or not), and a set X results, $X \subsetneq S_E$, all new survivors but the deviator, vote Y in the second stage, as specified by (2), or (3) and we do not care what the deviator does. This is an equilibrium profile. Indeed, by (1), no deviation by a single founder is profitable when it starts in the second stage. Consider a deviation by a survivor i of S_E that results with a set X in the first period, different from S_E. Since $X \subsetneq S_E$, i is a survivor also after the deviation and there must be other survivors of X. Thus, all i can achieve in the second period is a subset Z of Y, which he does not prefer, by (2). Let i be a non-survivor of S_E. If he deviates and a set X results, which is different from S_E, then he is still not a survivor and there must be other survivors, because $S_E \subsetneq F^0$. So, Y results in the second period, which member i does not prefer, by (3). ■

If we are interested in pure-strategy subgame-perfect outcomes, we must also specify moves when more than one voter deviates and impose further conditions. This is summed up in the following corollary.

Corollary 4.5. *Under the conditions of the previous proposition, S is a pure-strategy subgame-perfect outcome for the (EA) protocol if, in addition to (1)–(3), each one of the sets Y, mentioned in the proposition, satisfies the conditions of Proposition 4.1.*[27]

Proof: Indeed, since Y satisfies the conditions of Proposition 4.1, the strategies for the subgame derived after a deviation by a single agent, as described in the second part of the proof of Proposition 4.4, form an equilibrium profile for the second period subgame. For subgames derived after a deviation of two, or more voters, the remaining voters, if any, are instructed to implement the trivial equilibrium profile (Example 3.1). ■

III. Equilibria for $k = 1$ in the (AE) protocol. In the (AE) protocol, an outcome $S := S_A \cup S_E$ means admitting S_A at the first period and expelling the set of *founders* S_E in the second period. We call the members of $F^0 \cup S_A$ the *interim voters*.

In the (AE) protocol the voters can 'punish' a first-period deviator not only by preventing the expulsion of members that he would like expelled in the second period, but also by actually expelling him. As a result, if the utility of not belonging to the society is sufficiently small, every possible admission in the first period can be enforced by threatening to expel a deviator in the second period. Thus, in this case, Condition (1) in Proposition 4.6 below is both necessary and sufficient for a pure-strategy equilibrium. Otherwise, condition (2) is also needed.

Proposition 4.6. *Let Γ be a game representing a 1-stage voting scheme in the*

[27] No degenerate situation can occur in the (A)-period, but some subgames may have a single survivor. He should maximize his utility in the A-period.

(AE) protocol, having at least three founders.[28] *A set* $S := S_A \cup S_E$ *is a pure-strategy equilibrium outcome, if and only if*

(1) S_E *is an equilibrium outcome for the interim voters* $F^0 \cup S_A$*, who are allowed to expel only members of* F^0*.*[29]
(2) *For every founder i and every proper subset X of* S_A*, there exists a subset Y of founders (which may depend on i and X), such that for every subset Z of* $Y \backslash \{i\}$*, agent i does not prefer* $X \cup Z \cup (Y \cap \{i\})$ *to S.*

Proof: Let $S := S_A \cup S_E$ be the outcome of an equilibrium profile σ. Then, S_E must satisfy (1), to protect from a profitable single-person deviation in the E-period. A deviation by founder i in the admission period would lead to the admission of a subset X, $X \subsetneq S_A$. Thereafter, σ, implemented by the other interim voters, would result in the expulsion[30] of the set of founders Y. All agent i can do is to replace $Y \backslash \{i\}$ with a proper subset Z, at the second stage, which he does not prefer, because σ is an equilibrium profile. Thus condition (2) is also satisfied.

Conversely, let S be a set satisfying the conditions of the proposition. Consider the profile where all founders vote for S_A and all interim voters vote S_E, except under the following deviation: If a single founder deviates and a set X results, $X \subsetneq S_A$, all new survivors but the deviator, vote Y in the second stage.[31] This is an equilibrium profile. Indeed, by (1), no deviation by a single founder is profitable when there is no change in the first stage. Consider a deviation by a single member i of F^0 that results with a set X being voted in the A-period, different from S_A. Then, all other interim voters vote Y and agent i can only prevent expulsion of members from $Y \backslash \{i\}$. This he does not prefer, by (2). ∎

Corollary 4.7. *Under the conditions of the previous proposition, S is a pure-strategy subgame-perfect outcome for the (AE) protocol, if, in addition to (1)–(2), each one of the sets Y mentioned in Proposition 4.6, satisfies the conditions of Proposition 4.1.*

The proof is similar to that of Corollary 4.5 and will be omitted. Note that in some voting schemes, at some subgames, degeneracy similar to that of Example 3.2 may occur. In such cases no pure-strategy subgame-perfect equilibrium profile exists.

There is a kind of duality between the unanimity rule of admission or expulsion and the quota-1 rule treated in Barberà et al. [1]. Indeed, in the unanimity case, a single voter i has *veto power*: He can veto the admission of any candidate and he can veto the expulsion of any voter. Thus, the unanimity rule can be viewed as a 1-quota rule for vetoing. This observation allows us to

[28] Similar statements can be made if there are fewer founders, using reasoning similar to those in the footnotes of subsection I. We omit the details.
[29] This restriction is required in (5d) of Section 2. Necessary and sufficient conditions for this to happen are similar to those provided in Proposition 4.1 and its footnotes (where no restriction is imposed on the members that can be expelled). They will be omitted.
[30] Agent i is expelled if $Y \cap \{i\} \neq \emptyset$.
[31] Thus, they vote to expel $Y \backslash \{i\}$ and expel i, if $i \in Y$.

extend results obtained in the 1-quota rule to our case.[32] One such result is presented below and will also be found in Section 6.

Consider a game Γ, representing a 1-stage voting scheme in the (A), (E), or (SIM) protocol with $|F^0| \geq 3$ [$|F^0| \geq 2$ in the (A) protocol]. Let S have the property that, if it is an outcome allowed by the protocol, no founder i would prefer to replace $S \backslash \{i\}$ with a proper subset thereof. Only sets having this property are candidates for pure-strategy equilibrium outcomes in a 1-stage voting scheme (Proposition 4.1) and we shall refer to them as *relevant sets*. Denote $S^c := N \backslash S$. In the proof of Proposition 4.1, we showed that an equilibrium profile is achieved if each voter *vetoes* exactly S^c, that is, he votes S. But there are other interesting equilibrium profiles. For example, the strategy of *sincere voting*[33] in the separable case (Condition 9c). This is a special case of the following:

Proposition 4.8. *Let S be a relevant set in a game Γ, representing a 1-stage voting scheme in the (A), (E), or (SIM) protocols. Let $P_i \subseteq S^c$ be a best response set to veto*[34] *for member i, if the other members veto the set $S^c \backslash P_i$. Let $C := S^c \backslash \bigcup_{j \in F^0} P_j$ and let $V_i = P_i \cup C$, all $i \in F^0$. Then, each founder i in F^0 vetoing V_i is an equilibrium profile for Γ.*

The proof requires two lemmas:[35]

Lemma 4.9. *Suppose that vetoing the set P_i is a best response of founder i when all other founders veto $S^c \backslash P_i$, where S^c is an arbitrary given set of agents containing P_i. If $Q \subseteq S^c \backslash P_i$, then vetoing $P_i \cup Q$ is also a best response of i, when all others veto $S^c \backslash P_i$.*

Proof: Q is covered anyhow by $S^c \backslash P_i$, so it makes no difference whether i includes Q in his veto set. ■

Lemma 4.10. *Let P_i be a best response set to veto by founder i, when all the other founders veto together $S^c \backslash P_i$, where S^c is an arbitrary set of candidates containing P_i. If $R \subseteq P_i$ then vetoing the set $P_i \backslash R$ is a best response of i when all others veto $(S^c \backslash P_i) \cup R$.*

Proof: Vetoing the set $P_i \backslash R$, when all other founders veto $(S^c \backslash P_i) \cup R$, would yield player i the utility gained when S^c is being vetoed. If vetoing another set, Q, would yield him a higher utility, then vetoing $Q \cup R$, rather than P_i, would be a better response when the others veto $S^c \backslash P_i$, because $(Q \cup R) \cup (S^c \backslash P_i) = Q \cup (R \cup (S^c \backslash P_i))$. ■

---

[32] In Barberà et al. we considered only 1-quota for admission. So, strictly speaking, duality occurs only for the (A) protocol. For this reason we are forced to supply proofs, whenever these ideas are applied to other protocols as well. Note that this duality can be applied only for 1-stage schemes, and only in the (E), (A) and (SIM) protocols. Duality breaks down when several stages or periods are considered. Indeed, if one admits a person in the 1-quota case, the admitted agent becomes a voter as of the next stage, whereas if one prevents an agent from being admitted, this agent does not become a voter in the next stage, or period.

[33] Namely, voting to admit all of one's friends and expel all of one's enemies, as the protocol allows.

[34] $P_i = \varnothing$ is always such a best response set.

[35] Taken from Barberà et al. We reproduce the short proofs for completeness sake and in order to save the reader from producing the necessary alterations.

Proof of Proposition 4.8: Vetoing the set P_i is a best response of i against the others vetoing $S^c \setminus P_i$; therefore, vetoing V_i is a best response of i against all others vetoing $S^c \setminus P_i$ (Lemma 4.9). Now, some of the various P_j's may intersect P_i, so that these other players' votes may cover more that $S^c \setminus P_i$. We invoke Lemma 4.10 to deduce that vetoing $V_i \setminus \bigcup_{j \in F^0 \setminus \{i\}} P_j$ is a best response of i against the others vetoing $\bigcup_{j \in F^0 \setminus \{i\}} V_j$. Invoking Lemma 4.9 once more, we find that vetoing V_i is a best response of i against all others vetoing $\bigcup_{j \in F^0 \setminus \{i\}} V_j$. ∎

5. Subgame-perfect equilibrium outcomes for a multi-stage voting scheme

In this section we characterize all pure-strategy subgame-perfect equilibrium outcomes for multi-stage voting schemes for each of the protocols discussed in this paper. We then show that, in the (A) protocol, every Nash equilibrium outcome that can be realized by pure-strategy profiles, can also be realized by a subgame-perfect pure-strategy profile. We then explain why this is not the case in the other protocols.

Even though games representing multi-stage voting schemes are not games of perfect information because votes are taken simultaneously, it was proved in Barberà et al. [1] that a kind of backward induction is possible in order to achieve all multi-stage subgame-perfect equilibrium outcomes. To explain this, we first describe the collation process.

The collation process.

Consider a multi-period voting scheme, represented as a game in extensive form. Denote by k the number of stages. The protocol can be any of those discussed in Section 2. Choose an arbitrary pure, or mixed strategy for the last subgames.[36] Having done that, one can calculate the expected payoff vector for each of these subgames, had the path from the root reached, in fact, that subgame.

Truncate now these subgames from the extensive form, placing the calculated expected payoffs at the corresponding payoff endpoints. The resulting game has $k - 1$ stages, or is a k-stage game with only one period left at its end stages. Repeat this procedure, choosing each time a pure or a mixed strategy for the last subgame and then truncating it, placing the (expected) payments at the endpoints. Eventually, one reaches a single-stage game, or a single-period game in extensive form. Any strategy chosen now yields the same expected payoff vector that would have resulted, had the agents played the original game, using the chosen strategies on each 1-stage or 1-period end subgame.

The following was proved in Barberà et al. [1]:[37]

Theorem 5.1.[38] *Let Γ be a game representing a multi-stage voting scheme with utilities obeying general stream dependence (Assumption 9a) and obeying any*

[36] These are stage subgames in the (A), (E), and (SIM) protocols and the last period of the last stage subgames in the (AE) and (EA) protocols.
[37] The model was different, but the proof of the following theorem remains the same.
[38] This theorem is valid also for protocols not discussed in this paper; e.g., having different rules for admission.

of the protocols considered in this paper. If, during collation, we always choose a Nash equilibrium profile for each 1-stage subgame,[39] *the resulting profile is a subgame-perfect equilibrium profile for Γ. Conversely, every subgame-perfect equilibrium profile can result in this fashion.*

Thus, by applying collation, we can obtain all subgame-perfect equilibrium outcomes for an arbitrary multistage voting scheme. In particular, Proposition 4.2 can now be used to narrow down the search for subgame-perfect outcomes that are "Pareto undominated" from the point of view, e.g., of the agents who survive throughout the play.

Corollary 5.2. *The following collation process yields all possible pure-strategy subgame-perfect equilibrium outcomes in each protocol: Starting with the last stage and continuing backwards, we select a set of votes, permitted by the protocols, that are candidates for pure-strategy subgame-perfect equilibrium outcomes, as described in Propositions 4.1 and Corollaries 4.5 and 4.7. Specify pure-strategy subgame-perfect equilibrium profiles that give rise to these outcomes and then perform truncations. (In the (AE) and (EA) protocols we have combined both periods). We continue in this fashion for all stages. If we encounter a 'degenerate situation' (see Example 3.2), we discard the choices that generated this case.*

In the (A) protocol all multi-stage Nash equilibrium outcomes can always be realized by subgame-perfect equilibrium profiles, as the following theorem shows:

Theorem 5.3. *Let σ be a pure-strategy Nash equilibrium profile for a voting game Γ in the (A) protocol. There exists a pure-strategy subgame-perfect equilibrium profile $\hat{\sigma}$ of the same protocol, leading to the same stream of members, and therefore to the same payoff to the agents.*

Proof: We define $\hat{\sigma}$ to be equal to σ on the equilibrium path and off the equilibrium path every voter votes in accordance with the trivial equilibrium described in Example 3.1; namely, voting for $\emptyset$ as soon as there are at least two voters.

To see why this is an equilibrium, observe that if a voter could benefit by a deviation from $\hat{\sigma}$, by preventing the admission of a few candidates, he could have performed the same deviation in σ, and then vote $\emptyset$ until the end of the play. This is contrary to the fact that σ is a Nash equilibrium profile. By Example 3.1, $\hat{\sigma}$ is subgame-perfect.[40] ■

In all other protocols, there may exist pure-strategy Nash equilibrium outcomes that cannot be realized as subgame-perfect pure-strategy profiles, due to the fact that, e.g., a Nash equilibrium profile may incorporate a punishment – the expulsion of a deviator, even though a voter may want to keep him in the society, which makes this equilibrium profile not subgame-perfect.

---

[39] We can use pure, or mixed strategies. We restrict ourselves to behavioral strategies, which is permissible, because the game is of perfect recall (See Kuhn [3], or Selten [6]).

[40] As long as there is a single voter, there is only one subgame at each such node and therefore there is no distinction between the equilibrium requirement and the subgame-perfect requirement.

6. Pure-strategy perfect equilibria in single-stage single-period voting schemes

Nash equilibrium outcomes are sometimes criticized by the claim that off the equilibrium path, agents may threaten to act in a non-credible way. For this reason, subgame-perfect profiles are often preferred, albeit, sometimes at the expense of being dominated by Pareto superior Nash equilibria. Even subgame-perfect equilibria often entail non-credible behavior, as was shown in Selten [6], so one resorts to (trembling-hand) perfect equilibria.

In this section we explore criteria for the existence of pure-strategy perfect equilibrium profiles for the single-stage single-period voting schemes (A), (E) and (SIM), both when utilities are separable and when they are not. In Section 8 we study the existence of such perfect equilibrium profiles for multi-stage single-period schemes as well as the single-stage (AE) and (EA) schemes.

If no restrictions are placed on the utilities, other than general stream dependence (Assumption 9a), pure-strategy perfect equilibrium profiles may not exist. In Proposition 6.5 and Remark 6.6 below we provide a sufficient condition for the existence of such profiles for the (SIM), (A) and (E) protocols. We later use this result to prove that if utilities satisfy the weighted agents condition (Assumption 9d), then there always exists a pure-strategy perfect equilibrium profile in essentially every 1-stage generic[41] voting scheme in the (EA) protocol. This is in contrast to the (AE) protocol, wherein such a perfect equilibrium profile may not exist (Example 8.8).

When utilities are separable, there exists a natural pure-strategy perfect equilibrium profile for the (A), (E) and (SIM) protocols.

Proposition 6.1. *If utilities are separable, sincere voting constitutes a perfect equilibrium profile in the single-stage (A), (E) and (SIM) protocols, and it is essentially unique.*

Proof: Voting to expel all of one's enemies, if the protocol permits, and admit all of one's friends, if the protocol permits (voting at will for members or candidates towards which one is neutral), is always a best response against any trembles of all other agents. ■

In Section 3 we saw that there are degenerate cases where even pure-strategy Nash equilibrium profiles do not exist in the (E) protocol. One can easily construct examples for all protocols with non-separable utilities, in which pure-strategy Nash equilibrium profiles exist, but none of them is perfect.[42]

In order to present our sufficient condition for the existence of a pure-strategy perfect equilibrium profile where utilities are not separable, we need to introduce two definitions and a lemma.

Definition 6.2. A voting scheme is called *generic* if, for each agent, the utilities for different streams are different.

Definition 6.3. For a set $S \subseteq N$ and agents $i \in F^0$ and $x, x \in N \backslash S$, we say that i

[41] See Definition 6.2.

[42] One example can be constructed e.g., from Example 8.8 by considering the single period that results after the moves made at the second period.

supports x with respect to S if i prefers the outcome $S \cup \{x\}$ to S; namely, he would like to expel a founder x or admit a candidate x, as the case may be.[43]

Also, we need the following lemma which is proved by induction.

Lemma 6.4. *For all $n \geq 1$ and for all ε, $0 \leq \varepsilon \leq 1$, $1 - (1 - \varepsilon)^n \leq n\varepsilon$.*

Denote by $V_i(S)$, for $i \in F^0$ and $S \subseteq N$, the union of S and all agents supported by i with respect to S.

Proposition 6.5. *Let Γ be a generic 1-stage voting game with at least three founders. The protocol is (SIM). If S is a set of agents satisfying*

(1) $\{V_j(S)\}_{j \in F^0}$ *is an equilibrium profile,*
(2) $\bigcap_{i \in F^0} (V_i(S) \cup \{i\}) = S$,

then $\{V_i(S)\}_{i \in F^0}$ is a perfect equilibrium profile of Γ.

Proof:[44] Note that by (1), all founders want to expel all members of $S \cap F^0$ and all of them want to admit $S \cap C^0$ since otherwise, a founder could beneficially have vetoed the outcome. Denote $c = |C^0|$, $f = |F^0|$, $n = |N|$. Denote by d the minimum payoff difference for any two different sets of agents and any founder. Similarly, denote by M the maximal payoff difference for any two different sets of agents and any founder; i.e.,

$$d = \min_{\substack{i \in F^0, T_1 \neq T_2 \\ T_1, T_2 \subseteq N \setminus \{i\}}} |u_i(T_1) - u_i(T_2)|, \quad M = \max_{\substack{i \in F^0, T_1 \neq T_2 \\ T_1, T_2 \subseteq N \setminus \{i\}}} |u_i(T_1) - u_i(T_2)|. \quad (6.1)$$

The voting scheme is generic and $n \geq 3$; therefore, $d > 0$ and $M > 0$.

Assume fixed positive ε_1 and ε_2. Assume initially that they are each less than $\frac{1}{4n}$ 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 $x \in N \setminus (V_i(S) \cup \{i\})$, vote for $V_i(S) \cup \{x\}$ with probability ε_1.
(2) For any other set of agents, except $V_i(S)$, vote for this set with probability $\frac{\varepsilon_2}{2^n}$.
(3) Vote for $V_i(S)$ with the residual probability. This probability is greater than $1 - n\varepsilon_1 - \varepsilon_2$ and from the restrictions already imposed on the epsilons it is greater than $\frac{1}{2}$.

[43] Note that this term is relative to S. An agent i may support x relative to S and not support him relative to another set T. Moreover, it may well be the case that an agent i supports both agents x and y, but if he is given the choice of either S, or $S \cup \{x, y\}$ he will prefer S. Thus, this is a technical term and one should not attribute any intrinsic personal attitude of i towards x. Of course, if utilities are separable, i will always support candidate friends and enemy members (Proposition 6.1).
[44] The technique employed in this proof is similar to the one used in Barberà et al. [1]. However, this theorem holds for different and more general cases.

As ε_1 and ε_2 tend to zero, this completely mixed strategy tends to $V_i(S)$ for every founder i.

Let i be an arbitrary fixed founder. The proof will conclude if we show that $V_i(S)$ is his best reply against the others using these mixed strategies, provided the epsilons are small enough. Note that whether i survives the votes, or not, does not depend on i's own vote. His fate depends solely on the votes of the other players.

Consider two possible types of deviation by agent i. The first is a deviation that makes a difference against $V_{-i}(S)$, and the second is a deviation that makes no difference against $V_{-i}(S)$.

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

$$1-(1-n\varepsilon_1-\varepsilon_2)^{f-1} \leq (f-1)(n\varepsilon_1+\varepsilon_2) \leq nf(\varepsilon_1+\varepsilon_2), \tag{6.2}$$

as at least one founder $j \neq i$ must vote for a set different from $V_j(S)$. The first inequality is implied by Lemma 6.4.

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

$$\frac{d}{2^{f-1}} \geq Mnf(\varepsilon_1+\varepsilon_2), \tag{6.3}$$

and this always holds if $\varepsilon_1 < \frac{d}{Mnf2^f}$ as $\varepsilon_2 \leq \varepsilon_1$.

We now investigate the other type of deviation, and find restrictions on the epsilons to ensure that it will not be profitable either. Consider a deviation by agent i to $(V_i(S)\backslash R) \cup A$, where $R \cap S = \emptyset$, $R \subseteq V_i(S)$, $A \cap V_i(S) = \emptyset$, and $A \cap (\bigcap_{j \neq i} V_j(S)) = \emptyset$. Thus, player i removes members of R from his vote and adds members of A, and none of these agents is in the intersection of all the other founders' $V_j(S)$'s. Thus, such a deviation by player i makes no difference against $V_{-i}(S)$. This captures all deviations with the required properties. Denote $Q := A \cup R \neq \emptyset$ and $V_i' := (V_i(S)\backslash R) \cup A$.

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

The first case is when the intersection of the votes of the other founders and $S \cup Q$ is $S \cup \{x\}$ 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 $V_i(S)$. For each $x \in Q$ denote the probability of this subcase by $\eta_1(x)$.

The second case (possibility of gain) is when the intersection of the votes of all the other founders and $S \cup Q$ contains x for some $x \in Q$, but is not equal to $S \cup \{x\}$. 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 included in the intersection of the votes of the other founders.

If the intersection of the votes of the other founders contains no agents in Q, then i gets the same payoff from both $V_i(S)$ 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

$$d\sum_{x\in Q}\eta_1(x) > M\sum_{x\in Q}\eta_2(x). \tag{6.4}$$

For $x \in Q$, let $m(x)$ be the number of founders that do not support x with respect to S, not including agent i. For all $x \in Q$ it is true that $m(x) \geq 1$. Indeed, otherwise, if all founders other than i support x, then x's fate depends on i's vote, so i's deviation V_i' makes a difference against $V_{-i}(S)$ – a case dealt with earlier.

The following bounds hold for all $x \in Q$, as we explain:

$$\eta_1(x) \geq \varepsilon_1^{m(x)}(1 - n\varepsilon_1 - \varepsilon_2)^{f-m(x)-1} \geq \frac{\varepsilon_1^{m(x)}}{2^{f-m(x)-1}} \geq \frac{\varepsilon_1^{m(x)}}{2^f}. \tag{6.5}$$

The first inequality holds, as the event $\eta_1(x)$ includes the event that each founder that does not support x with respect to S votes $V_j(S) \cup \{x\}$ and all others, excluding i vote $V_j(S)$. The second inequality is implied by $1 - n\varepsilon_1 - \varepsilon_2 > \frac{1}{2}$. We further have:

$$\begin{aligned}\eta_2(x) &\leq 1 - (1-\varepsilon_2)^{f-1} + \varepsilon_1^{m(x)}(1 - (1 - n\varepsilon_1 - \varepsilon_2)^{f-m(x)-1})\\ &\leq (f-1)\varepsilon_2 + \varepsilon_1^{m(x)}(f - m(x) - 1)(n\varepsilon_1 + \varepsilon_2).\end{aligned} \tag{6.6}$$

The first inequality holds, as for this case to occur, at least one of the events [at least one founder $j, j \neq i$ votes for neither $V_j(S)$ nor $V_j \cup \{y\}$ for any agent y] which has probability no greater than $1 - (1-\varepsilon_2)^{f-1}$, or [all the founders j that do not support x with respect to S vote for $V_j \cup \{x\}$ and at least one of the other founders j' votes for a set different from $V_{j'}(S)$] must occur.[45] The second inequality holds from two applications of Lemma 6.4.

If we now assume that $\varepsilon_2 \leq \dfrac{\varepsilon_1^{m(x)+1}}{f-1}$ then (6.6) implies

$$\eta_2(x) \leq \varepsilon_1^{m(x)+1} + \varepsilon_1^{m(x)}(f-1)(n+1)\varepsilon_1 \leq 2nf\varepsilon_1^{m(x)+1}. \tag{6.7}$$

[45] To prove that the first inequality is indeed correct, let us prove that the complement of the union of these two events imply that $S \cup Q$ is not of the required form. That is, either x is not in the outcome or the outcome is precisely $S \cup \{x\}$. So, one can verify that the complement of the union of the above two events is the intersection of the following two events, the second of which is a union of two subevents: [All members j either vote $V_j(S)$ or $V_j(S) \cup \{y\}$] and [either there is a member who does not support x who votes for a set different than $V_j \cup \{x\}$ or every member who supports x votes for $V_j(S)$.] Now, if the first event and the first subevent of the second event hold then x is not in the outcome, while if the first event and the second subevent of the second event hold then the outcome either does not contain x, which occurs when one member j who does not support x either voted $V_j(S)$ or $V_j(S) \cup \{y\}$ for $y \neq x$, or is precisely $S \cup \{x\}$, when all members j who do not support x voted $V_j(S) \cup \{x\}$.

Inequalities (6.5) and (6.7) together imply

$$\eta_2(x) \le \eta_1(x) \frac{2nf\varepsilon_1^{m(x)+1} 2^f}{\varepsilon_1^{m(x)}} = \eta_1(x) 2^{f+1} nf\varepsilon_1. \tag{6.8}$$

Now, using inequality (6.8), inequality (6.4) is implied by

$$d \sum_{x \in Q} \eta_1(x) > M 2^{f+1} nf\varepsilon_1 \sum_{x \in Q} \eta_1(x). \tag{6.9}$$

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

$$\varepsilon_1 < \frac{d}{Mnf 2^{f+1}}. \tag{6.10}$$

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

$$\varepsilon_1 < \frac{d}{Mnf 2^{f+1}}, \quad \varepsilon_2 \le \frac{\varepsilon_1^f}{f-1}, \tag{6.11}$$

imply that $V_i(S)$ 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 $\{V_i(S)\}_{i \in F^0}$. ■

Remark 6.6: Proposition 6.5 is also valid for the (A) and the (E) protocols.

Proof: The (E) protocol is just the (SIM) protocol with $C^0 = \varnothing$. If the protocol is (A), we can modify the utilities and transform it essentially to a (SIM) protocol by stipulating that for each founder i, the utility of any set of members in which any founder is missing is lower than the utility of any set of agents that contains all the founders, with or without i himself. Then, in any pure-strategy perfect equilibrium profile in the (SIM) protocol, no member is going to vote to expel a founder, even if he is sure that another founder will veto his expulsion, because off the equilibrium path his vote can materialize in some subgames, thereby hurting him in the subgames, contrary to the subgame perfectness requirement. Thus, effectively, the (SIM) game is equivalent to the original (A) game. ■

Henceforth we use the expression 'veto S' to be synonymous with the expression 'vote $N \backslash S$'. To veto a founder means preventing his expulsion. To veto a candidate means preventing his admission. 'Vetoing' rather than 'voting for' makes some results more transparent, as can be seen from the following.

The next proposition is an important application of Proposition 6.5, which sheds some light on the structure of an important class of pure-strategy perfect equilibrium profiles. We shall use it in a nice way in the proof of Theorem 8.5.

Proposition 6.7. *Consider a generic 1-stage game Γ, representing a single-stage voting scheme in the (SIM) [or (E), or (A)] protocol. Suppose that in equi-*

librium it turns out that each member i vetoes a set of agents P_i, not containing himself, and that these sets are disjoint, then Γ has a pure-strategy perfect equilibrium profile consisting of each voter possibly vetoing some additional agents, in addition to P_i.

Proof: Denote by S the set $N\setminus\bigcup_{i\in F^0} P_i$. The vetoed sets P_i form a *generalized partition*[46] of $N\setminus S$. Thus, S is the set of members on which there is a unanimous consent to admit and an almost unanimous consent to expel. Observe that each member of P_i is not supported by i with respect to S. Indeed, if an agent x, $x \in P_i$, were supported by i, then i would have preferred to add x to S and he would have been benefited in doing so, because none of the other voters veto x. This would contradict the fact that the votes $\{N\setminus P_i\}_{i\in F^0}$ constitute an equilibrium profile. We now instruct each voter i to veto, *in addition*, all members of $(N\setminus S)\setminus P_i$ that he does not support with respect to S. This will bring him to the set $V_i(S)$ of Proposition 6.5 [Remark 6.6]. Equilibrium will persist with the outcome S, as any deviation that makes the outcome change is equivalent to a deviation from P_i. Condition (2) of Proposition 6.5 is also satisfied. Thus, the profile $\{N\setminus V_i\}_{i\in F^0}$ constitutes a pure-strategy perfect equilibrium profile. ■

7. The collation theorem for perfect profiles and its converse

In Section 5 we have seen that using a collation procedure, one can derive *all* pure-strategy subgame-perfect equilibrium outcomes from all 1 stage pure-strategy Nash equilibrium profiles. Deriving pure-strategy perfect equilibrium outcomes from the 1-stage/period ones can be done in a similar way. However only a weak version of the converse theorem is valid, as is elaborated below.

Theorem 7.1. *Let Γ be an extensive-form game representing a generic*[47] *multi-stage/period*[48] *voting scheme. Proceeding backward: We select at each end-subgame a pure-strategy perfect equilibrium profile for the subgame and then collate. We repeat this procedure until all end-subgames are covered. The resulting choice will be a pure-strategy perfect equilibrium profile for Γ.*

Remark 7.2: This theorem holds for other protocols, not necessarily those protocols discussed in this paper, such as having different sequences of voting schemes and different rules for admission and expulsion.

Proof: We regard Γ as being composed of 1-stage/period trees $\Gamma_{r,s}$, where r enumerates the stages/periods and s counts them, say from left to right. The endpoints of $\Gamma_{r,s}$ are assigned various payoffs, as determined via the collation, using an appropriate strategy profile, or trembling from it.

Denote by $\sigma_{r,s}$ the pure-strategy perfect equilibrium profile chosen for $\Gamma_{r,s}$ by the collation process. Denote by $(\sigma^n_{r,s})_{n=1}^{\infty}$ the trembling hand sequence of completely mixed profiles associated with $\sigma_{r,s}$. For each n, $\sigma_{i;r,s}$ is then a best

[46] Generalized, because some P_i's may be empty sets.
[47] Definition 6.2.
[48] By 'multi-stage/period' we mean either a multi-stage of a single period protocol (A), (E), or (SIM), or a single-stage, or a multi-stage voting scheme in the (AE) or (EA) protocols.

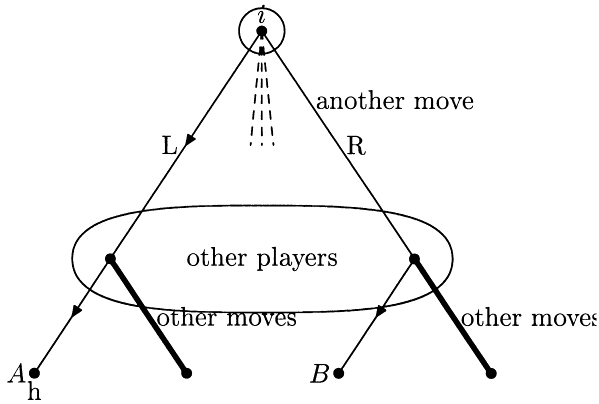

Fig. 1. The tree $\Gamma_{r,s}$

reply to $\sigma^n_{-i;r,s}$, and $\sigma^n_{r,s}$ converge to $\sigma_{r,s}$. We have to show that the collection $\sigma := \{\sigma_{r,s}\}$ is a perfect equilibrium profile for Γ. To this end we construct tremble sequences, basically composed of modifications of the $\sigma^n_{r,s}$.

To prove that a strategy profile is a perfect equilibrium, we have to view Γ as representing the agent normal form game[49] and we have to produce a sequence $(\sigma^n)^\infty_{n=1}$ of behavioral completely mixed strategy profiles, converging to σ, and verify that for each agent i and each n, σ_i is a best reply to σ^n_{-i} (See Myerson [4]).

Consider a particular tree $\Gamma_{r,s}$ and a particular agent i. In Figure 1 he is placed as the top player and the other players are symbolically enclosed in one ellipse. Without loss of generality we also assume that each agent in $\sigma_{r,s}$ is assigned the pure-strategy "left" (L). If the payments to agent i at the endpoints of $\Gamma_{r,s}$ are those that result by collation from the selected pure strategies[50] then, for each n, $\sigma_{i;r,s}$ is a best reply to $\sigma^n_{-i;r,s}$. However, this need not be the case if the payments at the endpoints are generated from trembles at lower levels. If, for each[51] n, agent i *strictly* prefers $\sigma_{i;r,s}$ to any other move against $\sigma^n_{-i;r,s}$, then, by passing to subsequences of $\sigma^n_{\tau,\rho}$, $\tau > r$, for games $\Gamma_{\tau,\rho}$ below $\Gamma_{r,s}$, one can assure that strict preference will remain even if collation is performed on the perturbed payments. Thus, with this choice of subsequences, agent i will continue to prefer the L move; namely, $\sigma_{i;r,s}$.

A more careful analysis is required if there are two (or more) endpoints of $\Gamma_{r,s}$ with equal original payments to agent i, denoted h, which i achieves, given that the other agents play as in σ, and moreover, for almost all n, agent i is indifferent between these moves when he plays in $\Gamma_{r,s}$ and the other agents tremble in this game. In such cases, keeping the original subsequent trembles, or subsequences of them, may alter the payment to agent i at $\Gamma_{r,s}$, so that he may not prefer to choose L any more.

Since the game is generic, equal original payments at two endpoints can

---

[49] I.e., the various information sets are regarded as belonging to different "*agents*" of the players.
[50] We call them *the original payments*.
[51] It is sufficient that this happens for infinitely many n's, as we can pass to a subsequence of the original $\sigma^n_{r,s}$.

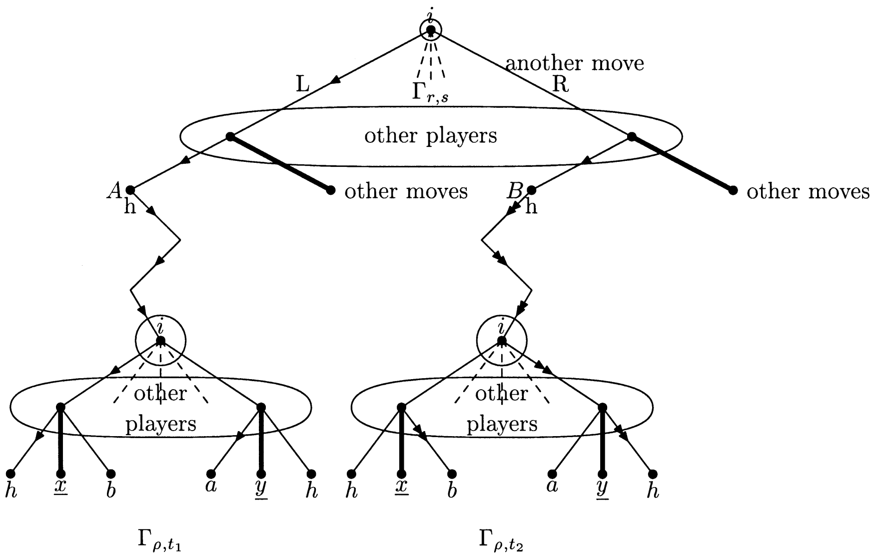

Fig. 2. Two replicas of a subgame

result only if the paths from the root of Γ to these endpoints, denoted A and B, represent *the same stream of members*, with different votes by the previous agents and, moreover, the contributions via the collation process continue to keep the same stream of members (see Figure 2).

Denote by $\Gamma^A_{r,s}$ and $\Gamma^B_{r,s}$ the multi-stage/period subgames starting at A and B, respectively. Then, these games are exact replicas of the same game. However, traveling along them via the collation process may be different and the trembles too may be different. See Figure 2 where the last trees, denoted Γ_{ρ,t_1}, Γ_{ρ,t_2}, $\rho > r$, are singled out, indicating the two different paths by single arrows and double arrows. Again, Γ_{ρ,t_1} and Γ_{ρ,t_2} are two replicas of the last stage/period. Without loss of generality we name the moves that lead to Γ_{ρ,t_2} and continue to the endpoint h via σ as "right" moves, although at some stages they may be identical to the left moves (but performed off the equilibrium path).

The proof will proceed along the following lines: We express the necessary and sufficient conditions that (σ^n_{ρ,t_1}) and (σ^n_{ρ,t_2}) are appropriate test sequences for Γ_{ρ,t_1} and Γ_{ρ,t_2}. These are inequalities (7.1) and (7.2) below. We modify $\sigma^n_{-i;\rho,t_1}$ and $\sigma^n_{-i;\rho,t_2}$, keeping the validity of (7.1) and (7.2) so that either the left side of (7.2) becomes less than or equal to the left side of (7.1), or so that the left side of (7.2) becomes strictly less than h. If one of these cases holds for infinitely many n's, we can pass to a subsequence on n's for which this happens. In the first case that means that the contribution of the perturbation towards (B) is less than, or equal to the contribution towards (A) and we can collate and proceed to higher levels with similar modifications so that eventually, at $\Gamma_{r,s}$, i will weakly prefer the move L, given the perturbations below. In the second case we pass to subsequences of $\sigma_{-i;\rho^*,t^*}$ of stage/period games along the path to (B) so that, when one reaches (B), the expectation of the perturbed payment is less than h. Having achieved this we then take a sub-

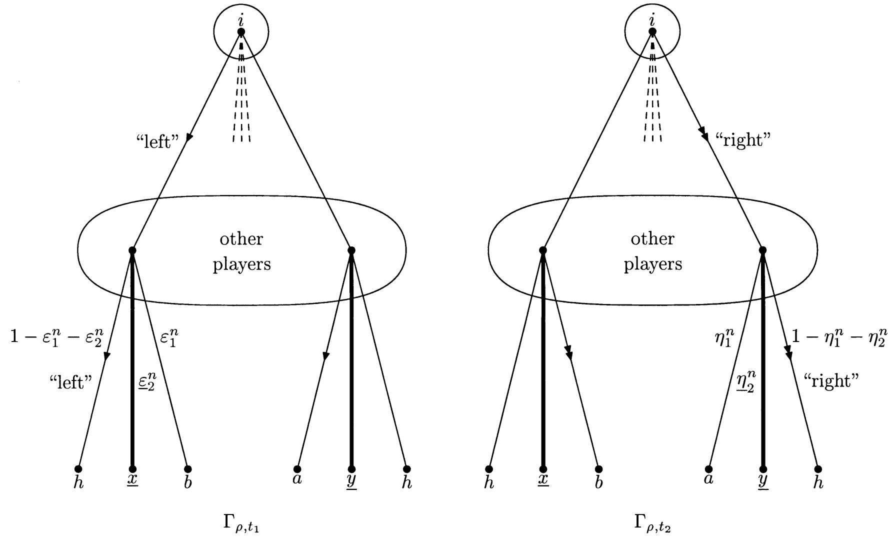

Fig. 3. The endgames in special cases

sequence of $\sigma^n_{-i;\rho^*,t^*}$ along the path to (A) that will reach (A) with an expected payment to i which is greater than the expectation at (B). This is true, because both expected perturbed payments tend to h as $n \to \infty$. In this case too, agent i will prefer the move L at $\Gamma_{r,s}$.

If "right" and "left" are the same moves for agent[52] i at Γ_{ρ,t_1} and Γ_{ρ,t_2}, we change σ^n_{ρ,t_2} and make it equal to σ^n_{ρ,t_1}. It will be a valid trembling sequence and the contributions from this level towards the payments at A and B will be identical. So we then collate and pass to the next higher level.

We now assume that "right" and "left" for player i are different at Γ_{ρ,t_1} and Γ_{ρ,t_2},[53] and we denote the payments to player i at these games as described in Figure 2. Thus, in Γ_{ρ,t_1}, agent i gets h if every agent went "left" and he gets a if he alone switched to "right". Similarly, in Γ_{ρ,t_2} he gets h if every agent moved "right", but gets b, if he alone switched to "left". The payments at the other endpoints form the vectors $\underline{x}$ and $\underline{y}$, written symbolically in Figure 3.

The original trembles $\sigma^n_{-i;\rho,t_1}$ and $\sigma^n_{-i;\rho,t_1}$ induce the following distribution vectors on these endpoints: $\varepsilon^n := (1-\varepsilon_1^n-\varepsilon_2^n, \underline{\varepsilon}_2^n, \varepsilon_1^n)$ at Γ_{ρ,t_1} and $\eta^n := (1-\eta_1^n-\eta_2^n, \underline{\eta}_2^n, \eta_1^n)$ at Γ_{ρ,t_2} (see Figure 3). Here $\underline{\varepsilon}_2^n$ is a vector of probabilities for the payments at the various endpoints that were collectively assigned the payment $\underline{x}$; the sum of its components is equal to the number ε_2^n. Similarly, $\underline{\eta}_2^n$ is a probability vector for the payments $\underline{y}$; the sum of its components is equal to the number η_2^n. In order for the trembles to be valid, these probabilities must converge to $(1, \underline{0}, 0)$ and $(0, \underline{0}, 1)$, respectively, and satisfy:

[52] We use i to denote the player and we identify his agents by indicating to which stage/period game the agent belongs.

[53] But they may still be the same for the other agents at this end game. This is discussed in Case D below.

$$(1 - \varepsilon_1^n - \varepsilon_2^n)h + \underline{\varepsilon}_2^n * \underline{x} + \varepsilon_1^n b \geq (1 - \varepsilon_1^n - \varepsilon_2^n)a + \underline{\varepsilon}_2^n * \underline{y} + \varepsilon_1^n h, \tag{7.1}$$

$$(1 - \eta_1^n - \eta_2^n)h + \underline{\eta}_2^n * \underline{y} + \eta_1^n a \geq (1 - \eta_1^n - \eta_2^n)b + \underline{\eta}_2^n * \underline{x} + \eta_1^n h. \tag{7.2}$$

Here, $*$ means ‘scalar product’. In particular, from $n \to \infty$, it follows that

$$h \geq a, \quad h \geq b. \tag{7.3}$$

These conditions are also sufficient for any sequence of completely mixed perturbations to be test sequences for agent i’s part in Γ_{ρ, t_1} and Γ_{ρ, t_2}.

We distinguish four cases.

Case A. $h > a$ and $h > b$.

In this case (7.1) and (7.2) always hold for all values $\varepsilon_1^n, \underline{\varepsilon}_2^n, \eta_1^n, \underline{\eta}_2^n$, provided they are small enough. Moreover, the left side of (7.2) is strictly smaller than h if we modify, when necessary, the components of $\underline{\eta}_2^n$ and make them small compared to η_1^n, thus satisfying one of the two cases mentioned in the overview of the proof.

Case B. $h > a$ and $h = b$.

Here, again, (7.1) is satisfied whenever ε_1^n and ε_2^n are small enough. Inequality (7.2) becomes

$$\underline{\eta}_2^n * (\underline{y} - \underline{x}) \geq \eta_1^n (h - a). \tag{7.4}$$

We now modify $\sigma_{-i;\rho,t_2}^n$ by increasing η_1^n and decreasing the components of $\underline{\eta}_2^n$ in such a way that (7.4) becomes an equality. We then replace ε_1^n and $\underline{\varepsilon}_2^n$ with the modified values of η_1^n and $\underline{\eta}_2^n$, respectively. Inequality (7.1) will remain valid for small enough η_1^n and $\underline{\eta}_2^n$ and the difference between the left sides of (7.1) and (7.2) will become

$$\underline{\eta}_2^n * (\underline{x} - \underline{y}) + \eta_1^n (h - a) = 0. \tag{7.5}$$

With these modifications the contributions of both perturbations towards (A) and (B) are the same.

Case C. $h = a$ and $h > b$.

Now (7.1) becomes

$$\varepsilon_2^n * (\underline{x} - \underline{y}) \geq \varepsilon_1^n (h - b), \tag{7.6}$$

and (7.2) is satisfied whenever η_1^n and $\underline{\eta}_2^n$ are small enough. By choosing η_1^n and $\underline{\eta}_2^n$ to be equal to ε_1^n and $\underline{\varepsilon}_2^n$, respectively, we find that the difference between (7.1) and (7.2) becomes

$$\underline{\varepsilon}_2^n * (\underline{x} - \underline{y}) + \varepsilon_1^n (b - h) \geq \underline{\varepsilon}_1^n (h - b) + \varepsilon_1^n (b - h) = 0. \tag{7.7}$$

Thus, the contribution of the perturbation towards (A) is larger than, or equal to the contribution towards (B).

Case D. $h = a$ and $h = b$.[54]

Now (7.1) and (7.2) become

$$\underline{\varepsilon}_2^n * (\underline{x} - \underline{y}) \geq 0, \quad \underline{\eta}_2^n * (\underline{y} - \underline{x}) \geq 0. \tag{7.8}$$

From (7.8) it follows that $\underline{x} \not> \underline{y}$ and $\underline{y} \not> \underline{x}$. Then, decreasing the coordinates of $\underline{\varepsilon}_2^n$ and $\underline{\eta}_2^n$, if necessary, until equalities hold in (7.8), by modifying $\sigma^n_{-i;p,t_1}$ and $\sigma^n_{-i;p,t_2}$, we find that the difference between the left sides of (7.1) and (7.2) vanishes, yielding the same contribution towards (A) and (B). ■

The converse theorem is essentially valid. Before introducing it we need a new definition:

Definition 7.3. A pure-strategy profile σ for an extensive-form game representing a one stage/period voting scheme is called *almost perfect*, if there is a sequence of completely mixed strategy profiles σ^n, called *perturbations*, tending to σ, such that for each n, and each agent i, his original pure-strategy in σ is either a best reply against the perturbations of the others, or an irrelevant deviation of that agent is such a best reply. By an *irrelevant deviation* we mean a move that does not change the outcome[55] for the deviating agent in σ.

Theorem 7.4. *Let Γ be an extended form game representing a generic multi-stage/period voting scheme. Let σ be a pure-strategy perfect equilibrium profile for Γ. Denote by $\sigma_{r,s}$ the restriction of σ to a single*[56] *stage/period $\Gamma_{r,s}$. Let the payoffs attached to $\Gamma_{r,s}$ be the ones obtained via the collation process, using σ. Under these conditions $\sigma_{r,s}$ is an almost perfect equilibrium profile for $\Gamma_{r,s}$. It is a perfect equilibrium profile if $\Gamma_{r,s}$ is an end subgame of Γ. It is a perfect equilibrium profile in general if the players are restricted to use only strategies which are functions of the streams alone.*

Proof: Since $\sigma = \{\sigma_{r,s}\}$ is a pure-strategy perfect equilibrium profile in Γ, there exists a test sequence $\{\sigma^n\}_{n=1}^{\infty}$, where $\sigma^n = \{\sigma^n_{r,s}\}$, such that for each n and each agent i, playing in some $\Gamma_{r,s}$, $\sigma_{i;r,s}$ is a best response to σ^n_{-i}. Equivalently, for each n and each one stage/period game $\Gamma^n_{r,s}$, $\sigma_{i;r,s}$ is a best response to $\sigma^n_{-i;r,s}$, where $\Gamma^n_{r,s}$ has the same tree game as $\Gamma_{r,s}$, but the payoffs at its endpoints are derived by collation from the expected payoffs computed at the one stage/period successive endgames below. These payoffs will be called *the perturbed payoffs*. Passing to the limit, we notice that agent i's payoff is not less than any payoff he would have obtained in $\Gamma_{r,s}$ by deviating alone from σ. Because of the genericity assumption, he can receive the same payoff only if his deviation from σ is irrelevant in $\Gamma_{r,s}$.

For the game $\Gamma_{r,s}$ we use as a test sequence the original perturbations re-

[54] This case occurs also if "right" and "left" for the agents different from i are identical.

[55] This does not mean that choosing the irrelevant deviation yields a perfect profile.

[56] We are employing the same notation that was used in the previous theorem.

stricted to this game; namely, $\sigma^n_{r,s}$. Call "left" the original move of agent i of $\Gamma_{r,s}$ in σ. We know that it is a best response against $\sigma^n_{-i;r,s}$ in $\Gamma^n_{r,s}$. If n is large enough, agent i will therefore also lose in $\Gamma_{r,s}$ against $\sigma^n_{-i;r,s}$ if he makes a relevant deviation from his L move. Indeed, in such a move, the perturbed payoff to him will be strictly smaller (being close to the payoffs in $\Gamma_{r,s}$) and the trembles $\sigma^n_{-i;r,s}$ will also not change this fact if n is large enough. On the other hand, if agent i makes an irrelevant deviation, it may well be the case that although "left" was the best response against σ^n_{-i} in $\Gamma^n_{r,s}$, it is not so in $\Gamma_{r,s}$. It may be that the trembles below make it worth his while to move "left", but when there are no trembles except those generated by $\sigma^n_{-i;r,s}$, he might prefer to make the irrelevant deviation when he plays in $\Gamma_{r,s}$.

Of course, the profile restricted to a last stage/period form a perfect equilibrium profile, because there is no tremble in the payoffs from below.[57]

If the players are restricted to strategies that are functions of the streams alone, then the restriction to each $\Gamma_{r,s}$ is a perfect equilibrium profile. Indeed, because of genericity, subgames starting at any two endpoints of $\Gamma_{r,s}$ with equal payoffs to any of the agents therein, are replicas of each other. The trembled payments at the two endpoints are equal for all agents so there is no advantage to a deviation by agent i. ■

8. Existence of pure-strategy perfect equilibrium profiles for multi-stage and multi-period voting schemes

In this section we study the existence of pure-strategy perfect equilibrium profiles for multi-stage and multi-period voting schemes. This includes the (A), (E) and (SIM) protocols with $k > 1$ and the (AE) and (EA) protocols with $k = 1$.

Recall from Proposition 6.1 that when utilities are separable and $k = 1$, sincere voting is essentially the unique pure-strategy perfect equilibrium profile for the (A), (E) and (SIM) protocols. We prove in this section that if utilities are separable and additivity across stages prevails, sincere voting remains a pure-strategy perfect equilibrium profile in the (A) protocol and $k > 1$. However, this natural perfect equilibrium profile is not necessarily unique anymore. Indeed, in Example 8.2 we demonstrate that there may exist two pure-strategy perfect equilibrium profiles, one of which is preferred by all founders. Subsequently, we show, in Example 8.3, that the existence of two such perfect equilibria allow player i to force player j, in a perfect equilibrium, to admit a candidate who is an enemy of j.

The introduction of dynamics to the (E) protocol turns out to have quite a different effect than what it had in the (A) protocol. Indeed, as it is shown in Example 8.4, a pure-strategy perfect equilibrium profile may not exist in the (E) protocol with separable utilities and additivity across stages, already if $k = 2$.

In general, when expulsion is possible and utilities are additive and separable, sincere voting is not necessarily a perfect equilibrium profile in either the (EA) and (AE) protocols with $k \geq 1$, or in the (E) and (SIM) protocols with $k > 1$. It may not even be a Nash equilibrium profile in these protocols.

---

[57] It also follows from the fact that a perfect equilibrium profile induces a perfect equilibrium profile in each subgame. Thus, genericity is not required here.

One reason is that after expelling one's enemies, a member may later find himself expelled by the survivors. Another reason is due to the fact that one may want to expel a friend in order to be able to expel an enemy at a later stage (in the (E) and (SIM) protocols), or in order admit a friend in a later stage in the (EA) and (SIM) protocols.

In the two-period protocols, a pure-strategy perfect equilibrium profile may not exist when $k = 1$, even if we restrict ourselves to weighted agents. Indeed, it is shown in Example 8.7 that pure-strategy perfect equilibrium profiles may not exist in the (AE) protocol when $k = 1$. Somewhat surprisingly, we prove that under the same conditions, a pure-strategy perfect equilibrium profile always exists in the (EA) protocol when $k = 1$.

Theorem 8.1. *Let Γ be a game representing a multistage voting scheme in the A protocol, where utilities are separable and additivity across stages prevails (Assumption 9c). Assume also that friendship is strict at all stages.*[58] *Under these conditions, voting for all of one's friends in every stage is a perfect equilibrium profile.*

Proof: Denote by σ the strategy profile wherein every member votes for all his friends at each stage. Then, after the first stage all common friends are admitted and the composition of the society remains fixed thereafter. If a member deviates alone from σ at some stage game of stage t, he may either lose friends and/or gain enemies. Thus, σ is an equilibrium profile.

Now, consider the following completely mixed (behavioral) strategy profile τ: As long as there are candidates, every member j votes at each stage t for the set of all of his friends with probability $1 - \varepsilon$ and for any other set of candidates with probability $\frac{\varepsilon}{2^{c^t} - 1}$, where $c^t = |C^t|$. We have to show that the pure strategy $\sigma_{i^\star}$, under which $i^\star$ votes for all his friends is a best reply to $\tau_{-i^\star}$ whenever ε is small enough and $i^\star$ is a single agent of player i. That is, we need to show that any pure strategy deviation by $i^\star$ from $\sigma_{i^\star}$ is not profitable for him. Without loss of generality, it is enough to consider deviations of $i^\star$ from $\sigma_{i^\star}$ at the first stage, and that $f := |F^0| \geq 2$.

Denote $U_i := \mathrm{fr}(i)$. Any deviation at the first stage means voting for $(U_i \backslash R) \cup T$, where $R \subseteq U_i$ and $T \cap U_i = \emptyset$, $R \cup T \neq \emptyset$. If such a deviation is profitable, then voting for $U_i \backslash R$ is at least as profitable, because bringing members of T into the society (if successful) would add enemies, or neutrals, who, on their part, may even prevent the admission of some friends; This covers also the case $U_i = \emptyset$ and we may assume that in his deviation i votes for $U_i \backslash R$, where $R \subseteq U_i$, $R \neq \emptyset$.

Denote by d the minimum utility per stage gained by introducing a friend (at any stage and for every member of the society); i.e.,

$$d = \min_{t \in \{1,\dots,k\}} \min_{i \in F^{t-1}, x \in \mathrm{fr}(i)} \{u_i(t, Q \cup \{x\}) - u_i(t, Q) : Q \subseteq N, x \notin Q\}. \quad (8.1)$$

By the assumption of strict friendship, it follows that $d > 0$. Denote by M the

[58] I.e., adding a friend of i to the society strictly increases the utility to i at each stage. (Actually, it is enough to make this assumption for periods in which i is in the society.)

maximum utility per stage gained by introducing any set of friends (at any stage and for every member); i.e.,

$$M = \max_{t \in \{1,\ldots,k\}} \max_{i \in F^{t-1}, P \subseteq \mathrm{fr}(i)} \{u_i(t, Q \cup P) - u_i(t, Q) : Q \subseteq N, P \cap Q = \emptyset.\} \tag{8.2}$$

We want to study the effect of i voting for $U_i \backslash R$. We may condition our calculations on the event that $F^0 \backslash \{i\}$ vote at stage 1 for a set containing R. If i refrains from voting for members of R, his immediate loss is at least d (since $R \neq \emptyset$). His motivation for refraining from voting to admit the members of R stems from his concern that these members may prevent the admission of a set of friends Q at later stages. These candidates can be of two types: Those, denoted by P, are friends of each member of F^0 and those who are not. Members of P were not admitted at the first stage because some member(s) of $F^0 \backslash \{i\}$ trembled. This can happen with probability of at most $1 - (1 - \varepsilon)^{f-1}$. Thus, even if P were admitted in the next stage, (and i refraining from voting for members of R, who might prevent admitting members of P), the gain would be at most $M(k-1)(1 - (1-\varepsilon)^{f-1}) \leq M(k-1)(f-1)\varepsilon$, by Lemma 6.4.

Members of $Q \backslash P$ would be admitted at a future stage only if all members of $F^0 \backslash \{i\}$ tremble to vote for them. Each member x therein would be admitted with probability $\frac{2^{(c^t-1)}\varepsilon}{2^{c^t}-1} \leq \varepsilon$ at stage t, and, therefore, with probability at most $(k-1)\varepsilon$ at any future stage. Such a member will contribute at most $k(k-1)\varepsilon M$. The total contribution of $Q \backslash P$ is at most $c^0 k(k-1)M\varepsilon$. Thus, a deviation from U_i to $U_i \backslash R$ is not profitable if $d > (k-1)(f-1)M\varepsilon + c^0 k(k-1)M\varepsilon$, which is indeed the case if ε is small enough. Since $\tau(\varepsilon) \to \sigma$ when $\varepsilon \to 0$, we have proved that σ is indeed a perfect equilibrium profile. ■

Recall from Proposition 6.1 that voting sincerely in the (A), (E) and (SIM) protocols is essentially the unique pure-strategy perfect equilibrium profile in the 1-stage generic case. However, this is not the case if $k > 1$. Indeed, in the following example we describe two pure-strategy perfect equilibrium profiles for the (A) protocol with $k = 2$. One of these two profiles is voting sincerely at each stage, while in the other, each player votes for a proper subset of his friends in the first stage and sincerely in the second stage.

Example 8.2. The protocol is (A). The population consists of $F^0 = \{1, 2\}$, $C^0 = \{a, b\}$, $A_i = -100$ for all players i and $k = 2$. The voting scheme obeys Assumption 9d (weighted agents) and $w_1(a) = 10$, $w_1(b) = 8$, $w_2(a) = 8$, $w_2(b) = 10$, $w_a(b) = w_b(a) = -1$. It follows from Proposition 6.1 and the footnote to Theorem 7.4 that in any perfect equilibrium, every voter must vote for all his friends (that have not yet become members of the society) in the second stage. We can therefore compute all the possible outcomes that result from the votes at stage 1. These outcomes, for the original founders,[59] are given by the following matrix:

[59] We exclude constant utilities derived from the presence of the original founders in all possible streams.

	$\varnothing$	a	b	ab
$\varnothing$	18, 18	18, 18	18, 18	18, 18
a	18, 18	20, 16	18, 18	20, 16
b	18, 18	18, 18	16, 20	16, 20
ab	18, 18	20, 16	16, 20	36, 36

.

Removing weakly dominated rows and columns, we find that the only weakly undominated 1-stage moves are a and ab for player 1 and b and ab for player 2. The only equilibrium profiles using undominated strategies are (a, b) and (ab, ab) at the first stage, and since they employ undominated strategies in a two-person game for the original founders,[60] it follows from Kohlberg and Mertens [2], that both are perfect equilibrium profiles. Since the game is generic, it follows from the Collation Theorem that voting either (a, b) or (ab, ab) by the original founders in the first stage and voting sincerely by everyone in the second stage are perfect equilibrium profiles for this game.

Note that one of the two pure-strategy perfect equilibrium is preferred to the other by both founders. This feature will be used in the next example to introduce, effectively, threats of punishment. It is well known that in ordinary Nash equilibria players can use the threat of punishment to force other players to adhere to an agreed upon strategy profile which is inferior to them. Thus, for example, two members can jointly force a third member to vote in favor of an enemy of his, since otherwise, the two members will not vote to invite a very good common friend of all three members. However, as is well known, such equilibria are often not subgame perfect and therefore not even perfect. Nevertheless, we demonstrate in the following example that a member i can still effectively employ a punishment that harms him, in order to "convince" another member j to vote to admit a candidate who is j's enemy even in a perfect equilibrium. The punishment in this case is the implementation of an inferior perfect equilibrium for both players.

Example 8.3. The population consists of $F^0 = \{1, 2\}$, $C^0 = \{a, b, c\}$ and $k = 3$. The voting scheme obeys Assumption 9d (weighted agents) and the protocol is (A). Further, let

$$w_1(a) = 10, \quad w_1(b) = 8, \quad w_1(c) = 7;$$

$$w_2(a) = 8, \quad w_2(b) = 10, \quad w_2(c) = -3;$$

$$w_c(a) = 1, \quad w_c(b) = 5,$$

$$w_i(j) = -1, \text{ otherwise.} \quad A_i = -100 \text{ for all } i.$$

[60] We regard the first stage as a two-person game, because the moves of agents a and b do not affect the payoffs.

To ensure that the voting scheme is generic, we perturb the utilities of the agents slightly, e.g., by adding $\varepsilon, \varepsilon^2$, etc., to the weights of the agents, where ε is *positive* and small enough. Thus, e.g., $w_1(a) = 10 + \varepsilon$, $w_1(b) = 8 + \varepsilon^2$ and $w_1(c) = 7 + \varepsilon^3$, etc. Then, for ε sufficiently small, the utilities of two streams for each agent are equal if and only if the streams are identical.

Consider the following strategy profile:

Stage 1. Both original founders vote to admit c.

Stage 2. If some candidates are admitted in Stage 1, all voters vote henceforth to admit all their friends. If nobody is admitted, player 1 votes to admit ac and player 2 votes to admit b.

Stage 3. Every member votes to admit all his friends.

Knowing the continuations after stage 1, we can compute the game, Γ^1, collated until the first stage. The payoffs, ignoring the ε's, are given in the matrix below.[61] It can be seen therein that the row and column corresponding to a vote for player c are not dominated by other rows and columns, even when the ε's are taken into account. Furthermore, even with the ε's taken into account, (c, c) is an equilibrium profile performed by undominated strategies as equal payoffs in the matrix only result from the same streams of members and therefore, for ε sufficiently small undomination does not depend on the

	$\varnothing$	a	b	ab	abc	bc	ac	c
$\varnothing$	18, 18	18, 18	18, 18	18, 18	18, 18	18, 18	18, 18	18, 18
a	18, 18	30, 24	18, 18	30, 24	30, 24	18, 18	30, 24	18, 18
b	18, 18	18, 18	24, 30	24, 30	24, 30	24, 30	18, 18	18, 18
ab	18, 18	30, 24	24, 30	54, 54	54, 54	24, 30	30, 24	18, 18
abc	18, 18	30, 24	24, 30	54, 54	75, 45	45, 21	51, 15	57, 27
bc	18, 18	18, 18	24, 30	24, 30	45, 21	45, 21	57, 27	57, 27
ac	18, 18	30, 24	18, 18	30, 24	51, 15	57, 27	51, 15	57, 27
c	18, 18	18, 18	18, 18	18, 18	57, 27	57, 27	57, 27	57, 27

[61] We exclude constant utilities derived from the presence of the original founders in all possible streams.

size of ε. By Kohlberg and Mertens [2], (c, c) is a perfect equilibrium profile for the first stage when ε is small enough.

To apply the Collation Theorem (Theorem 7.1), we have to check the various 1-stage subgames obtained by collation. By Theorem 8.1, all strategy profiles for the subgames starting with stage 2 are perfect, except, perhaps, if the first stage results with no member being invited. It remains to show that the strategy prescribed in stage 2 when nobody is admitted in stage 1, is indeed a perfect equilibrium. Knowing the continuations in the last stage, we can construct the payoff matrix facing the two founders in stage 2 when no new members were admitted in stage 1. It is given below, ignoring the ε's.

	$\varnothing$	a	b	ab	c	ac	bc	abc
$\varnothing$	18 18	18 18	18 18	18 18	18 18	18 18	18 18	18 18
a	18 18	20 16	18 18	20 16	18 18	20 16	18 18	20 16
b	18 18	18 18	16 20	16 20	18 18	18 18	16 20	16 20
ab	18 18	20 16	16 20	36 36	18 18	20 16	16 20	36 36
c	18 18	18 18	18 18	18 18	32 12	32 12	32 12	32 12
ac	18 18	20 16	18 18	20 16	32 12	34 10	32 12	34 10
bc	18 18	18 18	16 20	16 20	32 12	32 12	30 14	30 14
abc	18 18	20 16	16 20	36 36	32 12	34 10	30 14	50 30

From the payoff matrix above one observes that player 1 voting to admit $\{a, c\}$ and player 2 voting to admit b is a pure-strategy Nash equilibrium profile. None of the two founders employs a weakly dominated strategy and therefore, by Kohlberg and Mertens [2] theorem, it is a perfect equilibrium profile. Thus, we have established that the strategy profile, specified above, in which player 1 "convinces" player 2 to invite his enemy (player c) is indeed a perfect equilibrium.

Unlike the (A) protocol, the following example shows that in the (E) protocol there are 2-stage voting schemes for which no pure-strategy perfect equilibrium profile exists even if utilities are separable.

Example 8.4. The protocol is (E) and $k = 2$. The utilities obey weighted agents (Assumption 9d). $F^0 = \{1, 2, 3, a, b, c\}$ and $A_i = -100$ for each founder i. The weights of the various agents are:

$$w_1(a) = -1, \quad w_1(c) = -10, \quad w_1(i) = -1, \text{ otherwise;}$$

$$w_2(a) = -10, \quad w_2(b) = -1, \quad w_2(i) = -1, \text{ otherwise;}$$

$$w_3(b) = -10, \quad w_3(c) = -1, \quad w_3(i) = -1, \text{ otherwise;}$$

$$w_a(1) = 300, \quad w_a(2) = 300, \quad w_a(i) = -1, \text{ otherwise;}$$

$$w_b(2) = 300, \quad w_b(3) = 300, \quad w_b(i) = -1, \text{ otherwise;}$$

$$w_c(3) = 300, \quad w_c(1) = 300, \quad w_c(i) = -1, \text{ otherwise.}$$

It is a dominant strategy for agents a, b and c to vote sincerely, because a, b, or c, admitting a friend overcomes the harm done if this friend manages to expel them at the next stage, and voting to expel an enemy has the advantage of eliminating one who may later vote to expel the agent. That is, agent a will veto the expulsion of agents 1 and 2, agent b will veto the expulsion of agents 2 and 3 and agent c will veto the expulsion of agents 1 and 3. As to agents 1, 2, 3 – each one of them would like to keep exactly one of his 'admirers'; namely, the one whose weight is -1, so as not to be kicked out of the society at the second stage. Also, it is dominant to vote to expel an enemy who is not an "admirer". In any case, it is a dominant strategy for 1, 2 and 3 to vote to expel agents b, c and a, respectively. Given that a, b and c vote sincerely, we can determine the payoffs to 1, 2 and 3, for any of their relevant strategies, thereby determining all pure-strategy Nash equilibria, conditioned on sincere votes of agents a, b and c and everyone voting sincerely in the second stage. This yields the following matrices (agent 1 chooses the row, agent 2 chooses the column and agent 3 chooses the matrix), where the strategies are defined as *whom to veto; i.e., whom to prevent from expulsion*. The constant utility numbers derived from the fact that 1, 2 and 3 are in the society in the first stage were ignored, since in all cases, agents 1, 2 and 3 are not expelled in the first stage. Thus, excluding normalization, all entries in the arrays below will have to be reduced by 2.

	$\varnothing$	a	b	ab
$\varnothing$	−100 −100 −100	−2 −11 −101	−101 −20 −11	−4 −13 −13
a	−2 −11 −101	−2 11 −101	−4 −13 −13	−4 −13 −13
c	−11 −101 −2	−13 −13 −4	−13 −4 −13	−14 −14 −14
ac	−13 −13 −4	−13 −13 −4	−14 −14 −14	−14 −14 −14

Agent 3 vetoes $\varnothing$

	∅	*a*	*b*	*ab*
∅	−101 −2 −11	−4 −13 −13	−101 −2 −11	−4 −13 −13
a	−4 −13 −13	−4 −13 −13	−4 −13 −13	−4 −13 −13
c	−13 −4 −13	−14 −14 −14	−13 −4 −13	−14 −14 −14
ac	−14 −14 −14	−14 −14 −14	−14 −14 −14	−14 −14 −14

Agent 3 vetoes *b*

	∅	*a*	*b*	*ab*
∅	−11 −101 −2	−13 −13 −4	−13 −4 −13	−14 −14 −14
a	−13 −13 −4	−13 −13 −4	−14 −14 −14	−14 −14 −14
c	−11 −101 −2	−13 −13 −4	−13 −4 −13	−14 −14 −14
ac	−13 −13 −4	−13 −13 −4	−14 −14 −14	−14 −14 −14

Agent 3 vetoes *c*

	∅	*a*	*b*	*ab*
∅	−13 −4 −13	−14 −14 −14	−13 −4 −13	−14 −14 −14
a	−14 −14 −14	−14 −14 −14	−14 −14 −14	−14 −14 −14
c	−13 −4 −13	−14 −14 −14	−13 −4 −13	−14 −14 −14
ac	−14 −14 −14	−14 −14 −14	−14 −14 −14	−14 −14 −14

Agent 3 vetoes *bc*

Inspecting these matrices, we find that the only Nash equilibrium profiles for the first stage that do not employ a dominant strategy and could possibly therefore be perfect, are: (a,b,b), (a,ab,b), (a,a,c), (c,c,b), (ac,a,c), (c,b,bc) and (ac,ab,bc). In each of these profiles, at least one agent is protected by two of his enemies and he vetoes at least one of them. He will benefit against any tremble, if he removes his veto from one of his enemies, provided the tremble is small enough. Indeed, the probability that at least one of his protectors remains in such a tremble is much larger than the probability that both will be expelled and this offsets his loss if he is kicked out. We conclude that the above profiles are not perfect and therefore there is no pure-strategy perfect equilibrium profile for the entire game. ■

We next demonstrate the existence of a pure-strategy perfect equilibrium profile for the one-stage (EA) protocol when utilities satisfy the condition of weighted agents (Assumption 9d).

Theorem 8.5. *Let Γ be a game representing a 1-stage generic voting scheme in the (EA) protocol, having at least three members. If utilities obey the weighted agents assumption (Assumption 9d), then there always exists a pure-strategy perfect equilibrium profile.*

Proof: The votes in the A-period are sincere (footnote to Theorem 7.4). Then, we can regard the scheme as following the (E) protocol, when the founders compute their utilities knowing what will happen in the A-period. Observe that no founder will veto an enemy in the E-period. Indeed, such an enemy causes him a loss if he is not expelled, and may also prevent the admission of some of his friends.

To proceed with the proof we now describe a procedure which generates an equilibrium profile yielding an outcome set S determined by mutually disjoint veto sets.

A referee approaches a founder i and asks him whom does he wish *to veto* (from expulsion). The founder provides him with the best option he has, assuming (hypothetically) that the vote is final. By genericity, his veto set P_i^1 is unique. The referee now approaches another founder, j, telling him that founder i already vetoed P_i^1. Does he wish to veto additional founders? Founder j provides the best available unique answer denoted P_j^1. The referee then goes to another founder, or, perhaps returns to founder i, informing him about the set vetoed so far and asks for the best addition the founder may have. This continues until no founder wishes to veto any additional agent. The result for each founder i is a set $P_i = P_i^1 \cup P_i^2 \cup \cdots \cup P_i^{m_i}$, where P_i^t is the, possibly empty, set founder i vetoed when approached at the tth time and m_i is the number of times founder i was approached. Denote $S := N \setminus \bigcup_{i \in F^0} P_i$. Then, $\{P_i\}_{i \in F^0}$ are mutually disjoint sets whose union is $N \setminus S$. By Proposition 6.7, it can be extended to a pure-strategy perfect equilibrium profile if it is an equilibrium profile. To prove that this is indeed the case, we need the following result:

Lemma 8.6. *Under the conditions of Theorem 8.5, let A and B be sets of founders satisfying $A \subset B$. Let C be a set of friends of founder i in the society* *which is disjoint from B. If i prefers that $A \cup C$ is vetoed to A being vetoed, then he would also prefer that $B \cup C$ is vetoed to B being vetoed.*[62]

Proof: From the data it follows that when A is contemplated, the harm caused to i because agents in C prevent the admission of some candidates in the (A) period, is less than the benefit he gains by vetoing C. When B is contemplated, the gain from vetoing C is the same, and the harm is either the same, or less, if members of $B \backslash A$ already prevent a few admissions that would be prevented by C. ■

Continuation of the proof of Theorem 8.5. Suppose that $\{\text{veto } P_i\}_{i \in F^0}$ is not an equilibrium profile. Let i be a founder who can benefit by deviating alone. Let the profitable deviation be vetoing $(P_i \backslash Q) \cup R$, where $Q \subseteq P_i$ and $R \cap P_i = \emptyset$. At least one of the sets Q, or R is not empty and both consist of friends of i. Denote $Q := Q^1 \cup Q^2 \cup \cdots \cup Q^{m_i}$, where $Q^t := P_i^t \cap Q$ is the set of friends in Q that i deleted at the tth approach. We shall show that vetoing $P_i \cup R$ would yield founder i even higher profits. This is impossible, because he told the referee that he does not wish any further member vetoed. To this end concentrate on the first encounter of founder i with the referee. He stated that, given what the other members already vetoed, he prefers to veto P_i^1 instead of, say, $(P_i^1 \backslash Q^1)$, when $Q^1 \neq \emptyset$, because P_i^1 was his best set. By Lemma 8.6 he would also prefer to veto $(P_i^1 \backslash Q^1) \cup Q^1 \cup R$ to $(P_i^1 \backslash Q^1) \cup R$, given the same preceding votes of the other members. Invoking Lemma 8.6 once more, agent i would prefer vetoing $\bigcup_{j=1}^{m_i} (P_i^j \backslash Q^j) \cup Q^1 \cup R$ to $\bigcup_{j=1}^{m_i} (P_i^j \backslash Q^j) \cup R$. Thus, he would prefer to continue vetoing Q^1 in his original deviation. So, let us modify the original deviation by keeping Q^1 in his veto set.

We continue in this fashion m_i steps until we reach a more profitable deviation for i; namely, to enlarge his veto set P_i with the inclusion of only the set R. This, as we have shown, is manifestly a contradiction if $R \neq \emptyset$. We have demonstrated that $\{\text{veto } P_i\}_{i \in F^0}$ is a Nash equilibrium profile. By Proposition 6.7, it can be extended to a pure-strategy perfect equilibrium profile for the E-period, after collation in the A-period, and by the Collation Theorem (Theorem 7.1), we conclude that there always exists a pure-strategy perfect equilibrium for the (EA) protocol and $k = 1$, when utilities obey the weighted agents assumption. ■

The next intriguing example shows that some of the perfect equilibria, whose existence was assured above, are questionable. Thus, perfectness does not guarantee plausibility.

Example 8.7. Consider a 1-stage voting scheme in the (EA) protocol, where $F^0 = \{1, 2, 3\}$, $C^0 = \{a\}$ and utilities obey weighted agents (Assumption 9d). Suppose further that

$$w_1(2) = 1, \quad w_1(3) = 1, \quad w_1(a) = 10;$$

$$w_2(1) = 1, \quad w_2(3) = -1, \quad w_2(a) = -1;$$

$$w_3(1) = 1, \quad w_3(2) = -1, \quad w_3(a) = -1$$

[62] We compare two hypothetical events: A [resp. B] remain for the A period vis $A \cup C$ [resp. $B \cup C$] remain for the A period.

and $A_i = -100$ for each agent. Since 2 and 3 will veto admitting a at the A-period, agent 1 will vote to expel them in the E-period. He will succeed, because 2 and 3 "hate" each other and vote to expel each other. They, in turn, will not vote to expel agent 1, as it will do them no good. They are out anyhow. If any of them is not expelled, due to a tremble, there is no benefit for him if 1 is expelled. Similarly, it will do 2, or 3 no good to veto the expulsion of the other. They will be out anyhow, although 1 may succeed to recruit a.

This is the unique perfect equilibrium. It is also Pareto undominated. Nevertheless, we are convinced that the voters will not accept it. Agent 2, for example would declare: I am going to veto the expulsion of 3. Now act as you like. Then, if 1 believes him (and there is an obvious ground to this belief), he stands to merely lose a friend if he votes to expel him. Thus, 1 voting sincerely and 2 and 3 voting $\varnothing$ in the E-period and then voting sincerely in the A-period seems to us a safer Nash equilibrium profile. It is also Pareto undominated, but it is not a perfect profile.

It is interesting to realize that although there exists a pure-strategy perfect equilibrium profile for the (EA) protocol when utilities obey the weighted agents assumption (Assumption 9d) and $k = 1$, there need not exist a pure-strategy perfect equilibrium profile in the (AE) protocol under the same conditions.

Example 8.8. The population consists of $F^0 = \{1, 2, 3\}$, $C^0 = \{a, b\}$. The protocol is (AE) and $k = 1$. The utilities obey weighted agents (Assumption 9d) and are:

$$\begin{array}{llll} w_1(2) = -2, & w_1(3) = -10, & w_1(a) = 14, & w_1(b) = 9; \\ w_2(1) = 10, & w_2(3) = -9, & w_2(a) = -8, & w_2(b) = 15; \\ w_3(1) = 2, & w_3(2) = -1, & w_3(a) = 4, & w_3(b) = 8; \\ w_a(1) = 1, & w_a(2) = 2, & w_a(3) = 4, & w_a(b) = 8; \\ w_b(1) = 1, & w_b(2) = 2, & w_b(3) = 4, & w_b(a) = 8. \end{array}$$

The utility for being out of the society is taken to be -100 for each agent. Here, each member has 4 possibilities at the A-period. Taking into account that the votes in the E-period are "sincere" (footnote to Theorem 7.4 and Proposition 6.1), we can compute the 2-period payoffs that result from the votes in the A-period.

We first find all pure-strategy equilibrium profiles and then show that none of them is perfect. To this end, notice that voter 3 has a dominant strategy $\{a, b\}$, which is the only one he can employ in a pure-strategy perfect equilibrium profile. We therefore assume that he votes $\{a, b\}$.

Considering the strategies for agents 1 and 2, we observe that if no one is admitted, then 2 and 3 are expelled, giving the outcome $\{1\}$. If a is admitted, then no one is expelled, giving $\{1, 2, 3, a\}$; if b is admitted, no one is expelled, so $\{1, 2, 3, b\}$ is the outcome; if $\{a, b\}$ is admitted, no one is expelled, so $\{1, 2, 3, a, b\}$ results.

The utilities for these four outcomes are, respectively:

1:	0,	2,	−3,	11.
2:	−100,	−7,	16,	8.
3:	−100,	5,	9,	13.

The payoff matrix for agents 1 and 2 is therefore:

	$\varnothing$	a	b	ab
$\varnothing$	0, −100	0, −100	0, −100	0, −100
a	0, −100	2, −7	0, −100	2, −7
b	0, −100	0, −100	−3, 16	−3, 16
ab	0, −100	2, −7	−3, 16	11, 8

The only pure-strategy equilibrium profile are $(\varnothing, \varnothing, \{a,b\})$, $(\varnothing, \{b\}, \{a,b\})$ and $(\{a\}, \{a\}, \{a,b\})$. The first two profiles are not perfect, because a vote of $\{a\}$ by agent 1 weakly dominates a vote of $\varnothing$ by him. Similarly, the third profile is not perfect, because voting $\{a,b\}$ weakly dominates voting $\{a\}$ by agent 2. Thus, after collating using sincere voting in the last period there is no Nash equilibrium in the first period. This implies that in this example there does not exist a pure-strategy perfect equilibrium.

9. The importance of the agenda

Members of a society may wish to select the protocol under which the society will operate. It is therefore of interest to study who will benefit and who will lose under the various protocols. In this section we compare the 1-stage (AE), (EA) and (SIM) protocols, assuming that the members use perfect equilibrium profiles. We assume that the protocols obey the weighted agents assumption (Assumption 9d) and that not to be in the society is the worst outcome for each agent. For simplicity we assume that $w_i(j) \neq 0$ whenever $i \neq j$ and that $|F^0| \geq 3$. We also assume that $w_i(i) = 0$ so that we can write $u_i(\{i\} \cup S)$ instead of '$u_i(S)$ conditioned on i being a member of the society' (see (2.6)).

We show that in any of the pure-strategy perfect equilibrium profiles under the above restrictions, the survivors under the (EA) protocol are always also survivors under the other protocols and, moreover, they are not worse off under the (EA) protocol.

So far we assumed that in the (AE) protocol one is not allowed to expel a newly admitted candidate and in the (EA) protocol one is not allowed to admit a recently expelled member. It will be shown that the first restriction is redundant for the analysis carried out in this section and it is therefore relaxed. The second restriction, however, must be maintained.

Lemma 9.1. *In any pure-strategy perfect equilibrium profile under the (AE) protocol, in which utilities satisfy the weighted agent assumption (Assumption 9d), a founder i will vote to admit his enemy j in the A-period only if this vote avoids i's expulsion. Consequently, if there is a founder in $F^0\backslash\{i\}$, who likes i, then candidate j, who is an enemy of i, will be invited by i only if i is sure that j will be expelled at the E-period;*[63] *i.e., if j is an enemy of all the interim voters.*

Proof: An enemy, if invited and not expelled, adds negative utility which will occur with positive probability, at least under a tremble, and moreover, this enemy may prevent the expulsion of enemies in the E-period. If i is not in danger of being expelled, he is better off vetoing the admission of j. ■

A veto by member i to expel his enemy in the E-period of an (EA) protocol, adds a negative utility to i whenever all other members vote for such an expulsion. Moreover, such an enemy may prevent i from admitting some of his friends in the A-period. Therefore we conclude:

Lemma 9.2. *In any pure-strategy perfect equilibrium profile under the (EA) protocol, a founder will always vote to expel his enemies.*[64]

Notation. We denote by σ_X an arbitrary pure-strategy perfect equilibrium profile under a protocol (X). Similarly, we denote by E_X and A_X the set of expelled and admitted agents, respectively, under protocol (X), $X=$ AE, SIM or EA, as the case may be.

Proposition 9.3. *Let Γ be a game representing a voting scheme satisfying the requirements stated at the beginning of this section. Then, for every pure-strategy perfect equilibrium profiles σ_{AE}, σ_{EA} and σ_{SIM},*

$$F^0 \cap E_{AE} \subseteq E_{EA} \quad \textit{and} \quad E_{SIM} \subseteq E_{EA}. \tag{9.1}$$

Thus, every founder that survives under σ_{EA} also survives under both σ_{SIM} and σ_{AE}.

Proof: A. Under (SIM) only $F^0 \cap \text{comen}(F^0)$ are expelled,[65] because the vote is sincere. By Lemma 9.2, these members are expelled under the (EA) protocol and perhaps also others.

B. Since the second period is sincere,

$$F^0 \cap E_{AE} = F^0 \cap \text{comen}(F^0 \cup A_{AE}). \tag{9.2}$$

Consequently,

$$F^0 \cap A_{AE} \subseteq F^0 \cap \text{comen}(F^0) \subseteq F^0 \cap E_{EA} = E_{EA}. \tag{9.3}$$

The last inclusion follows from Lemma 9.2. ■

---

[63] This, of course, cannot happen if expulsion of an admitted candidate is prohibited.
[64] This claim is incorrect if admission of an expelled candidate is permissible (see Example 9.5).
[65] 'comen' and 'comfr' stand for 'common enemies' and 'common friends'.

Proposition 9.4. *Let Γ be a game representing a voting scheme satisfying the requirements stated at the beginning of this section. Let σ_{AE}, σ_{EA}, and σ_{SIM} be arbitrary pure-strategy perfect equilibrium profiles. Then, for every founder i that survives σ_{EA},*

$$u_i(\sigma_{EA}) \geq u_i(\sigma_{SIM}) \quad \text{and} \quad u_i(\sigma_{EA}) \geq u_i(\sigma_{AE}). \tag{9.4}$$

Proof: A. By (9.1), if i is a survivor under the (EA) protocol, he is not expelled under the (SIM) protocol. That is, $i \notin F^0 \cap \text{comen}(F^0)$. Let i deviate from σ_{EA} by vetoing the expulsion of all founders not in $F^0\backslash\text{comen}(F^0)$, and voting sincerely in the A-period. He cannot gain by the deviation, because σ_{EA} was an equilibrium profile. The outcome from such a deviation would be precisely $F^0 \cap \text{comen}(F^0)$, because the others vote to expel sets containing $F^0 \cap \text{comen}(F^0)$, which will give him the payoff he would have received under (SIM).

B. Since σ_{EA} is an equilibrium profile, a surviving founder i cannot gain if he deviates alone. Let him deviate to a pure-strategy σ_i', which is sincere vote in both periods. Our proof will be concluded if we show that even after such a deviation he would receive at least as much as he receives in σ_{AE}.

Dividing the members of F^1 into founders and agents newly admitted, the outcomes under the two protocols are:

$$\begin{aligned} u_i(\sigma_{EA;-i}, \sigma_i') = {} & u_i(F^0\backslash(E_{EA} \cap \text{en}(i))) \\ & + u_i(\{i\} \cup [\text{comfr}(F^0\backslash(E_{EA} \cap \text{en}(i))) \cap C^0]), \end{aligned} \tag{9.5}$$

$$u_i(\sigma_{AE}) = u_i(F^0\backslash\text{comen}(F^0 \cup A_{AE})) + u_i(\{i\} \cup [A_{AE}\backslash\text{comen}(F^0 \cup A_{AE})]). \tag{9.6}$$

Note that i belongs to both sets of (9.5), because he is a survivor under σ_{EA} and is not expelled at E_{EA}. By Proposition 9.3 he survives also (AE) and therefore belongs to both sets of (9.6). Observe also that in (9.6) we allow the expulsion of a newly admitted agent. Now, any founder i who is a survivor of σ_{EA}, has at least one founder that likes him. Therefore, by Lemma 9.1, he will agree to admit an enemy under the (AE) protocol only if he is sure that he can expel him at the E-period. Consequently, $A_{AE}\backslash\text{comen}(F^0 \cup A_{AE})$ consists only of members who are friends of every founder that survived σ_{EA}. Now, by Lemma 9.2, $F^0 \cap (\text{comen}(F^0 \cup A_{AE})) \subseteq F^0 \cap \text{comen}(F^0) \subseteq E_{EA} \cap \text{en}(i) \subseteq \text{en}(i)$; therefore,

$$u_i(F^0\backslash(E_{EA} \cap \text{en}(i))) \geq u_i(F^0\backslash\text{comen}(F^0 \cup A_{AE})). \tag{9.7}$$

Also,

$$A_{AE}\backslash\text{comen}(F^0 \cup A_{AE}) \subseteq \text{comfr}(F^0\backslash(E_{EA} \cap \text{en}(i))) \cap C^0. \tag{9.8}$$

Indeed, let x be a member of the left side of (9.8). Then, by Lemma 9.1 and Proposition 9.3, he is a friend of every surviving founder under σ_{EA}, and therefore also under its modification. Therefore x is also a member of the right side of (9.8), since he is also a candidate. Since both sets consist of friends of i, it follows that

$$u_i(\{i\} \cup [\text{comfr}(F^0 \backslash E_{EA} \cap \text{en}(i)) \cap C^0])$$

$$\geq u_i(\{i\} \cup [A_{AE} \backslash \text{comen}(F^0 \cup A_{AE})]). \quad (9.9)$$

By (9.5), (9.6), (9.7) and (9.9), $u_i(\sigma_{EA}) \geq u_i(\sigma_{EA;-i}; \sigma_i') \geq u_i(\sigma_{AE})$. ∎

We have seen that in the (AE) protocol we can relax the restriction of not expelling an admitted agent in the same stage. The next interesting example shows that under the (EA) protocol we must require that no expelled agent is admitted in the same stage in order for Proposition 9.4 to be true.

Example 9.5. Let $F^0 = \{1, 2\}$, $C^0 = \{a\}$, $k = 1$ and the utilities satisfy weighted agents (Assumption 9d) with

$$w_1(2) = 1, \quad w_1(a) = 2,$$

$$w_2(1) = -1, \quad w_2(a) = -1,$$

$$w_a(1) = -1, \quad w_a(2) = -1,$$

$$A_i = -10, \quad \text{all agents.}$$

The following is the unique pure-strategy perfect equilibrium profile under the modified (EA) protocol which allows the admission of a previously expelled founder: Founder 1 expels founder 2 in order to be able to admit candidate a in the A-period. Founder 2 votes $\emptyset$ in both periods. (He does not expel his enemy, because he wants agent 1 to readmit him.) At the "sincere" A-period, both 2 and a are admitted and the outcome $F^1 = \{1, 2, a\}$ yields the founders, 1 and 2, the payments 3 and -2, respectively.

Under the (AE) protocol, no one is admitted in the first stage and only founder 2 survives. Thus, Proposition 9.3 and 9.4 are not valid in an (EA) protocol which allows admission of previously expelled members.

We conjecture that there is no comparison between the (AE) and the (SIM) protocols; namely, that for an appropriate game, some common survivors will prefer a perfect equilibrium of the (AE) protocol and others – a perfect equilibrium of the (SIM) protocol.

10. Overview and discussion

In this section we review the main results of this paper and attempt to judge them critically. We indicate what has been achieved so far, but, more importantly, we try to point out what is still missing. One aim is to offer suggestions for further research.

This study and the study by Barberà et al. [2001] show that dynamic issues concerning developing societies can be studied using game theoretical techniques. The models studied are general on the one hand and particular on the other. They are are quite general, because no assumption is made on the utility functions of the agents (except that additivity is assumed when we discuss perfect equilibrium profiles). These models are highly particular however, because we endow each agent with the power of either admitting any set of candidates at any stage, in the Barberà et al.'s paper, or vetoing the admission and expulsion of any agent other than himself in the present paper. Undoubtedly, voting rules that do not impose these restrictions could have other interesting features. For example, one would be able to analyze more thoroughly logrolling and its dependance on the number of agents a member needs in order to push forward various issues.

We did not introduce another option that can play an important role in the determination of the set of equilibrium outcomes; namely, the option of resignation. In the real world, resignations often take place both as threats and as parts of the equilibrium path (e.g., when the agent is available only for a limited time period).

Closely related to the option of resignation is the option of several agents resigning and forming other societies, or joining other existing societies. Research in this direction is probably difficult, as it embodies both non-cooperative and cooperative considerations.

Section 2 specifies several other simplifying assumptions, each of which can be relaxed, giving rise to many other models having perhaps completely different properties. It would be interesting to study, for example, models with different admission/expulsion rules. Obviously, many such models can be constructed. One can then attempt to look for properties that are common to classes of voting rules or study who stands to gain/lose under the various rules.

One major limitation of the present research is the assumption of a finite horizon. Study of infinite horizon models, or finite horizon models whose length is not known, seems attractive and important.

Even though the games associated with our voting protocols are not games of perfect information, their structure is simple enough to permit a kind of backward induction (called *collation*). Backward induction works also when we consider perfect equilibria (Sections 5 and 7). Therefore, when discussing pure-strategy subgame-perfect and perfect equilibrium profiles and equilibrium outcomes, we can limit the study to 1-stage games. Here, we find duality relations between the two papers. Theorems that are stated in the Barberà et al.'s paper have their duals in the present paper, where "vetoing" replaces "voting". This is an example of a comparison between two voting models. Another comparison of a different nature is carried out in Section 9, where we study who stands to gain/lose in perfect equilibrium outcomes under various protocols.

We characterize in Section 4 all 1-stage equilibrium outcomes that result from pure-strategy profiles. We implement these outcomes by strategies that involve common-voting. These profiles could be criticized on the basis that it is hard to achieve common voting in the real world. We therefore provide in Proposition 4.8 other profiles that involve common voting only on a subset of the available agents. Undoubtedly, there may exist many other equilibrium profiles and it is interesting to discover them.

One topic missing from this paper is a characterization of equilibrium profiles which are not subgame perfect. These often occur in the real word when threats are employed off the equilibrium path. They have the advantage that they yield higher payoffs, even Pareto dominating ones. Studying such equilibria requires deeper analysis of the interrelation between the stages. Some examples in Sections 3 and 8, as well as in the Barberà et al.'s paper reveal some features of such interrelations.

In this paper we study pure-strategy equilibrium profiles. As game theorists, we should ask if allowing mixed strategies, or even correlated-strategy profiles would yield more attractive profiles. One may think that this is an unnecessary luxury, as people never employ mixed strategies when deciding how to vote. Nevertheless, we produced an example in the Barberà et al.'s paper in which a correlated-strategy profile made much more sense than the others. Perhaps, if other examples could be constructed, voters in the real world would be inclined to consider such strategies as options.

Sections 6–9 deal with trembling-hand pure-strategy perfect equilibrium profiles. This required a deeper analysis and the results are not complete. For example, we were only able to provide sufficient conditions for a strategy-profile to be perfect and even then, we had to restrict the utilities to be additive and generic. The perfect profiles that we found are interesting as they show clearly what type of added safety they provide. Furthermore, they are instrumental in proving that under the single-stage (EA) protocol, a pure-strategy perfect-equilibrium profile always exists when the game is generic and the utilities are additive (Theorem 8.5). The strategy profile that accomplishes this is attractive as it can be stated in a language that may appeal to a layman. In the (AE) protocol it may well be the case that pure-strategy prefect-equilibrium profiles do not exist. The (EA) protocol has another interesting advantage: If staying out of the society is always not desirable and the players insist on pure-strategy perfect-equilibrium profiles, then any agent that survives any such perfect-equilibrium profile under the (EA) protocol will prefer the corresponding outcome to an outcome obtained in any pure-strategy perfect-equilibrium profile under either the (SIM) or the (AE) protocol (Section 9).

The various examples in the paper are an important and integral part of the present study. They point to features that are not covered by the general propositions and theorems. Thus, in Section 3 we show that it may be beneficial to expel a common friend or keep a common enemy. In Section 8 we show that voting sincerely at each stage in not the only perfect-equilibrium profile under the (A) protocol. There may be cases in which several perfect-equilibria exist, one dominating another. This is instrumental in constructing a perfect-equilibrium in which the punishment off the equilibrium path is a deviation to another dominated perfect profile.

One main lesson that can be drawn from some of the examples is the fact that finding an equilibrium profile is not an end in itself. Some equilibrium profiles and outcomes, even those involving perfect profiles, simply do not make sense (e.g., Example 8.7). Thus, if one wants to make a recommendation in a real case, one must examine all equilibrium outcomes before deciding which one to recommend.

An important line of research is to restrict the utilities to a particular class. If enough requirements are placed on the utility functions then the set of equilibrium profiles diminishes, perhaps reduces to a single attractive equilibrium profile. We can then "plot" the stream of members of the society, study its properties under various agenda and see where the society is heading or at least determine interesting properties of such a stream. This task can be achieved in two ways:

1. Imposing restrictions in order to obtain theorems. This may be an easy, yet an important mathematical contribution.

2. Making restrictions to fit a real case. For example, we pointed out that the EU operates under rules similar to those in the present model. Can we

estimate the utility of each country for the various compositions of a EU? Can we then single out particular equilibrium profiles that make sense economically and narrow down the set of reasonable equilibrium profiles?

We started this paper by a motivation drawn from biology. We would like to end this overview with a suggestion, also motivated from biology: The study of stable evolutionary processes in biology is greatly enhanced by dynamic models such as the replicator dynamics. We suggest that a dynamic model be developed that takes into account social and economic forces that explain how agents may be led to a particular equilibrium outcome, or to cycling among several ones.

References

1. Barberà S, Maschler M, and Shalev J (2001) Voting for Voters: A Model of Electoral Evolution, *Games and Economic Behavior* **37**:40–78
2. Kohlberg E and Mertens JF (1986) On the stability of equilibria, *Econometrica* **54**:1003–1039
3. Kuhn HW (1953) Extensive games and the problem of information, Contributions to Game Theory II (Kuhn HW and Tucker W, eds.), Annals of Mathematics Studies, 28, Princeton University Press, Princeton, NJ, pp. 193–216
4. Myerson RB (1991) Game Theory, Analysis of Conflict, Harvard University Press, Cambridge, Massachusetts, London
5. Ouchi GW (1981) Theory Z: How American Business Can Meet the Japanese Challenge, Addison Wesley, Reading, Massachusetts
6. Selten R (1975) Reexamination of the perfectness concept for equilibrium points in extensive games, *Int. J. Game Theory* **4**:25–55