Voting for voters : a model of electoral evolution

Author(s): Barberà, Salvador
Date 1998
Type Working Paper
Series Estudios = Working papers / Instituto Juan March de Estudios e Investigaciones, Centro de Estudios Avanzados en Ciencias Sociales 1998/115
City: Madrid
Publisher: Centro de Estudios Avanzados en Ciencias Sociales

VOTING FOR VOTERS: A MODEL OF ELECTORAL EVOLUTION

Salvador Barberà

Estudio/Working Paper 1998/115
June 1998

Salvador Barberà is Professor of Economics at the Universidad Autónoma de Barcelona. This paper, co-authored with Professor M. Maschler of the Hebrew University of Jerusalem and Professor J. Shalev of the Université Catholique de Louvain, is based on a series of four public conferences that Professor Barberà gave at the Juan March Institute on 11, 13, 18 and 20 March 1997. The conferences were entitled, respectively, "Estrategia y elección social"; "La agregación de preferencias"; "Problemas dinámicos en teoría de la elección social"; and "Modelos formales de justicia distributiva".

ABSTRACT

We model the decision problems faced by the members of societies whose new members are determined by vote. We adopt a number of simplifying assumptions: the founders and the candidates are fixed; the society operates for k periods and holds elections at the beginning of each period; one vote is sufficient for admission, and voters can support as many candidates as they wish; voters assess the value of the streams of agents with whom they share the society, while they belong to it. In spite of these simplifications, we show that interesting strategic behavior is implied by the dynamic structure of the problem: the vote for friends may be postponed, and it may be advantageous to vote for enemies. We discuss the existence of different types of equilibria in pure strategies and point out interesting equilibria in mixed strategies.

1. Introduction

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

Clubs and learned societies are examples of human groups that fit our description exactly. Others may only meet part of the features we require here. For example, parliaments are elected according to well specified rules, but their size is fixed, while our focus will be on the forces that determine the growth or the stagnation of groups. In other cases, entry and exit are the result of informal procedures, whose description as voting rules might be too simplistic even as an approximation. Our model, thus, only applies to a restricted set of societies.

Election rules are social constructs: they may come from an agreement among different founders, they may reflect the will of a unique founder or they may be the result of successive amendments, but they must be set purposely. Once the rules for election to a society are set, participants in the election are bound to engage in strategic considerations that involve non-myopic behavior. In particular, voters cannot overlook the fact that newly elected members will become voters in later elections: this may lead to postpone the election of individually attractive candidates who might vote in unattractive ways, or to accelerate the election of a poor candidate whose vote is needed. We are interested in the evolution of groups which results from considerations of this type being made by rational agents under well specified voting rules. The features we have emphasized should make it clear that electoral evolution is the result of nonmyopic behavior which is quite typical to human societies.

Since this paper is a first attempt at modeling such facts, we allow ourselves some strong simplifying assumptions. The founders and the rules of election of a society are fixed in advance (we don't explain why they join to create the society or why they agree on these rules). The candidates to enter the society are fixed as well (we don't explain why they don't try to create other societies, or any other process by which eligible candidates could change from election to election). We assume that nobody leaves the society once admitted (thus concentrating on entry and not on exit). We study finite horizon situations where members of the society know at all times when it will be dissolved and voting takes place at a finite number of periods (when in fact many societies operate under an uncertain horizon). We assume a specific voting method, whereby each member can vote for as many candidates as he wishes, and it is enough for a candidate to receive a vote in order to be admitted (this is the method of 'voting by quota one'; many others are worth considering). We postulate that agents' preferences are defined over streams of members in the society, and that they are additive across stages. Under these assumptions, we provide theorems on the existence and the characteristics of different types of equilibria of the games generated in such dynamic voting contexts. Although clearly restricted by our assumptions, these results bear witness to the abundance of possibilities within our model.

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

Our closest reference is "Voting by Committees", by Barberà, Sonnenschein and Zhou [1991], where the question of electing members for a society is treated as a one period problem. That paper characterizes the set of all strategy-proof mechanisms respecting the sovereignty of voters when their preferences over sets of candidates satisfy one of two alternative restrictions, called additivity or separability: they are the methods of voting by committees. We shall not describe the general class, but simply say that they contain an interesting subclass of methods, which in addition to the preceding properties will also respect anonymity and neutrality; i.e., will treat all voters and all candidates alike. This subclass consists of the methods based on voting by quota: each agent can vote for as many candidates as he wishes, and all candidates who get at least q votes are elected, where q is fixed a priori. Our main interest in the present paper is on phenomena that only arise when the society's horizon is greater than one period, and this is why we have chosen to work with multiperiod models whose one period version takes the form of voting by quota. Since these methods are strategy-proof in their one shot version, we can be sure that whatever strategic behavior arises when several periods are considered must have a dynamic source.

As already mentioned, our ambition is to study the evolution of societies who resort to voting as a means to include or to exclude members. It has both a normative and a positive viewpoint. Many interesting questions come to mind. Just to mention one topic on the descriptive side, we would like to understand why some societies maintain their defining features along their history, while others change so much that their own founders would not recognize them. However, our ambition must be tempered by the fact that the game theoretic analysis quickly becomes complex and presents several alternative routes. Accordingly, the paper contains examples, which point at the complexities of the analysis, as well as technical results on how to solve for equilibria and what types of equilibria to look for. It is structured as follows. In Section 2 we present the model, based on a gallery of assumptions. Section 3 contains examples. These examples show that the simplicity of the one period model is immediately lost if we have several periods. They also prove that some counterintuitive phenomena, like strategic voting for enemies, can occur if the number of periods is not too small. They also indicate that it will be worth analyzing not one but several solution concepts, because each one of them can provide some insight on the phenomena we try to model. One example shows that, although we concentrate on pure-strategy equilibria, the use of mixed strategies, or even correlated strategies, may be most reasonable in some cases. In Section 4 we analyze subgame-perfect equilibria and 'quasi-strong equilibria',[1] and we discuss the fact that the streams of members for a society can be attained in equilibrium, given the rules, through different distributions of the individual votes. In this section we also look for Pareto-undominated equilibria. Unfortunately, Pareto undominated equilibrium profiles are often not perfect equilibria. Thus, the members may wish to adopt less profitable outcomes in order to gain the stability that a perfect equilibrium yields. Section 5 is devoted to the existence of perfect equilibria in pure strategies: we provide a sufficient condition under which there will exist such equilibria, and examples showing that the condition is not necessary. We also show by examples that quite natural cases exist in which perfect equilibrium profiles can only be reached by using mixed strategies. In this paper agents are satisfied in employing only history-independent strategies (which we formally define in Section 2). The merit and the limitations of these strategies are discussed in the Appendix.

[1]i.e., equilibria that have the additional property that no deviator can benefit if the set of deviators does not include the set of all voters at the start of a deviation.

2. The model

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

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

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

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

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

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

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

The decision on how to vote at each stage, that every voter i faces, should take into consideration the priorities that each agent has over the various streams.[2]

[2]One can think of complicated priorities on events that may even be concealed. For example, a voter might not like

As mentioned in the introduction, we make many simplifying assumptions in order to render the model simple and yet still capture some dynamic aspects of the workings of the voting scheme. In fact, we suppress many aspects in order not to 'blur' the purely dynamic issues. Obviously, other, more complicated and more realistic models should be studied. As we show, even the present simple model possesses enough intricacies to render the analysis interesting.

Some simplifying assumptions.

1. FIXED POPULATION. *We assume that the population is finite and fixed and includes the nonempty set of the original founders* F^0. Therefore, we can denote the set of *agents* by *N*. $N \setminus F^0$ is called *the set of the original candidates* and is denoted by C^0. Similarly, we write C^t for $N \setminus F^t$. Members of C^{t-1} are the candidates from whom the voters F^{t-1} can choose at stage *t*.
2. NO FIRING. *We assume that an elected candidate will stay in the society all the time.* There are no provisions to fire him. Later we shall add an assumption that guarantees that no agent will want to resign.
3. 1-QUOTA VOTING. The rule for electing a candidate into the society is simple: every *voter can bring any number of candidates into the society at any stage, simply by casting* a *vote for them at the beginning of that stage.* This rule is known as *voting by quota 1.*
4. STREAMS OF MEMBERS ARE ALL THAT MATTER. We assume that *each agent cares only about the streams of members in the society* and does not care, for example, about who voted for each member. Thus, his priorities are functions of the streams $\mathcal{F}$.
5. COMMON HISTORIES. We assume that *at each stage the elected candidates* are *known to everyone.* Thus, for every agent *i* the relevant $(t-1)$-stage histories are the same;[3] namely,

an agent *j,* if he knew that agent *p* also voted for *j,* but otherwise he might have loved to have *j* in the society. Perhaps he does not even know who elected *j.* We shall not consider such complications in this paper.

[3]Actually, if ballots are not secrets, histories may be more complicated than simply past stream of members. They

subsequences of the streams terminating at F^{t-1}. These will be denoted h^t, $t = 1, 2, \ldots, k$.

We now have all the ingredients to convert the above setup into a *game form*: The set of *players* is N, the *pure strategies* available to player i are choices of sets that specify at each stage t the candidates that he votes for at that stage as a function of the history at that stage.

Most of the time we shall restrict ourselves to pure strategies which are *history-independent*; namely, strategies that depend only on the stage number[4] and on the population of voters at that stage[5] and neither on the precise sequence of votes that led to that stage,[6] nor even on the sets of members that were elected in previous stages.[7]

With this restriction, we can denote a pure strategy for agent i by $\sigma_i := (\sigma_i^1, \sigma_i^2, \ldots, \sigma_i^k)$, where the *stage strategy* σ_i^t is a function from the domain $\{1, 2, \ldots, k\} \times 2^N$ into 2^N. Here, $\sigma_i^t(t, F^{t-1})$ is interpreted as the set of votes that i casts at stage t when F^{t-1} is the current set of voters.[8]

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

may include information such as who voted for whom, and when. In this paper we shall not employ such histories except when we show under what conditions one can do without them (Theorems 6.1 and 6.3).

[4]Or rather on the number of stages left.

[5]Thus, also on the set of candidates left.

[6]As mentioned previously, often this sequence may be unknown, or partly unknown to the agents.

[7]In the appendix, we shall discuss the merits and the limitations of this assumption. In particular, we prove there that any equilibrium outcome that can be obtained by pure strategies can also be obtained by history-independent pure strategies. We shall also show that history-independent pure-strategy equilibria are robust; namely, remain equilibria even if we allow responses that employ mixed and history-dependent strategies. However, when mixed strategies are feasible, the history-dependent set of equilibrium outcomes is definitely richer.

[8]Or a probability distribution on such sets, if we are interested in behavioral strategies. We find it useful to denote by σ_i^t the strategy of a player i in the subgame that starts at stage t. It has its own players (the voters at this stage), which may be different from the players in a different stage game. Here, the superscript t is just a part of the name of the strategy. σ_i^t is a function of two variables, one of which is t, because the same player may act differently at different stages even if he faces the same sets of voters and candidates. Moreover, $\sigma_i^t(\tau, F^{\tau-1})$, $\tau \geq t$ would be his action when he reaches stage τ and the set of voters is $F^{\tau-1}$.

given by

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

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

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

6. KNOWN UTILITIES. We assume that *the priorities of agent i are given by complete and transitive binary relations on the set of streams* and therefore they can be represented by a *utility function* u_i. Later, when we deal with mixed strategies, we shall assume that *these utilities are, in fact, Von Neumann Morgenstern utilities.*[9]

The last assumption is not needed for a great part of the paper. We assume that all utilities are known to every agent and, in fact, are common knowledge.

We want the utilities to express the desire that each agent, above all, wishes to be in the society. We normalize his utility for staying alone in the society to be zero.[10]

Accordingly, we state:

7. STRONG PREFERENCE FOR PARTICIPATION. NORMALIZATION. *Serving in the society is preferable than staying out in all circumstances.* Staying alone in the society has a zero utility.

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

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

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

We now present, several *possible simplifying assumptions on the utilities,* ranging from simple to more complicated considerations. Some of them will be employed in the examples of the next section, to illustrate some of the issues. Others will be needed for the proofs.

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

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

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

We then assume that *each agent wishes to spend as much time as possible with friends and as little time as possible with enemies and that this is all he cares for.* If the stages are equally spaced in time, it then makes sense to denote by 1 the utility of having a friend in the committee for one stage and by $(-1 - \epsilon)$ – the utility of having an enemy for one stage, where ϵ is a small positive number, added to break ties.[11]

---

[11]We decided to require a positive ϵ in order to express the fact that, other things being equal, the members would

If the voters are not sophisticated and only durations of time spent with 'friends' and 'enemies' matter, it makes sense to choose additive utilities. We summarize the above formally:

8a. PURE FRIENDSHIP AND ENMITY. *The utility for a stream of members, given by (2.1) for an agent who succeeds in entering the society is given by*

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

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

In a more sophisticated model we can still assume that whether or not to vote for a person is decided on purely personal grounds; namely, only on the merits of the person and not, e.g., on who is already in the society, but now we let agents also take into consideration *how much* they like/dislike each person.

Individual considerations may be quite complicated: a voter may like one person and dislike another. He may want a person in the society, because he thinks that his views should be heard. He may want a person in to balance an extreme stand of a founder, etc. Here we make the strong assumption that whatever these considerations are, they can be summed up by *each agent providing each individual with a time-independent and society-independent "weight function", so that the sum of the weights reflects the utility of the voter for one stage.*

Naturally the weights still allow us to distinguish between friends and enemies. Friends will be agents with positive weight and enemies -- with negative weights. If the weight is zero, we can call him neutral to the voter.

like to have a society with as few conflicts as possible: it is worse to have a friend and an enemy for a certain period of time than to have none of them for that period.

We couple the above assumption with the idea that *a voter wants to spend as much time as possible with friends and as little time as possible with enemies.* Together, the above brings about the next simplifying assumption:

8b. FRIENDS AND ENEMIES. ADDITIVITY WITHIN EACH STAGE AND ACROSS STAGES. *Every agent i has a weight function $w_i : N \to \Re$. His utility $u_i(\mathcal{F})$ for a stream of members $\mathcal{F}$ serving in the society is given by:*

$$u_i(\mathcal{F}) = \sum_{\{t\geq 1:\ i\in F^t\}} \sum_{a\in F^t} w_i(a). \tag{2.3}$$

Thus, $w_i(a)$ can be interpreted as the utility that i accumulates from spending one stage in the society together with a.

How a weight function w_i is determined in real life is hard to tell. Presumably it reflects player i's opinion on the importance that the agent belongs to the society. As indicated previously, a friend may carry a high weight and yet not be invited to join.

On a higher level of sophistication we consider a model in which not only individuals but also groups matter. Thus, we now assume only that agents have priorities over the various groups that may compose the society and these priorities need not be sums of weights for individual members. We still assume additivity across stages. Formally:

8c. ADDITIVITY ACROSS STAGES. *Each member i has a utility function $u_i : \{1, 2, \ldots, k\} \times 2^N \to \Re$, that may depend on t, so that his utility for a stream[12] $\mathcal{F} = \{F_1, F_2, \ldots, F_k\}$ is*

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

[12] By abusing slightly the language, we use the same expression u_i to denote both 'utility gain for a stage' (or 'utility per stage') and 'utility of a stream'. This is justified because of assumption (2.4) below.

Again, additivity across stages makes sense if the stages are equally spaced in time. Note that now we no longer assume 'time independence': We allow that the same set of members adds a different utility per stage to a player if it appears at a different stage. This may be the case, e.g., if some of the agents are experts, whose services are important only at a late stage in the life of the society.

To complete the descriptions above we make a last assumption:

9. COMMON KNOWLEDGE. *All utilities* as *well as all the descriptions above are common knowledge.*

Who are the players? We have set up the society protocol and we have converted it into a game. Clearly, the way we formulated it, the set of players is *N.* Yet, we can regard the situation as *a sequence of several games,* one starting at each stage, with different players, where the players at each stage t are the set of voters F^{t-1} and the other agents are considered extraneous entities. Indeed, agents do not really become players until they enter the society. The only votes that count are those of agents who are members by that stage. They create the continuation and it is their interest that matters.[13] Thus, if we want to talk about refinements of equilibria, we sometimes prefer to make them relative to the set of voters at each stage. Accordingly, we shall employ the following definition:

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

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

The concept of 'strong equilibrium' was introduced in Aumann [1959]. We shall encounter in the next section games for which strong equilibria do not exist. Nevertheless, we shall show in Section 4 that it is often possible to achieve ‘quasi-strong equilibria’ as defined below:

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

This concept is in a sense weaker than Aumann’s, because it does not allow for deviations involving all the voters. In another sense it is stronger, because it tells us that no voter can gain even if others lose.

3. Some interesting simple examples

A universal equilibrium profile. One equilibrium profile always exists in pure strategies:[14]

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

If there is only one founder he chooses that stream that maximizes his utility given that as soon as there are at least two voters, each will vote for every candidate. For example, under pure friendship and enmity (Assumption 8a),[15] he will vote for all his friends in the first stage, if he has more friends than enemies (and every candidate will be brought in at the second stage) and if the number of friends does not exceed the number of enemies he will vote for nobody until the last stage, whereupon he will bring all his friends.

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

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

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

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

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

Complications can occur if additivity does not prevail, as the following example shows:[16]

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

$$u_1(\emptyset) = 2, \quad u_1(a) = 3, \quad u_1(b) = 1, \quad u_1(ab) = 0,$$

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

POSSIBLE SCENARIO: Founder 1 likes to stay alone.[17] He thinks it is a good idea to bring a to the society and it is a bad idea to bring b. It is a disaster to bring both, because the two will fight all the time. Founder 2 does not like a's views. He somewhat prefers b, but would above all like to stay alone. Bringing both is a 'compromise' between the previous two undesirable events.

---

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

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

The pure-strategy equilibrium points are (b,b), (a,ab) and (ab,ab). None of them is perfect — they are all eliminated by weak domination. The only perfect equilibrium is mixed, in which Founder 1 votes for $\emptyset$ and a with equal probabilities and Founder 2 votes for $\emptyset$ and b with equal probabilities.

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

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

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

Since friendship here is vacuously transitive, the following is a perfect equilibrium profile: *a votes for c at both stages and b votes for d at both stages, regardless of the histories.* Nevertheless, there is another equilibrium profile that is preferred by *both* players: *players a and b bring their friends only in the second stage and if anyone deviates in the first stage, both a and b invite all the remaining candidates in the second stage.* In this strategy each founder ties the hands of the other founder: "If you do not abide, we shall punish you by bringing in all the enemies." This is even a subgame-perfect equilibrium and sequentially-Pareto-undominated,[18] but it is *not* perfect: Whatever the action of the other person, voting only for one's friend in the last stage is never worse and in some cases better than the prescribed action.

[18]Another variant, in which the deviator is punished only by the other person, in case of deviation, is *not* subgame-perfect but is more convincing: why should the deviator agree, and abide by punishing himself? This is another manifestation of the known dilemma: Why should one trust a promise of a person, who already proved that he does not keep his promises, because he deviated in the first stage.

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

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

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

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

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

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

The utilities (not including utilities for time spent with the original founders and ignoring multiples of ϵ) are: $u_1 = -14, u_2 = -10, u_3 = -2, u_g = -10, u_e = u_f = u_a = u_b = u_c = -12, u_d = -10, u_p = u_q = u_r = u_s = -10$.

It can be checked that this is indeed an equilibrium profile and, moreover, it is perfect.[19]

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

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

When many common enemies exist. We have seen in the previous example how a punishment can force an equilibrium. In fact, if there are enough common enemies, then *any agreement* between the current founders, at any stage other than the last, can be enforced by a strategy that stipulates that out of the agreement all voters will vote for all common enemies as soon as they recognize that they are off the equilibrium path. This is even subgame-perfect.

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

We keep the above in mind but we wish to make the following two observations: (1) In real life one

[19] Any "tremble" can be observed only in the last stage when it is still to one's advantage to bring all his friends.

can usually extend the set of candidates so as to include as many common enemies as one 'wishes'. (2) Nevertheless, a threat to bring these common enemies is often not credible as a general procedure. It often would be considered unthinkable, because it would undermine the very foundations upon which the society rests. Thus, although such threats may be feasible, often they are not viable, which brings us again to the recognition that a model does not usually capture all the intricacies of a real situation.

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

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

The founder would like to bring all his friends, but if he simply does so at the first stage then each b_i will bring c_i in the next stage. This is because the b_i's will not fear[20] that c_i will bring d before the last stage, knowing that if c_i does so, d will bring e. To prevent this from happening, the founder can vote for e in the first stage. A complete strategy profile is this:

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

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

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

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

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

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

$$\sigma^3_{b_i} = \begin{cases} \{c_i\}, & \text{if } d \in F^2, \\ \emptyset, & \text{otherwise,} \end{cases} \qquad i \in \{1,\dots,5\};$$

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

$$\sigma^t_a = \emptyset, \quad t = 2,3,4;$$

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

One can verify that this is indeed an equilibrium profile.

Example 3.5. The game of chicken. In this example, $F^0 = \{1,2\}, C^0 = \{x_1,x_2,y_1,y_2\}$, $k = 3$. Founder 1 likes only x_1, who likes only y_1. Founder 2 likes only x_2, who likes only y_2. Agents y_1 and y_2 like only each other.

Skipping formalities, each founder can essentially either choose his friend in the first stage, or refrain from doing so. (He strictly loses by voting for an enemy at this stage.) Unfortunately, if player 1 votes for his friend at the first stage, player 2 will lose if he too votes for his own friend. The reason is that in this case it is clear that both y_1 and y_2 will be present in stage 3, so there will be no reason for both x_1 and x_2 to refrain[21] from voting for their friends in stage 2. These friends are enemies of the founders. Putting together the relevant information and ignoring ϵ, we get the following payoff as functions of the choices in the first stage:

	$\emptyset$	x_2
$\emptyset$	0, 0	−3, 1
x_1	1, −3	−4, −4

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

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

Even more sensible for the players is to decide by a flip of an unbiased coin who will bring his friend to the society. The expected payoff will then be $(-1, -1)$.

This example comes to show that mixed strategies should not be ignored.

4. Common voting and partial common voting

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

We then proceed to characterize and, at least theoretically, construct all the equilibrium streams, and therefore all equilibrium outcomes that can be achieved by pure strategies. We also indicate where to look when we want to get all the sequentially-Pareto-undominated equilibrium streams, as well as all the subgame-perfect streams.

In the last part of this section we provide interesting procedures that produce equilibrium profiles that only 'partially' employ common voting, or even some in which the voters vote for distinct sets.

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

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

Proof. Indeed, had he preferred so, he could benefit by adding these members in his vote, contrary to the fact that the profile was an equilibrium one. ∎

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

Consider a game Γ that represents a voting scheme, as described in Section 2. Every subgame of this game is itself a game that can result from a voting scheme. Only the set of founders and candidates and the number of stages differ. This enables us to work by induction. Suppose that the play in all the proper subgames is *known and fixed*. One can then construct a one-stage game Γ^1 whose tree is the subtree for the first stage of Γ and whose endpoint payoff vectors are calculated from those of Γ: The payoff vector at an endpoint of Γ^1 is the payoff vector that results in Γ from reaching the corresponding node in Γ and continuing along the fixed path of the subgame attached to that node.[22] Note that the players in Γ^1 are the original founders, and each one of them has exactly one information set. In fact, Γ^1 is a game played simultaneously by all the founders. Note also that endpoints of Γ^1 at which the same candidates were elected have the same payoff vectors if the continuations are history-independent.

Let σ be an arbitrary equilibrium profile in Γ. Its first stage σ^1, regarded as a strategy profile in Γ^1, is an equilibrium profile for this game.[23] Indeed, if a player can benefit by deviation in Γ^1, he could also benefit in Γ by choosing the same deviation in the first stage.

We can now construct another equilibrium profile $\bar{\sigma}^1$ for Γ^1, by *instructing every original founder to vote for the union of all the votes of the founders under* σ; i.e.,

$$\bar{\sigma}_i^1 = \cup_{j \in F^0} \sigma_j^1, \qquad \forall i \in F^0. \tag{4.1}$$

[22]This construction can be extended also to cases when the continuations are neither pure nor history-independent. Most results in this section, however, will not be true in such cases.
[23]This statement is true also if the profile is not history-independent, and not pure.

We call $\bar{\sigma}^1$ the *first stage profile generated from σ^1 by common voting.* Analogously, we say that $\bar{\sigma}$ is a *strategy profile generated from a history-independent strategy profile σ by common voting*, if each stage, on and off the equilibrium path, is generated from the corresponding stage of σ by common voting. Strategy $\bar{\sigma}$ is well defined and σ and $\bar{\sigma}$ yield the same stream of members. The same stream of members for σ and $\bar{\sigma}$ occurs also in all the subtrees off the equilibrium path; therefore, *$\bar{\sigma}$ is an equilibrium profile yielding the same payoffs as σ and it is subgame-perfect if σ is subgame-perfect.*

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

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

One should be a bit careful when one tries to generalize Proposition 4.2 to multi-stage games: At future stages 'new' players may enter the game and one has to take into account possible agreements involving them. Consider the following:

Example 4.3. Let $F^0 = \{1,2\}, C^0 = \{a,b\}$. Under pure friendship and enmity (Assumption 8a), agents 1 and 2 like agent a. Agent a likes agent b. For all other pairs (i,j), j is an enemy of i.

$k = 2$. The following strategy profile is subgame-perfect:

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

It is already in common voting for the original founders: they always vote the same way. Nevertheless, this profile is not immune to deviation involving a proper subset of the founders: Agents 1 and a can deviate by 1 voting for a already in the first stage and a voting for $\emptyset$ at the second stage. By this deviation agent 1 gains and agent a also gains, because he becomes elected.[24]

This example indicates that one should require that common voting involves all members that enter the game on and off the equilibrium path. Indeed, if we augment the above example and request that both 1 and 2 vote for b at the second stage, when a is elected in the first stage, then no profitable deviation can take place by a proper subset of the founders. For example, it will do no good that a will refrain from voting b, because founder 2 will still vote for b.

We keep in mind that when we talk about a deviation of a set of voters we allow all kinds of agreements involving future candidates. Candidates off the equilibrium path will agree to anything, because they prefer to be in the society under all circumstances (Assumption 7). It stands to reason to request that candidates along the equilibrium path should not lose when we claim a profitable deviation, although this is not important for the next theorem.

Theorem 4.4. *Let σ be a pure-strategy history-independent subgame-perfect equilibrium profile for a game Γ representing a voting scheme defined on preferences satisfying additivity across stages (Assumption 8c). Let $\bar{\sigma}$ be the profile generated from σ by common voting. Then $\bar{\sigma}$ is an equilibrium profile yielding the same stream of members and therefore the same payoff vectors as σ. Moreover, it is a quasi-strong equilibrium profile.*

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

Proof. The strategy profile $\bar{\sigma}$ yields the same stream of members as σ, so the payoffs are the same for every player. Moreover, it is a subgame-perfect equilibrium profile, because if one could benefit by deviation in a subgame, he could benefit by the same deviation in σ. We shall prove by induction on the number q of stages left since a possible deviation started to occur, that no proper subset of F^{k-q} can deviate in such a way that at least one deviator gains. Proposition 4.2 establishes this fact for $q = 1$. Suppose that this was verified for subgames with $q - 1$ stages and we are now facing a deviation starting in a subgame $\hat{\Gamma}$ having q stages. Let τ be a deviation from $\bar{\sigma}$, such that the set of deviators does not include all the voters F^{k-q}. Denote by Γ^* the first stage of $\hat{\Gamma}$ with payoff vectors at the endpoints calculated on the assumption that the continuation was as dictated by $\bar{\sigma}$ in Γ. Denote by Γ^{**} the first stage of $\hat{\Gamma}$ with the payoff vectors at the endpoints calculated on the assumption that the continuation was as dictated by τ in Γ.[25]

Denote by A the endpoint of Γ^*, that is reached if the relevant part of $\bar{\sigma}$ is played. Denote by B the endpoint of Γ^* that is reached if the relevant part of τ is played. Denote by A and B the corresponding endpoints in Γ^{**}. A and B correspond also to nodes of Γ, reached under $\bar{\sigma}$ and τ at the end of stage $k - q + 1$. We denote these nodes also A and B.

Let Δ be the subgame of Γ starting at B. It has $q - 1$ stages. We can regard τ as performed in two steps.

Step 1: The play changes from A to B and then continues as dictated by $\bar{\sigma}$.

The resulting payoff vector would then be the payoff vector at endpoint B of Γ^*, whereas the payoff vector at endpoint A of Γ^* is the payoff vector in Γ if the play is $\bar{\sigma}$.

Step 2: Perform further modification in Δ, if this is dictated by τ.

The proof will conclude if we show that no deviator gains either in Step 1 or in Step 2.

[25] We are creating these 1-stage games in a way similar to the way we created Γ^1 at the beginning of this section.

Indeed, the claim for Step 1 means that no deviator gains in Γ^* by moving from A to B. The claim for Step 2 means that no deviator gained in Δ by switching from $\bar{\sigma}$ to τ in this subgame. This implies that no deviator gained in passing from B in Γ^* to B in Γ^{**}, because these payoff vectors differ from those of Δ by the same constant payoff vector that was accumulated until B was reached.[26] Thus, ultimately, no violator gained in Γ by deviating from $\bar{\sigma}$ to τ.

The claim for Step 1 follows from Proposition 4.2, because $\bar{\sigma}$ induces a pure-strategy common-voting profile in Γ^*. The claim for Step 2 is simply the induction hypothesis. ∎

We have shown that all pure-strategy history-independent equilibrium outcomes can be generated by common voting.[27] The natural question that now comes to mind is how to characterize all streams that constitute such outcomes. Proposition 4.5 and Theorem 4.6 provide an answer.

Proposition 4.5. *Assume that there are at least two founders in a 1-stage game Γ^1. A set S of candidates chosen can result from a pure-strategy equilibrium profile iff S has the property that no founder would prefer to add members to S.*[28]

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

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

[26] We are invoking additivity across stages.
[27] By Theorem 6.1 we can drop the 'history-independent' requirement from this statement.
[28] Such an S always exists, for example $S = C^0$.

Theorem 4.6. *Consider* a game *Γ, representing* a *voting* scheme, whose utilities *obey additivity across stages (Assumption 8c). Work backwards from the final stages constructing strategy profiles, analogous to the one in Proposition 4.5, taking care to choose the same S, whenever the same voters appear at different starts of a* same stage. Continue, as long as there are at least *two voters. If you encounter* a *node with only one voter choose a path leading to a maximal payoff to this voter. The above construction results in a pure-strategy history-independent subgame-perfect equilibrium. All pure-strategy history-independent subgame-perfect equilibrium streams are obtained if one exhausts all possibilities of the above construction.*

Proof. Start from the last subgames and work backwards by common voting. At each stage you find yourself with a 1-stage game with fixed history-independent subgame-perfect continuations, for which Proposition 4.5 can be applied. This shows that you will thus construct a pure-strategy history-independent subgame-perfect strategy profile for the entire game. By Theorem 4.4, all pure-strategy streams will be reached if one exhausts all possibilities. ■

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

If we were only interested in equilibrium outcomes we could stop here. But we are also interested in other equilibrium profiles that lead to such outcomes, in particular those obtained by pure strategies. We shall close this section by producing a wider class of equilibrium profiles. These extend the common-voting class in that they involve only partial common voting, or even no common voting at all. This last type of profile will play an important role in Section 5, when we deal with perfect equilibria.

Proposition 4.7. *Let Γ^1 result from a game Γ having known and fixed play at all proper subgames. Assume that Γ has at least two founders. Let S be a set of candidates from C^0, having the property that, if elected, no original founder will prefer to add players to S. For each founder i, choose a set P_i, contained in S, that is a best response to*[29] *$S \setminus P_i$. Let $C = S \setminus \cup_{j \in F^0} P_j$. Finally, let $V_i = P_i \cup C$. Under these conditions, $\{V_i : i \in F^0\}$ is an equilibrium profile for Γ^1.*

The proof requires two lemmas:

Lemma 4.8. *Let Γ^1 be a first stage game, as previously described. Let P_i be a best response of founder i against $S \setminus P_i$, where S is an arbitrary given set of candidates from C^0 containing P_i. If $Q \subseteq S \setminus P_i$, then $P_i \cup Q$ is also a best response of i to $S \setminus P_i$*

Proof. Q is covered anyhow by $S \setminus P_i$, so it makes no difference whether i includes Q in his vote, or not. ∎

Lemma 4.9. *Let P_i be a best response of founder i, playing Γ^1, against $S \setminus P_i$, where S is an arbitrary set of candidates containing P_i. If $R \subseteq P_i$ then $P_i \setminus R$ is a best response of i to $(S \setminus P_i) \cup R$.*

Proof. Voting $P_i \setminus R$ against $(S \setminus P_i) \cup R$, would yield player i the utility gained from S being

[29] Such a set always exists; for example $\emptyset$.

elected. If voting for another set, Q, would yield him a higher utility, then voting $Q \cup R$ would be a better response to $S \setminus P_i$ than voting P_i, because $(Q \cup R) \cup (S \setminus P_i) = Q \cup (R \cup (S \setminus P_i))$. ■

Proof of Proposition 4.7. P_i is a best response of i against $S \setminus P_i$; therefore, V_i is a best response of i against $S \setminus P_i = (S \setminus P_i) \cap (\cup_{j \in F^0 \setminus \{i\}} V_j)$ (Lemma 4.8). By Lemma 4.9, $(C \cup P_i) \setminus \cup_{j \in F^0 \setminus \{i\}} P_j$ is a best response of i against $\cup_{j \in F^0 \setminus \{i\}} V_j$. Invoking Lemma 4.8 once more, we find that V_i is a best response of i against $\cup_{j \in F^0 \setminus \{i\}} V_j$. ■

Theorem 4.10. *Consider a game Γ, representing a voting scheme, whose utilities obey additivity across stages (Assumption 8c). If a construction analogous to the one in Proposition 4.7 is done at every start of a subgame of Γ, starting from the final stages and working backwards as long as there are at least two voters, and choosing the move leading to a maximal payoff if encountering one voter, then, the resulting profile constitutes a pure-strategy subgame-perfect equilibrium.* ■

Proof. We may assume that $|F^0| \geq 2$. Let σ be a strategy profile as described in the theorem. Let τ_i be a deviation made by player i, starting at a certain subgame $\hat{\Gamma}$. If the deviation started at the last stage then player i could not have benefitted from it either because the equilibrium path was not affected by the deviation, or, by Proposition 4.7, if it was.

Assume by induction that a player cannot gain by deviating alone in any deviation whose length is at most $q - 1$. Let $\hat{\Gamma}$ be a q-stage subgame. As in the beginning of this section[30] construct 1-stage games Γ^* and Γ^{**}, for the subgame $\hat{\Gamma}$, derived from σ and from $\tau := (\sigma_{-i}, \tau_i)$, respectively. These games have the same trees, but their payoff vectors may be different, due to the different continuation by τ. Let A be the endpoint of these 1-stage games, as well as $\hat{\Gamma}$, reached via σ and let B be the endpoint of these games reached via τ.

[30] When we constructed Γ^1.

Denote by Δ the subgame of $\hat{\Gamma}$ starting at B. By the induction hypothesis, player i cannot gain from deviating to τ in the game Δ; therefore, his payoff at B in Γ^{*} is not smaller than his payoff at B in Γ^{**}, since a fixed "income" was accumulated in both cases; namely, the utility from the candidates at the first stage of $\hat{\Gamma}$ while reaching B.[31] But reaching A in Γ^{*} yields player i a utility, which is at least as much as reaching B in the same game, because the restriction of σ to Γ^{*} is an equilibrium for Γ^{*} by construction and Proposition 4.7. Consequently, the payoff to i at A in Γ^{*} is not smaller than the payoff to i at B in Γ^{*-}, which proves that σ is indeed subgame-perfect. ∎

[31] We are invoking additivity across stages.

5. Perfect equilibria in pure strategies

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

We are able to prove the main theorems under the assumption that the voting scheme is *generic*; namely, it is such that different streams yield different utilities for each player. Example 5.3 shows that this assumption is necessary for the result.

Proposition 5.1. *Let Γ^1 be a first stage game, derived from a game Γ representing a generic voting scheme as defined in Section 4, given a fixed continuation at the proper subgames of Γ. Suppose that there exists a set of votes $\mathcal{P} = \{P_j\}_{j \in F^0}$, where $P_j \subsetneq C^0$, whose union is denoted by S, that satisfies:*

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

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

Define

$$V_j := \{x \in S : S \succ_j S \setminus \{x\}\}. \tag{5.1}$$

Under these conditions,[32] *$\mathcal{V} := \{V_j\}_{j \in F^0}$ is a perfect equilibrium profile for Γ^1.*

[32] Here, $\succ_j$ means: 'Preferred by j'.

Terminology: Profile $\mathcal{P}$, satisfying (1) and (2) above will henceforth be called *a generalized-partition equilibrium profile* (of S, when needed).

Proof. The theorem is certainly true[33] if $|C^0| \leq 1$, or if $|F^0| = 1$, so we assume that $|C^0| \geq 2$ and $|F^0| \geq 2$. We call members of V_i *desirable* for i. Other members of S called *undesirable* for player i.

For all i in F^0, denote $\mathcal{P}_{-i} := \cup\{P_j : j \in F^0 \setminus \{i\}\}$. As P_i is a best response of agent i to $\mathcal{P}_{-i}$, and $P_i \cap \mathcal{P}_{-i} = \emptyset$, it follows that each member of P_i is desirable for i. Consequently,

$$P_i \subseteq V_i \subseteq S, \qquad \forall i \in F^0. \tag{5.2}$$

Conversely, if $\hat{V}_i$ satisfies $P_i \subseteq \hat{V}_i \subseteq S$, then $\hat{V}_i$ is a best response to $\mathcal{P}_{-i}$ (Lemma 4.8).

For all j in F^0, denote

$$M_j := \{T \subseteq C^0 : T = P_j \setminus \{x\} \text{ and } x \in P_j\}, \tag{5.3}$$

$$H_j := \{T \subseteq C^0 : T \neq P_j,\ T \neq V_j,\ T \notin M_j\}. \tag{5.4}$$

Note that $M_j = \emptyset$ iff $P_j = \emptyset$ and $H_j \neq \emptyset$, because $|C^0| \geq 2$, as can easily be checked. For any positive $\epsilon_1, \epsilon_2, \epsilon_3$, such that $\epsilon_1 + \epsilon_2 + \epsilon_3 < 1$, define $\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3$. For each j in F^0, define ϵ^j as $\epsilon_1 + \epsilon_2 + \epsilon_3$, if $\mathcal{P}_{-j} \neq \emptyset$ and as $\epsilon_1 + \epsilon_3$, if $\mathcal{P}_{-j} = \emptyset$. We construct the following completely mixed strategy σ_j for player j, $j \in F^0$:

[33] If $|C^0| = 1$, and $|F^0| > 1$, then all the players vote their preferred outcome out of the pair $\{\emptyset, C^0\}$. Since there are only two possible outcomes, voting for the more preferred outcome against any strictly mixed strategy profile always gives a higher probability of it occurring than voting for the less preferred one. This proves that voting for the most preferred outcome is a perfect equilibrium. Other cases are even simpler.

With probability $1-\epsilon^j$ vote V_j, $\quad 0 < \epsilon^j < 1$; (5.5)

With probability ϵ_1 vote P_j, $\quad 0 < \epsilon_1 < 1$; (5.6)

(If $P_j = V_j$ vote for V_j with probability $1 - \epsilon^j + \epsilon_1$.)

If $P_j \neq \emptyset$ vote with probability $\frac{\epsilon_2}{|M_j|}$ for $P_j \setminus \{x\}$, for each member x of P_j; (5.7)

Vote with probability $\frac{\epsilon_3}{|H_j|}$ for each member of H_j. (5.8)

Additional conditions on ϵ^j, ϵ_1, ϵ_2, ϵ_3 will be placed later, but we can now state that

$$\epsilon^j = \epsilon_1 + \epsilon_2^j + \epsilon_3, \tag{5.9}$$

where we set $\epsilon_2^j = \epsilon_2$ if $P_j \neq \emptyset$ and $\epsilon_2^j = 0$, otherwise. Consequently, $\epsilon_1, \epsilon_2^j, \epsilon_3 \to 0$, if $\epsilon^j \to 0$; namely, $\sigma_j \to V_j$ for each $j \in F^0$ and $\{\sigma_j\}_{j \in F^0}$ is a valid test sequence. The proof will be concluded if we show that the epsilons can be chosen in such a way that V_i will be a best response to σ_{-i} for all the members of the sequence.

For a fixed i in F^0, let T be a set chosen by $F^0 \setminus \{i\}$ under $\sigma_{-i} := \{\sigma_j;\ j \in F^0 \setminus \{i\}\}$. Denote by η, η_1, $\eta_2(x)$, η_3, the following probabilities:

η : The probability that $\mathcal{P}_{-i} \subseteq T \subseteq S$ and at least one member j in $F^0 \setminus \{i\}$ did not vote P_j.

(Note that η could be zero. This happens, for example, if $S = \emptyset$.)

η_1 : The probability that each j in $F^0 \setminus \{i\}$ voted P_j.

$\eta_2(x)$: The probability that $T = \mathcal{P}_{-i} \setminus \{x\}$, for some x in $\mathcal{P}_{-i}$.

Note that no $\eta_2(x)$ is defined if $\mathcal{P}_{-i} = \emptyset$.

η_3 : The probability that any other set is chosen.

Clearly,

$$\eta + \eta_1 + \sum_{x \in \mathcal{P}_{-i}} \eta_2(x) + \eta_3 = 1, \tag{5.10}$$

where summation over an empty set is defined as zero.

We can place the following bounds on these probabilities as follows:

$$\eta_1 \geq \epsilon_1^{q-1}, \qquad \text{where } q = |F^0|; \tag{5.11}$$

$$\eta_2(x) \geq \epsilon_1^{q-2}\epsilon_2/|S|, \qquad \text{all } x \text{ in } \mathcal{P}_{-i} \tag{5.12}$$

$$\eta_2(x) \leq \epsilon_2 + \epsilon_3, \qquad \text{all } x \in \mathcal{P}_{-i}; \tag{5.13}$$

$$\eta_3 \leq ((\epsilon_2)^2 + \epsilon_3) \cdot w, \qquad \text{where } w = 2^{|C^0|(q-1)}. \tag{5.14}$$

Indeed, (5.11) follows from the definitions of ϵ_1 and η_1. (Strict inequality prevails if $P_j = V_j$ for some j in $F^0 \setminus \{i\}$.) (5.12) follows from the fact that the event that $\mathcal{P}_{-i} \setminus \{x\}$ is chosen by $F^0 \setminus \{i\}$, whose probability is measured by $\eta_2(x)$, occurs, for example, if agent j, whose P_j contains x, votes $P_j \setminus \{x\}$ and every other player ℓ, in $F^0 \setminus \{i\}$, votes P_ℓ.[34] (5.13) follows from the fact that this event $\mathcal{P}_{-i} \setminus \{x\}$ is included in the event: The above player j voted for neither P_j nor V_j (probability $\epsilon_2^j + \epsilon_3 \leq \epsilon_2 + \epsilon_3$) (see 5.9)), and all other players in $F^0 \setminus \{i\}$ voted according to σ_{-i} (probability not larger than 1). To prove (5.14), notice that this event occurs only if one of the following elementary events happened:

(1) One player j in $F^0 \setminus \{i\}$ did not vote either P_j or V_j, or $P_j \setminus \{x\}$, for some x in P_j — an event whose probability is at most ϵ_3;

(2) Two players j and ℓ in $F^0 \setminus \{i\}$ voted $P_j \setminus \{x\}$ and $P_\ell \setminus \{y\}$, for some $x \in P_j$ and $y \in P_\ell$ — an event whose probability is at most $\epsilon_2^j \cdot \epsilon_2^\ell$.

[34] Here we use property (2) of the proposition.

The number of such events is at most w and $\epsilon_2^j \cdot \epsilon_2^\ell \le (\epsilon_2)^2$. (5.14) follows now from the observation that $\max\{(\epsilon_2)^2, \epsilon_3\} \le (\epsilon_2)^2 + \epsilon_3$.

In order to show perfectness of $\mathcal{V}$, we have to show that V_i is a best response to σ_{-i} for each i, if one chooses $\epsilon, \epsilon_1, \epsilon_2, \epsilon_3$ appropriately, and that such a choice can be maintained for $\epsilon \to 0$. To this end we denote:

h_i: The payoff to i if S is chosen.

a: The minimum loss to i if a set T results, $T \neq S$, such that $T \supseteq \mathcal{P}_{-i}$.

$b(x)$: The gain to i if $S \setminus \{x\}$, results, $x \in \mathcal{P}_{-i}$, and x is undesirable for i; i.e., $x \notin V_i$.

b: The minimum of all the $b(x)$ for $x \in \mathcal{P}_{-i} \setminus V_i$.

c: The minimal loss to i if $S \setminus \{x\}$, $x \in \mathcal{P}_{-i}$ results, and x is desirable for i; i.e., $x \in V_i$.

M: The maximal payment in Γ^1.

m: The minimal payment in Γ^1.

Note that a, $b(x)$, c are positive, because the voting scheme is generic. Thus, only a vote giving the outcome S is a best reply to $\mathcal{P}_{-i}$. $b(x)$ and c are undefined if $\mathcal{P}_{-i} = \emptyset$. They are not necessary for this case as (5.19) will not be used.

We now give bounds on the payoffs to i under various pure strategies of his, when all agents in $F^0 \setminus \{i\}$ vote according to σ.

If i votes V_i, his payoff is *at least*

$$(\eta + \eta_1)h_i + \sum_{x \in \mathcal{P}_{-i} \cap V_i} \eta_2(x)h_i + \sum_{x \in \mathcal{P}_{-i} \setminus V_i} \eta_2(x)(h_i + b(x)) + \eta_3 m. \tag{5.15}$$

Indeed, S is certainly covered if all members j in $F^0 \setminus \{i\}$ vote either V_j or P_j, or if they vote $\mathcal{P}_{-i} \setminus \{x\}$ and $x \in V_i$. If, on the other hand, $\mathcal{P}_{-i} \setminus \{x\}$ results and $x \notin V_i$, i will gain $b(x)$. In all cases, however, he will get at least m.

If i votes for a set Q_1 which is not a best reply to $\mathcal{P}_{-i}$, his payoff is *at most*

$$\eta h_i + \eta_1(h_i - a) + (\eta_3 + \sum_{x \in \mathcal{P}_{-i}} \eta_2(x))M. \tag{5.16}$$

Indeed, he will not get more than h_i if $F^0 \setminus \{i\}$ vote T such that $\mathcal{P}_{-i} \subseteq T \subseteq S$. However, he will certainly get at most $h_i - a$ if $\mathcal{P}_{-i}$ is voted by the others, which happens with probability η_1 at least. In all other cases he will not get more than M.

If i votes for a set Q_2 that *is* a best reply to $\mathcal{P}_{-i}$, but is different from V_i, then since $P_i \subseteq Q_2 \subseteq S$, Q_2 results from V_i by omitting r members and adding s members from $S \setminus V_i$, with $r + s > 0$. Here,

$$r = |(\mathcal{P}_{-i} \cap V_i) \setminus Q_2|, \tag{5.17}$$

$$s = |(\mathcal{P}_{-i} \setminus V_i) \cap Q_2|. \tag{5.18}$$

The payment to i if he chooses such a Q_2 is *at most*

$$\begin{aligned}(\eta + \eta_1)h_i + \sum_{x \in (\mathcal{P}_{-i} \cap V_i) \setminus Q_2} \eta_2(x)(h_i - c) + \sum_{x \in (\mathcal{P}_{-i} \cap V_i) \cap Q_2} \eta_2(x) h_i + \sum_{x \in (\mathcal{P}_{-i} \setminus V_i) \cap Q_2} \eta_2(x) h_i + \\ \sum_{x \in (\mathcal{P}_{-i} \setminus V_i) \setminus Q_2} \eta_2(x)(h_i + b(x)) + \eta_3 M.\end{aligned} \tag{5.19}$$

Here, sums over an empty set are considered equal to zero. Indeed, S will certainly result with probability $\eta + \eta_1$ at least. He will lose at least c each time $F^0 \setminus \{i\}$ vote $P_{-i} \setminus \{x\}$ with x desirable for i and i does not vote for x in Q_2; otherwise, he will still get h_i if he himself voted for x. He will not gain $b(x)$ from the omission of an undesirable x from $\mathcal{P}_{-i}$, if he himself voted for x.

Expression (5.15) is not less than expression (5.16) whenever

$$a\eta_1 \geq (M - m)(\eta_3 + \sum_{x \in \mathcal{P}_{-i}} \eta_2(x)), \tag{5.20}$$

because $h_i \geq m$. This is all that is needed if $\mathcal{P}_{-i} = \emptyset$ for V_i to be a best reply to σ_{-i}.[35] Indeed, in this case $P_i = S = V_i$ and agent i does not have a strategy that is a best response to $\mathcal{P}_{-i}$, which is

[35] In this case, the summation in (5.20) is zero.

different from V_i. If, however $\mathcal{P}_{-i} \neq \emptyset$, then, in addition to (5.20) the following inequality, derived from (5.15) and (5.19) and the fact that $b(x) \geq b$, should also be satisfied for all sets Q_2 that are best replies to $\mathcal{P}_{-i}$ but differ from V_i:

$$c \sum_{x \in (\mathcal{P}_{-i} \cap V_i) \setminus Q_2} \eta_2(x) + b \sum_{x \in (\mathcal{P}_{-i} \setminus V_i) \cap Q_2} \eta_2(x) \geq \eta_3(M-m). \tag{5.21}$$

Now, recall that

$$\epsilon = \epsilon_1 + \epsilon_2 + \epsilon_3 \tag{5.22}$$

and note that if $\epsilon \to 0$ then also $\epsilon^j \to 0$, because $0 < \epsilon^j \leq \epsilon$; therefore, by letting ϵ trace a sequence tending to zero, our completely mixed strategies will converge to V.

It follows from (5.11), (5.13) and (5.14), that (5.20) will be satisfied if

$$\epsilon_1^{q-j} \geq \frac{M-m}{a}\left[w((\epsilon_2)^2 + \epsilon_3) + |S|(\epsilon_2 + \epsilon_3)\right]. \tag{5.23}$$

This inequality can certainly be maintained for all j, by letting ϵ_1 be near enough to ϵ. Having fixed ϵ_1 sufficiently near ϵ, we are still free to choose ϵ_2 and ϵ_3, as long as their sum remains constant (namely, $\epsilon - \epsilon_1$).

It follows from (5.12), (5.14) and the fact that $r + s > 0$, that (5.21) will be satisfied if Q_2's exist[36] and

$$(\epsilon_1)^{q-2} \geq (M-m)\frac{w|S|}{d(r+s)}\left(\frac{\epsilon_3}{\epsilon_2} + \epsilon_2\right). \tag{5.24}$$

where $d = \min\{b, c\}$.

To this end, for a fixed positive ϵ, choose ϵ_1 sufficiently near to ϵ, so that $\epsilon_2 + \epsilon_3$ becomes so small that for any choice of ϵ_2,

$$\frac{1}{2}(\epsilon_1)^{q-2} \geq (M-m)\frac{w|S|}{d(r+s)}\epsilon_2. \tag{5.25}$$

[36] If no such Q_2 exist, (5.21) need not be satisfied.

Having fixed also ϵ_1, choose ϵ_2 sufficiently near to $\epsilon - \epsilon_1$, so that ϵ_3 becomes so small that also

$$\frac{1}{2}(\epsilon_1)^{q-2} \geq (M - m)\frac{w|S|}{d(r+s)}\frac{\epsilon_3}{\epsilon_2}. \tag{5.26}$$

Adding (5.25) and (5.26) yields (5.24).

We have therefore proved that for any positive ϵ we can choose ϵ_1, ϵ_2 and ϵ_3 in such a way that V_i will be a best reply to σ_i, uniformly for all i in F^0. Letting $\epsilon \to 0$ concludes the proof. ∎

Proposition 5.1 raises two interesting questions:

(1) What conditions guarantee that a generalized-partition equilibrium profile exists?

(2) Is the existence of a generalized-equilibrium-partition necessary for the existence of pure-strategy perfect equilibrium?

We answer the second question negatively, by the following example:

Example 5.2. The population consists of:

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

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

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

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

The payoff matrix is given by[37]

	$\emptyset$	a	b	ab
$\emptyset$	2, 4	3, 2	4, 1	1, 3
a	3, 2	3, 2	1, 3	1, 3
b	4, 1	1, 3	4, 1	1, 3
ab	1, 3	1, 3	1, 3	1, 3

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

In this example the pure equilibrium profiles are $(\{a\},\{a,b\})$, $(\{b\},\{a,b\})$ and $(\{a,b\},\{a,b\})$. None of them is a generalized-equilibrium-partition, nevertheless, $(\{a\},\{a,b\})$ and $(\{b\},\{a,b\})$ are perfect equilibrium profiles.[38] This shows that Proposition 5.1 does not yield necessary conditions. On the other hand, Example 3.1 shows that voting schemes exist that do not have any pure-strategy perfect equilibrium. Providing a necessary and sufficient condition for the existence of pure-strategy perfect equilibrium in a 1-stage game remains an open question.

The next example will show that the requirement that the game is generic is needed for Proposition 5.1 to hold.

Example 5.3. The population is:

$$F^0 = \{1,2\}, \quad C^0 = \{a,b,x,y\}.$$

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

$$u_1(\emptyset) = 3; \quad u_1(\{a\}) = 4; \quad u_1(\{b\}) = 2; \quad u_1(\{a,b\}) = 1; \quad u_1(S) = 0, \text{ otherwise.}$$

$$u_2(\emptyset) = 4; \quad u_2(\{a\}) = 1; \quad u_2(\{b\}) = 3; \quad u_2(\{a,b\}) = 2; \quad u_2(S) = 0, \text{ otherwise.}$$

Founder 1 voting x and Founder 2 voting y is a generalized-partition-equilibrium profile but the game is not generic. Eliminating all weakly dominated pure strategies, which cannot be employed in a perfect equilibrium profile leaves us with the pure-strategy profiles $(\emptyset,\emptyset)$, $(\emptyset,\{b\})$, $(\{a\},\emptyset)$ and $(\{a\},\{b\})$, none of which is even an equilibrium profile. This shows that requiring genericity is needed in Proposition 5.1.

An interesting application of Proposition 5.1 is the following:

---

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

Theorem 5.4. *Let Γ be a game representing a 2-stage generic voting scheme, whose utilities obey additivity across stages and additivity within each stage (Assumption 8b). Under these conditions, Γ has a perfect equilibrium in pure strategies.*

Proof. Any perfect equilibrium profile for Γ must specify for each subgame of the second stage a profile under which each voter votes precisely for the set of his friends (who are not already in the society). This is a perfect equilibrium of the subgame (Section 2, case $k = 1$) and unique, by genericity. With this understanding, we can construct a 1-stage game Γ^1 as was done in Section 4. The proof will be concluded if we show that Γ^1 has a pure-strategy perfect equilibrium, as the combination of this strategy with the continuation is a perfect strategy[39] for Γ. To achieve that, it is sufficient, by Proposition 5.1, to exhibit a generalized-partition equilibrium profile for Γ^1. This we are about to do by a construction under which voters add candidates to the society piecewise: There will be a variable set of candidates, called a *current set*, that grows, or stays put, as the voters add to in during the construction, until it eventually becomes the outcome for Stage 1, as well as an outcome of Γ^1. We introduce the following terminology: Let A be a current set of candidates. We say that a, possibly empty, set of candidates taken from $C^0 \setminus A$, is *optimal for voter i w.r.t. A*, and denoted $X_i(A)$, if it is the best set of candidates that i could add to A, so as to increase his utility from the two stages. Note that $X_i(A)$ cannot contain enemies of i, since such candidates are enemies, and can only contribute more enemies at Stage 2. (The friends of i will be brought in anyhow by i at Stage 2.) In symbols, $X_i(A)$ is characterized by

$$\begin{aligned} &w_i(A \cup X_i(A)) + w_i(\mathrm{en}_i\,(F^0 \cup A \cup \mathrm{fr}\,(F^0 \cup A \cup X_i(A)))) \geq \\ &w_i(A \cup B) + w_i(\mathrm{en}_i\,(F^0 \cup A \cup \mathrm{fr}\,(F^0 \cup A \cup B))), \qquad \text{all } B \subseteq \mathrm{fr}_i\,(C^0 \setminus A). \end{aligned} \tag{5.27}$$

(In this calculation friends of i at the second stage are omitted from both sides of each inequality.) Here, $w_i(T) := \sum_{t \in T} w_i(t)$, $\mathrm{fr}_i\,(S) := \{j : j \in \mathrm{fr}\,(i) \cap S\}$, $\mathrm{en}_i\,(S) := \{j : j \in \mathrm{en}\,(i) \cap S\}$ and

[39] A proof for any k-stage game is given in Theorem 5.10.

$\mathrm{fr}(B) := \{\ell : \ell \in \mathrm{fr}(j)$ for some j in $B\}$. Sums over the empty set are considered equal to zero.

By genericity, the set $X_i(A)$ is unique.

THE CONSTRUCTION:

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

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

Lemma 5.5. *Assume the conditions and notations of Theorem 5.4. Let A and B be two sets of candidates, $A \subseteq B$. Let C be a set of friends of a voter i satisfying $C \cap B = \emptyset$. If $A \cup C \succ_i A$ then $B \cup C \succ_i B$.*

Proof. From the data it follows that the total weight of i from C exceeds the absolute value of the total weight of the new enemies that C brings at Stage 2.[40] When C is added to B he brings the same number of friends, namely $|C|$, and no new enemies. Perhaps even less — the previous ones that happen to be in $B \setminus A$. ■

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

(Continuation of the proof of Theorem 5.4). If $(P_j)_{j\in F^0}$ is not an equilibrium profile, then a voter i can benefit from a deviation. A deviation means that he deletes a set T of candidates from his vote P_i and adds a set Q of candidates not in S.[41] At least one of these sets is not empty. The set T, if not empty, is a union of nonempty sets $T_1, T_2, \ldots T_r$, which are, respectively, subsets of his votes $P_i^1, P_i^2, \ldots, P_i^r$ taken when i was approached at times that we enumerate chronologically $1, 2 \ldots, r$. Denote by $S_1, S_2, \ldots, S_r$ the current sets at these times after his addition. Consider a hypothetical sequence when all founders vote as in the construction except that agent i votes $P_i^1 \setminus T_1$ at time 1, $P_i^2 \setminus (T_1 \cup T_2)$ at time 2, $\ldots$, $P_i^r \setminus T$ at time r and each time he also adds the candidates of Q. The end of this sequence is the deviation, which, as we assumed, benefited player i. We now modify this sequence in such a way that player i will continue to benefit and at least as much. To this end, add T_1 to the hypothetical vote of voter i at all times, starting from time 1. This will benefit him at time 1. Indeed, he would benefit if the current set were $S_1 \setminus T_1$ because $X_i(S_1 \setminus P_i^1) = S_1$ is the unique optimal response and so, by Lemma 5.5, he would benefit by adding T_1 to $(S_1 \setminus T_1) \cup Q$. For the same reason i would benefit by adding T_1 at every part of the hypothetical sequence, since $S_1 \setminus T_1 \subseteq Q \cup (S_t \setminus (T_1 \cup T_2 \cdots \cup T_t))$ and $T_1 \cap (Q \cup (S_t \setminus (T_1 \cup T_2 \cdots \cup T_t)) = \emptyset$, $t \in \{1, 2, \ldots, r\}$. After adding T_1 we are in an improved deviation that starts at time 2. We make a similar modification and continue for r times. Eventually, we arrive at an improved deviation at which only Q is added. But this is impossible, since the original construction ended when no voter could beneficially add members outside the current set. The contradiction shows that we are indeed at equilibrium. ∎

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

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

described in Proposition 5.1 is the same, regardless of the order of approach.

Lemma 5.6. *If $A \subseteq B \subseteq C^0$, then $A \cup X_i(A) \subseteq B \cup X_i(B)$ for every agent i in F^0.*

Proof. Assume negatively, that for some i in F^0, $D := (A \cup X_i(A)) \setminus (B \cup X_i(B)) \neq \emptyset$. By optimality of $X_i(A)$ and genericity of Γ, it follows from (5.27), replacing B by $X_i(A) \setminus D$, and noting that $D \cap A = \emptyset$, that

$$w_i(D) + w_i(\text{en}_i(F^0 \cup A \cup \text{fr}(F^0 \cup A \cup X_i(A)))) - w_i(\text{en}_i(F^0 \cup A \cup \text{fr}(F^0 \cup A \cup (X_i(A) \setminus D)))) =$$
$$w_i(D) + w_i(\text{en}_i(\text{fr}(D) \setminus \text{fr}(F^0 \cup A \cup (X_i(A) \setminus D)))) > 0. \tag{5.28}$$

Using (5.27) once more, replacing $A, X_i(A), B$ by $B, X_i(B), X_i(B) \cup D$, respectively, we obtain:

$$w_i(D) + w_i(\text{en}_i(\text{fr}(D) \setminus \text{fr}(F^0 \cup B \cup X_i(B)))) < 0. \tag{5.29}$$

However, $(A \cup X_i(A)) \setminus D \subseteq B \cup X_i(B)$, and enemies of i carry negative utilities; therefore, the left side of (5.29) is not smaller than the left side of (5.28) — a contradiction. ∎

Corollary 5.7. *Changing the order of the referee's approaches leads to the same final set S, although the actual votes of the players may be different.*

Proof. Let $\emptyset = T^0, T^1, \ldots, T^r = T$ be the sequence of 'current sets' generated by a different order of approaches. We shall show that $T^m \subseteq S$ for every m and therefore $T \subseteq S$. Reversing the roles of S and T one gets $S \subseteq T$ and this concludes the proof. Proceed by induction: Certainly $T^0 \subseteq S$. Suppose $T^{m-1} \subseteq S$ and $T^m \not\subseteq S$. Then, some i in F^0 has $X_i(T^{m-1}) \not\subseteq S$. Thus, a candidate a exists in $X_i(T^{m-1})$, $a \notin S$. From Lemma 5.5, $a \in X_i(T^{m-1}) \subseteq X_i(S)$, which contradicts the fact that the construction terminates when $X_i(S) = \emptyset$ for all i. ∎

One may now ask whether a perfect equilibrium profile is always unique under the conditions of Theorem 5.4. The following example settles this question negatively.

Example 5.8. The set of founders is $F^0 = \{1, 2\}$. The set of candidates is $C^0 = \{a, b, c\}$. $k = 2$ and we assume pure friendship and enmity (Assumption 8a). $\mathrm{fr}(1) = \{a\}, \mathrm{fr}(2) = \{b\}$, $\mathrm{fr}(a) = \mathrm{fr}(b) = \{c\}$. The construction in Theorem 5.4 leads to $S = \emptyset$. However, it can be checked that 1 and 2 voting for their friends at all stages and a and b vote for their friend at Stage 2 is also a perfect equilibrium profile.

It is interesting to find conditions that guarantee that a pure-strategy *generalized partition* equilibrium profile always exists. For a while we thought that this will always be the case if additivity holds within a stage and across stages (Assumption 8b), as Theorem 5.4 seems to indicate. The following example shows that this is not true and, moreover, it may well be that no pure-strategy perfect equilibrium exists in this case.

Example 5.9. The population is:

$$F^0 = \{1, 2\}, \quad C^0 = \{a, b, x_1, x_2, m_1, m_2, y\}.$$

The number of periods is $k = 3$. The weights of each member of the population from being with each other member, per period, are:

$$w_1(a) = w_1(b) = 100, \quad w_1(x_1) = w_1(x_2) = -200,$$

$$w_2(a) = 500, \quad w_2(b) = 100, \quad w_2(m_1) = -600, \quad w_2(m_2) = -200,$$

$$w_a(x_1) = w_a(x_2) = w_a(m_1) = 100, \quad w_a(y) = -200,$$

$$w_b(x_2) = w_b(m_2) = 100,$$

$$w_{m_1}(y) = 1,$$

$$w_{m_2}(y) = 1.$$

All other weights are equal to -1. Utilities are calculated by (2.3).

The construction of the unique perfect equilibrium profile will be carried by backwards induction, starting from the third stage and working backwards towards the first stage. We shall demonstrate that it must employ a mixed-strategy profile.

The third and last stage. In any perfect equilibrium, at the last stage any member of the society (elements of F^2) invites exactly the set of his friends that are not members of the society.

Thus,[42] 1 and 2 invite a and b; a invites x_1; x_2 and m_1; b invites x_2 and m_2; m_1 invites y and m_2 invites y.

The second stage. Based on their knowledge of what will happen in the last stage for any configuration of F^2, the invitations at stage two in any perfect equilibrium can be calculated by using deletion of dominated strategies.

Agents m_1 and m_2 will invite y. Agent b will invite x_2 and m_2. Agent a will invite x_1 and x_2, and will also invite m_1 if b belongs to F^1 (since in this case y will be invited at stage 3 by m_2 who is invited by b). If a is in F^1, then 1 will invite b (since in this case both x_1 and x_2 will be invited by a at stage 3). If a is not in F^1 then no one will invite a and therefore 1 will not invite b if a is not in F^1.

The First Stage. Based on the above continuations, there are four possible configurations to consider for the first stage. These constitute all the subsets of $\{a, b\}$. Neither 1 nor 2 can gain by inviting any of the others.

If the empty set is invited at stage 1 (i.e. $F^1 = \{1,2\}$), then based on the previous analysis, the continuation will have $F^2 = \{1,2\}$, $F^3 = \{1,2,a,b\}$. If only a is invited at stage 1, then the stream will be $F^1 = \{1,2,a\}$, $F^2 = \{1,2,a,b,x_1,x_2\}$, $F^3 = \{1,2,a,b,x_1,x_2,m_1,m_2\}$.

[42] We remind the reader that the "invitations" are useless if the inviting agent is not a member of the society.

If only b is invited at stage 1, then the stream will be $F^1 = \{1,2,b\}$, $F^2 = \{1,2,b,x_2,m_2\}$, $F^3 = \{1,2,a,b,x_2,m_2,y\}$. And if both a and b are invited at stage 1, the stream will be $F^1 = \{1,2,a,b\}$, $F^2 = \{1,2,a,b,x_1,x_2,m_1,m_2\}$, $F^3 = \{1,2,a,b,x_1,x_2,m_1,m_2,y\}$.

Using the data about the weights, we can calculate the payoffs for the possible actions in the first stage. The row player is player 1 and the column player is player 2.

	$\emptyset$	a	b	ab
$\emptyset$	200 600	−302 896	−3 397	−205 195
a	−302 896	−302 896	−205 195	−205 195
b	−3 397	−205 195	−3 397	−205 195
ab	−205 195	−205 195	−205 195	−205 195

Since dominated strategies are not used in any perfect equilibrium of a normal-form game, we delete dominated strategies for the two players. We are left with the following normal-form payoff matrix for the first stage:

	$\emptyset$	a
$\emptyset$	200 600	−302 896
b	−3 397	−205 195

Clearly, there is no pure-strategy perfect equilibrium, and the only perfect equilibrium with pure moves at stages 2 and 3 consists of using mixed strategies in the first stage — Player 1 voting for both the empty set and $\{b\}$ with positive probabilities, and Player 2 voting for both the empty set and $\{a\}$ with positive probabilities.

This example demonstrates that assumption 8b of Section 2 is not enough to guarantee the existence of a pure-strategy perfect equilibrium.

We conclude this section by extending Proposition 5.1 to several-stage voting schemes.

Theorem 5.11. *Let Γ be a game representing a generic voting scheme, obeying additivity across stages (Assumption 8c). If one can work backwards on all subgames, as described in Section 4, finding sets of candidates obeying the conditions of Proposition 5.1, one obtains a perfect equilibrium for Γ in pure strategies.*

Proof. Let σ be the strategy profile constructed as in the theorem. Based on σ we can construct 1-stage games at every start of a subgame, as was done for Γ^1 at the beginning of Section 4. We then construct a test sequence for every 1-stage game that converges to the restriction of σ for that 1-stage game, such that σ_i is a best reply for each element of the sequence, for each voter. This can be done as shown constructively in the proof of Proposition 5.1. Note that by genericity, the best reply is unique. Moreover, it continues to be a best reply even if the payoff vectors at the endpoints are slightly modified. Such modifications are, in fact, created from the test sequences of the games at the next stages.

To construct a test sequence for the global game Γ, cut from each 1-stage test sequence enough elements so that, together, the remaining parts will cause perturbations so small that σ_i, for each i, will remain a best reply also for the perturbed payoff vectors. The existence of such a sequence shows that σ is a perfect equilibrium point. ∎

6. Appendix

In this appendix we shall discuss the merits and limitations of the requirement that the agents use only history-independent strategies. This assumption certainly simplifies our analysis. One can try and claim that it is appropriate if ballots are secret, but this is not good enough since part of the history is known by watching who was elected at each stage.

On the face of it this requirement looks innocuous: for example, you come to stage 3 and have 5 stages to go. You know who are the voters and who are the candidates. You have all information concerning priorities of each agent. You have to make your choice. Why should you care how you came to this situation? Isn't it spilt milk?

Well, – not always!

If one is interested only in equilibrium outcomes that can be achieved in pure strategies, Theorem 6.1 below shows that the same stream of members can be obtained as an equilibrium outcome using only history-independent pure-strategy profiles. Example 6.2 shows that this is not the case when mixed strategies are being considered.

One can claim that limiting the agents to history-independent pure-strategy equilibrium profile is not a good restriction, if an agent can profit by deviating to a mixed, history-dependent strategy. Theorem 6.3 proves that this cannot happen.

Theorem 6.1. *Any equilibrium outcome that can be achieved in pure-strategy [subgame-perfect] profiles can also be achieved with pure-strategy history-independent [subgame-perfect] profiles.*

Proof. If there is only one original founder, then, as long as he votes for nobody, we can regard his votes as history-independent since he can choose his votes without looking at what happened.

So, we can assume, without loss of generality, that there are at least two original founders. Let $\sigma = (\sigma_i)_{i\in N}$ be a Nash equilibrium profile obtained by pure strategies. Let $\tau = (\tau_i)_{i\in N}$ be a history-independent strategy profile defined by

$$\tau_i^t(t, F^{t-1}) = \begin{cases} \sigma_i^t(h^t(\sigma)), & \text{if } F^{t-1} = F^{t-1}(\sigma), \\ N, & \text{otherwise.} \end{cases} \tag{6.1}$$

Here, $h^t(\sigma) = \{F^0, F^1(\sigma), \ldots, F^{t-1}(\sigma)\}$ and $\sigma_i^t(h^t(\sigma))$ is the vote cast by agent i at stage t, given the history up to that stage.

The path followed by profile τ is the same as the one followed by σ, hence τ yields the same stream of members as σ, and therefore the same utility outcome. It remains to show that τ is a Nash equilibrium. Assume that agent i can profit by deviating alone from τ_i, using strategy τ_i'. Let t_0 be the first stage in which $\tau_i' \neq \tau_i$. Since there are at least two founders, τ' generates the stream $(F^0, F^1(\sigma), \ldots, F^{t_0}(\sigma_{-i}|\tau_i'^{t_0}), N, \ldots, N)$. Player i can deviate also from σ and obtain the same stream of members. Indeed, all he has to do is deviate from stage t_0, voting as in τ' at that stage and voting N afterwards. This will yield him a higher utility, contrary to the fact that σ was an equilibrium profile.

If σ is also subgame-perfect then so is τ and this completes the proof. ∎

Note that a similar theorem may not hold when we deal with other refinements. For example, in general τ, as constructed above, will not be a perfect equilibrium, even if σ was perfect.

Unfortunately, Theorem 6.1 does not hold, in general, if σ is a mixed strategy equilibrium profile. The next example will demonstrate this fact.

Example 6.2. The population is:

$$F^0 = \{1, 2\}, \quad C^0 = \{a_1, a_2, x_1, x_2, z\}.$$

The number of Stages is $k = 4$. The utilities of each member obey additivity within each stage and across stages (Assumption 8b). The weight functions are:

$$w_1(a_1) = 0, \quad w_1(a_2) = -1, \quad w_1(x_1) = 10, \quad w_1(x_2) = -10, \quad w_1(z) = -100,$$

$$w_2(a_1) = -1, \quad w_2(a_2) = 0, \quad w_2(x_1) = -10, \quad w_2(x_2) = 10, \quad w_2(z) = -100,$$

$$w_{a_1}(x_1) = -2, \quad w_{a_2}(x_1) = -2.$$

All other utilities not given above are equal to -1.

Consider the following strategy profile.

At Stage 1, Player 1 randomizes (with equal probabilities) between $\{a_1\}$ and $\emptyset$, while player 2 randomizes (with equal probabilities) between $\{a_2\}$ and $\emptyset$,

At Stage 2, all members vote for $\{a_1, a_2\}$.

At Stage 3, if $|F^1|$ was even, then all members invite $\{x_1\}$. Otherwise, they all invite $\{x_2\}$.

At Stage 4, all invite their own friends.

If there is any detectable deviation from this path before Stage 4, all members invite z and their own friends immediately after the deviation is detected.

This is a Nash equilibrium. Neither founder can gain with an undetectable deviation at the first stage, and a detectable deviation causes a loss. Any deviation that makes a difference at Stages 2 or 3 is detectable and causes a loss. There are no profitable deviations at the last stage. The strategies are history-dependent, as if $F^2 = \{a_1, a_2\}$, then members of F^2 have different actions for Stage 3, depending on whether the number of members in F^1 was even or odd.

The outcome of the equilibrium is one of the four following streams, each occurring with equal probability:

$$(F^1, \ldots, F^4) = (\emptyset, \{a_1, a_2\}, \{a_1, a_2, x_1\}, \{a_1, a_2, x_1, x_2\}$$

$$(F^1, \ldots, F^4) = (\{a_1\}, \{a_1, a_2\}, \{a_1, a_2, x_2\}, \{a_1, a_2, x_1, x_2\}$$

$$(F^1, \ldots, F^4) = (\{a_2\}, \{a_1, a_2\}, \{a_1, a_2, x_2\}, \{a_1, a_2, x_1, x_2\}$$

$$(F^1, \ldots, F^4) = (\{a_1, a_2\}, \{a_1, a_2\}, \{a_1, a_2, x_1\}, \{a_1, a_2, x_1, x_2\}$$

Note that at Stage 3 either x_1 or x_2 is invited, depending on the cardinality of F^1. However, since F^2 is the same for each of the four streams, history-independent strategy cannot specify different actions for the third stage. Therefore, this outcome, which was achieved in equilibrium with history-dependent strategies, cannot be supported with history-independent strategies.

Restricting the strategy space to pure history-independent strategy profiles raises the doubt whether equilibrium points in this space might cease to be in equilibrium when one extends the strategy space and allows for mixed history-independent strategies. That this is not the case follows from the following theorem:

Theorem 6.3. *Let $\hat{\Gamma}$ be a game representing a voting scheme in which the players are allowed to select only pure history-independent strategies. Let σ be an equilibrium profile for this game. Let Γ be a game obtained from $\hat{\Gamma}$ by extending the strategy space and allowing mixed and history-dependent strategies. Under these conditions, σ is still an equilibrium profile for Γ.*

Proof. Suppose player i can deviate from σ_i in Γ, against σ_{-i}. Then, he can benefit by choosing an appropriate pure-strategy best reply τ_i. This strategy may be history-dependent, so we define another strategy τ_i' by

$$\tau_i'(t, F^{t-1}) = \begin{cases} \tau_i(t, h^t(\sigma_{-i}, \tau_i)), & \text{if } F^{t-1} = F^{t-1}(\sigma_{-i}, \tau_i), \\ \emptyset, & \text{otherwise.} \end{cases} \tag{6.3}$$

Strategy τ_i' is pure, history-independent, because it selects the same set at stage t whenever the set of voters is $F^{t-1}(\sigma_{-i}, \tau_i)$ and selects the same (empty set), otherwise. Moreover, the path under

(σ_{-i}, τ_i') coincides with the path under (σ_{-i}, τ_i), thus yielding the same stream of members. We have proved that there exists a pure, history-independent strategy τ_i' that yields player i more than σ_i against σ_{-i}. This contradicts the fact that σ was an equilibrium point in $\hat{\Gamma}$. ∎

References

Aumann, R. J., *Acceptable points in general cooperative n-person games*, Contribution to the Theory of Games IV, Annals of Mathematics, Study 40, Princeton University Press, Princeton, N. J., 1959, pp. 287–324.

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

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

Kuhn, H. W., *Extensive games and the problem of information*, Contributions to Game Theory II (H. W. Kuhn and W. Tucker, eds.), Annals of Mathematics Studies, 28, Princeton University Press, Princeton, NJ, 1953, pp. 193–216.

Selten, R., *Reexamination of the perfectness concept for equilibrium points in extensive games*, International Journal of Game Theory **4** (1975), 25–55.