Social Choice and Welfare manuscript No.
(will be inserted by the editor)

Some chance for consensus

Voting methods for which consensus is an equilibrium

Jobst Heitzig · **Forest W. Simmons**

Final draft / Accepted: 22 October 2010
Published version: DOI 10.1007/s00355-010-0517-y, www.springerlink.com/content/5xr0220678805288

Abstract We introduce the following basic voting method: Voters submit both a "consensus" and a "fall-back" ballot. If all "consensus" ballots name the same option, it wins; otherwise a randomly drawn "fall-back" ballot decides. If there is one potential consensus option that everyone prefers to the benchmark lottery which picks the favourite of a randomly drawn voter, then naming that option on all "consensus" ballots builds a very strong form of correlated equilibrium. Unlike common consensus procedures, ours is not biased towards the status quo and removes incentives to block consensus. Variants of the method allow for large groups, partial consensus, and choosing from several potential consensus options.

Keywords consensus decision making · voting method · fall-back method · benchmark · lottery · random ballot · strong correlated equilibrium

Mathematics Subject Classification (2000) 91B14 · 91B12 · 91B08 · 91A80

1 Introduction

A major issue with consensus decision making is the question of what happens when no consensus can be reached, e. g. when someone (or, in case of *partial* consensus decision-making, a sufficiently large part of the group) "blocks". If in this case the issue is "laid down" and the status quo prevails, then all who favour that option have

Jobst Heitzig
Potsdam Institute for Climate Impact Research, Transdisciplinary Concepts and Methods
P. O. Box 601203, 14412 Potsdam, Germany
E-mail: heitzig-j@web.de

Forest W. Simmons
Liberal Arts & Mathematics Division, Portland Community College
Cascade Campus, 705 N. Killingsworth Street – TH 220, Portland OR 97217
E-mail: fsimmons@pcc.edu

incentives to block. If, as is done often in practice, some common form of voting is used as a *fall-back method* when consensus cannot be reached within a given time frame (e. g., Saint and Lawson 1994), then all who prefer the (expected) result of the fall-back method over the potential consensus option have incentives to block. From a game-theoretic point of view, the combined procedure (seeking consensus and using the fall-back method as needed) then often becomes equivalent to using only the fall-back method in the first place, making it unlikely for rational agents to reach any consensus different from the status quo and from the majority's favourite.

We begin this paper with suggesting a radically different type of fall-back method which will give all agents the right incentives to cooperate rather than block in situations where a potential consensus option exists. The suggested fall-back method is *Random Ballot,* that is, choosing the *expressed* favourite of a randomly chosen agent, a method known mostly for its unique property of being *strategy-free* (Gibbard 1977). Our notion of "potential consensus" is a pragmatic one: Each option qualifies which is preferred by everyone to some "benchmark". Obviously, we cannot use as this benchmark any option that is favoured by any agent, including the status quo option. Rather, our benchmark is the lottery whose result is the *true* favourite of a randomly chosen agent. Before stating things more formally, let us look at an example:

Example 1 *A body of ten must choose between options A (the status quo), B (some "main" motion), and C (an "amended" motion), where six favour A, four favour B, and all consider C almost as desirable as their favourite.*

If either the status quo (A) or the majority's favourite (also A) is known to be the result when no consensus will be reached, the six have incentives to make sure this happens indeed, and will block a proposed consensus C. But if the fall-back method is Random Ballot, agents will face a choice not between C and A, but between getting C for sure and getting A or B with 60% and 40% probability, respectively. If all prefer C to this "benchmark lottery", as assumed, they have incentives to agree when C is proposed as consensus.

That is, in situations where a potential consensus option exists, the effect of using Random Ballot as a fall-back method is to help the consensus to be found and realized. This is because Random Ballot "levels the playing field" by distributing the decision power in a completely egalitarian way, rather than giving it to a majority or to the proponents of the status quo, and hence all agents have an interest in taking into account all other's preferences. In addition, if agents have *procedural preferences* (as in Hansson 1996) in addition to those concerning the options only, e. g. for belonging to the "winning" coalition or against randomness, they will try to reach a consensus even more when facing the suggested fall-back method.

The paper is organized as follows. In Sec. 2, we formally define this group decision procedure as a basic *voting method* and note some properties, before studying its strategic implications in a certain game-theoretic framework in Sec. 3. We will assume that voters can communicate before voting but cannot enter binding contracts, so our framework is a non-cooperative game-theory with correlated strategies where the voters use some correlation device or mediator to coordinate their actions, as first introduced by Aumann (1974). We will see that in our situation, any solution concept that considers only the possibility of deviations by *individual* voters will leave us with far too large a number of equilibria, only some of which lead to consensus. Therefore, we will then take into account that coalitions might plan for correlated deviations from the correlated strategy. There are a number of existing solution concepts for that setting which can be classified according to (i) whether they allow coalitions to plan any deviation strategies before the correlation device is applied *(ex ante)* or after that *(ex post),* and (ii) whether they can use a new correlation device or mediator for a planned deviation or not.

For simplicity's sake, we will here utilize an *ex ante* version *with new devices* from Moreno and Wooders (1996), which we feel fits a typical voting situation quite well, but adapt it to a more general utility model than the usual von Neumann–Morgenstern model.[1] Our main result is that under this adapted solution concept of *very strong correlated equilibrium,* the voting method introduced here singles out only a small number of easily identified equilibria, each of which either realizes a potential consensus option or, if none such exists, the benchmark lottery.

In Sec. 4, we then shortly present some variants of the method which work better with large groups or allow for partial consensus and choosing from several potential consensus options, before discussing the performance of our approach in simulations in Sec. 5 and concluding with some outlook.

2 The basic method

Throughout, we assume that a finite set $N=\{1,\dots,n\}$ of $n>1$ *agents* or *voters* wants to choose exactly one out of a finite set X of $k>1$ mutually exclusive *options* or *alternatives,* including the *status quo option* x_0 if such exists. We define two types of *ballots* (which we will modify later): On a *consensus ballot,* a voter $i\in N$ marks exactly one option $c(i)\in X$ as "consensus", while on a *fall-back ballot,* she marks exactly one option $f(i)\in X$ as "favourite". Let us define the *fall-back lottery* p_f and the *sure-thing lottery* p_x by putting for all $x,y\in X$, $x\neq y$:

$$p_f(x)=|\{i\in N: f(i)=x\}|/n,\quad p_x(x)=1,\ p_x(y)=0.$$

Our basic voting method then assigns each option x a winning probability $p_{\text{basic}}(x)$ as follows:

Voting method 1 (Full consensus or Random Ballot fall-back)
Each voter secretly submits a consensus ballot and a fall-back ballot. If some option is marked on all consensus ballots, that option wins. Otherwise, the option marked on a randomly drawn fall-back ballot wins. Formally: If $c(i)=x$ for some $x\in X$ and all $i\in N$, then $p_{\text{basic}}=p_x$, otherwise $p_{\text{basic}}=p_f$.

[1] Alternatively, one could base the analysis on the *ex ante* version *with new devices* from Milgrom and Roberts (1996), the *ex ante* version *without new devices* from Ray (1996) or one of the *ex post* versions *with new devices* from Einy and Peleg (1995), Ray (1998), or Bloch and Dutta (2009). A discussion of the differences is, however, beyond the scope of this paper.

2.1 Properties

Despite being *anonymous* (treating all voters equally), the above method is also *neutral* (it also treats all options equally). In particular, it is not biased towards or against the status quo like other procedures for consensus decision-making are. Furthermore, it is *monotonic* in the sense that if one voter i changes her choice of $c(i)$ or $f(i)$ to $c(i) = x$ or $f(i) = x$, respectively, then this cannot decrease x's winning probability. As we will see below, it is also *partially strategy-free* in the sense that if a voter strictly prefers some option to all others, she has no incentives to mark a different option as "favourite".

Unlike most common voting methods, ours is *non-deterministic* in the sense that it potentially uses randomness not only for rare tie-breaking purposes. And it is *non-majoritarian* in the sense that a majority might not have a way of enforcing a specific outcome. Instead, any coalition M of voters can make sure a specific option x gets a winning probability at least as large as the coalition's relative size, no matter what the remaining voters do: By marking x on both ballots, they ensure that $p_{\text{basic}}(x) \geqslant |M|/n$. That is, the method gives *equal power of probability allocation* since each individual voter or coalition can control her or its proportional share of the winning probability. Still, when a potential consensus option can be found, rational voters will usually not exercise this power but will rather avoid the resulting lottery and support the consensus, as we will see in the following section in more detail.

3 Game-theoretic analysis

3.1 First game, before eliminating dominated actions

For any non-empty set A, let $\Delta(A) = \{p \in [0,1]^A : \sum_{a\in A} p(a) = 1\}$ denote the set of *probability distributions* or *lotteries* over A. We will first interpret the above voting method as an *n-player game* G_1 given in normal form, whose sets of *players* and *outcome lotteries* are N and $\Delta(X)$, and whose set of *player actions* (or *pure strategies*) is $A_i = A_0 = X \times X$ for each $i \in N$, where an element $(x,y) \in A_0$ is interpreted as marking x as "consensus" and y as "favourite". Let c_a and f_a denote these two components of a, i. e., $a = (c_a, f_a)$ for all $a \in A_0$. Note that, later in the analysis, we will also study a second game G_2 defined by smaller sets A_i where some dominated actions have been eliminated.

3.2 Preference model and assumptions

Our analysis is not restricted to any particular kind of utility model for the players. Rather, we explicitly formulate some conditions on the voters' preferences over certain lotteries over X. Let $\ell \succcurlyeq_i \ell'$ denote the fact that player i considers lottery $\ell \in \Delta(X)$ *equivalent or preferable* (aka *weakly preferable*) to lottery ℓ'. *Strict preference* $\succ$ is defined by $\ell \succ_i \ell'$ iff $\ell \succcurlyeq_i \ell'$ but not $\ell' \succcurlyeq_i \ell$.

Unlike in the von Neumann–Morgenstern expected utility model ("vNM" in the sequel), we only require a quite mild form of "rationality": we firstly assume that (1) weak preference $\succcurlyeq_i$ is *reflexive* and *transitive* (i. e., a *quasi-order*), but not necessarily *complete,* which would require $\ell \succcurlyeq_i \ell'$ or $\ell' \succcurlyeq_i \ell$ for all ℓ, ℓ' (for possible justifications of incomplete preferences, see e. g. Mandler 2005). Thus we have to distinguish *equivalence* (or *indifference*) $\sim$, defined by $\ell \sim_i \ell'$ iff both $\ell \succcurlyeq_i \ell'$ and $\ell' \succcurlyeq_i \ell$, from *incomparability* (or *indecisiveness*) $\|$, defined by $\ell \|_i \ell'$ iff neither $\ell \succcurlyeq_i \ell'$ nor $\ell' \succcurlyeq_i \ell$. In particular, one has to be careful not to confuse $\ell' \not\succcurlyeq_i \ell$ (a non-existing weak preference) with $\ell \succ_i \ell'$ (an strong preference in the other direction).

To be able to define the benchmark lottery and a notion of potential consensus, we assume in addition that (2) each voter $i \in N$ has a *true favourite* $f_0(i) \in X$ such that $\lambda f_0(i) + (1-\lambda)\ell \succ_i \lambda x + (1-\lambda)\ell$ for all $x \in X \setminus \{f_0(i)\}$, $\ell \in \Delta(X)$ and $\lambda \in (0,1]$. The *benchmark lottery* $p_0 \in \Delta(X)$ is then the fall-back lottery p_f of our basic method that arises if all voters mark their true favourites, that is,

$$p_0(x) = |\{i \in N : f_0(i) = x\}|/n.$$

A *potential consensus option* is then any option $x \in X$ that everyone weakly and at least one voter strictly prefers to this benchmark lottery, i. e., for which $x \succ_N p_0$.

Finally, let us assume two consistency conditions regarding comparisons of the benchmark lottery with other lotteries: (3) a strict preference between some option x and the benchmark lottery p_0 is retained when both are mixed with some other lottery, that is, $p_0 \succ_i x$ implies $\lambda p_0 + (1-\lambda)\ell \succ_i \lambda x + (1-\lambda)\ell$, and $x \succ_i p_0$ implies $\lambda x + (1-\lambda)\ell \succ_i \lambda p_0 + (1-\lambda)\ell$, for all $i \in N$, $x \in X$, $\ell \in \Delta(X)$ and $\lambda \in (0,1]$.

And (4): when a voter prefers to replace in a mixed lottery some amount of some lottery $p' \in \Delta(X)$ by the same amount of p_0, she must also prefer p_0 to at least one of the possible outcomes of p'; formally: for all $i \in N$, $p', \ell \in \Delta(X)$ and $\lambda \in (0,1]$, if $\lambda p_0 + (1-\lambda)\ell \succ_i \lambda p' + (1-\lambda)\ell$, then $p_0 \succ_i x$ for some $x \in X$ with $p'(x) > 0$.

Note that in the vNM model, all these assumptions (1)–(4) are implied by the simple condition that each voter has a unique utility-maximizing option $f_0(i)$, whereas in non-expected utility models, e. g. rank-dependent utility, this need not be the case (for an overview of non-expected utility, see e. g. Machina 1987).

Later on, we will need the notion of a *(proper) coalition,* which is any non-empty (proper) subset M of N. The relations $\succcurlyeq$, $\succ$, $\sim$, $\|$ are extended to coalitions in the canonical way by writing $\ell \succcurlyeq_M \ell'$ or $\ell \sim_M \ell'$ iff $\ell \succcurlyeq_i \ell'$ or $\ell \sim_i \ell'$ for all $i \in M$, respectively, and writing $\ell \succ_M \ell'$ or $\ell \|_M \ell'$ iff $\ell \succcurlyeq_M \ell'$ but not $\ell' \succcurlyeq_M \ell$, or neither $\ell \succcurlyeq_M \ell'$ nor $\ell' \succcurlyeq_M \ell$, respectively. In particular, $\ell \succ_N \ell'$ denotes the fact that ℓ' is *weakly Pareto-dominated* by ℓ, i. e., all voters prefer ℓ to ℓ' weakly and at least one voter strictly.

3.3 Solutions when only individuals can plan deviations

Let us shortly look at some classical solution concepts, suitably adapted for the above preference model, that consider possible deviations by individual voters only.

A *pure strategy equilibrium (PSE)* is a tuple $a \in \prod_{i \in N} A_i$, prescribing a specific voting action of all players, such that no player i strictly prefers to vote differently. Formally: a is a PSE iff there is no $i \in N$ and $b_i \in A_i$ such that $p_{\text{basic},b} \succcurlyeq_i p_{\text{basic},a}$, where $b = (a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_N)$, and $p_{\text{basic},a}$ is the outcome lottery of the basic method (as defined above) when the voters vote according to a.

It is easy to see that for the basic method, there are three types of PSE: if $n > 2$, a is a PSE if and only if one of the following holds:

1. All voters mark the same x on the consensus ballot and no-one strictly prefers the fall-back lottery to x, i. e., for some $x \in X$ and all $i \in N$, $c_{a_i} = x$ and $p_{f,a} \not\succ_i x$.
2. All mark their true favourite, all but one voter j mark the same x on the consensus ballot, and j does not strictly prefer x to $p_{f,a}$, i. e., $f_{a_i} = f_0(i)$ for all $i \in N$, and for some $x \in X$ and some $j \in N$, $x \not\succ_j p_{f,a}$ and $c_{a_i} = x$ for all $i \in N \setminus \{j\}$.
3. All mark their true favourite, and each x is marked by at most $n-2$ voters on the consensus ballot.

When there is a potential consensus option, the corresponding PSE of the first above type elects it, but the quite numerous PSEs of the other types fail to elect a consensus, hence we need a more restrictive solution concept. In the vNM model, both the solution concepts of *Nash equilibrium* and of *correlated equilibrium* (Aumann, 1974) are even more general than PSE, and adapting them for our more general preference model would also only enlarge the set of equilibria instead of reducing it. So, the unwanted equilibria of type 2 and 3 above only disappear when we strengthen the solution concept by not only allowing for correlated strategies but also for coordinated deviations by coalitions.

3.4 Correlated strategies and deviations

Our framework of correlated strategies for coalitions can be formalized like this. For a coalition M, we call $A_M = \prod_{i \in M} A_i$ the set of *coalition actions* for M, with $a(i) \in A_i$ for each $a \in A_M$ and $i \in M$, and $S_M = \Delta(A_M)$ is the set of *coalition strategies* for M. Each $a \in S_i = S_{\{i\}}$ is called a *mixed strategy* for player i.

A coalition action $a \in A_N$ for the *grand coalition N* prescribes a specific voting behaviour of all players, while a coalition strategy $s \in S_N$ for N encodes a possibly correlated random process by which voters may choose these actions. One way to achieve the correlation is that some trusted *mediator* (acting as a special form of "correlation device") draws an action profile a from the distribution s and privately tells each player i to take action a_i. If each player follows this advice, we say the grand coalition strategy $s \in S_N$ is *adopted.* The voting method then produces an outcome lottery $p_s \in \Delta(X)$ defined by

$$p_s(x) = \sum_{a \in A_N} s(a) p_{\text{basic,a}}(x) = \sum_{a \in A_x} s(a) + \sum_{a \in A_f} s(a) n_x(a)/n,$$

where the subsets $A_x = \{a \in A_N : c_a(i) = x \text{ for all } i \in N\}$ and $A_f = A_N \setminus \bigcup_{x \in X} A_x$ contain the grand coalition's *consensus-x actions* and *non-consensus actions,* respectively, and $n_x(a) = |\{i \in N : f_a(i) = x\}|$ is the number of voters marking x as "favourite" under a.

Following Moreno and Wooders (1996), we assume that before the mediator recommends an action *(ex ante),* any coalition M may plan to deviate from his advice by agreeing on a *deviation scheme*[2] δ that uses a new mediator (a "new device"). This deviation scheme will be formalized here as a function which assigns to each coalition action $a \in A_M$ a new, possibly correlated, coalition strategy $\delta(\cdot|a) \in S_M$ which they plan to use instead of a should the mediator tell them to use a. To apply the deviation scheme, each $i \in M$ would send the advice a_i received from the mediator privately to the new mediator who then draws a new coalition action b according to $\delta(\cdot|a)$ and privately sends back b_i to i, who finally uses b_i instead of a_i. If all $i \in M$ follow the advice of the new mediator, the original strategy $s \in S_N$ for the grand coalition is transformed into a new, effective strategy that we denote by $s/\delta \in S_N$. It can be written as $(s/\delta)(b \star a') = \sum_{a \in A_M} \delta(b|a)s(a \star a')$ for all $b \in A_M$ and $a' \in A_{N\setminus M}$, where $a \star a'$ denotes the grand coalition action that is the combination of the coalition action a and the action a' of the coalition's complement.

3.5 Dominated actions, partial strategy-freeness, and the second game

Before stating our solution concept, we first eliminate some dominated actions from the game G_1 to get a simplified game G_2. We call a coalition action *a dominated* if the coalition has incentives to replace a by some other action no matter what the grand coalition strategy is. Formally: $a \in A_M$ is *dominated by* $b \in A_M$ iff $p_{s/\delta_{a\to b}} \succcurlyeq_M p_s$ for all $s \in S_N$, and $p_{s/\delta_{a\to b}} \succ_M p_s$ for at least one $s \in S_N$, where $\delta_{a\to b}$ is the deviation scheme defined by $\delta_{a\to b}(b|a) = \delta_{a\to b}(c|c) = 1$ and $\delta_{a\to b}(b'|a) = \delta_{a\to b}(c'|c) = 0$ for all $c \in A_M \setminus \{a\}$, $b' \in A_M \setminus \{b\}$, and $c' \in A_M \setminus \{c\}$.

Our first result is that the voting method is strategy-free on the fall-back ballot:

Lemma 1 *The following holds in the game G_1 under the above assumptions:*

(a) Each individual voter has incentives to mark their true favourite no matter what the grand coalition strategy is. Formally: Let $i \in N$, $x, y \in X$ and $y \neq f_0(i)$. Then (x, y) is dominated by $(x, f_0(i))$.

(b) If a deviation scheme δ makes a difference for some coalition M but requires that some voter sometimes marks a different option as "favourite", then each such voter has incentives to deviate further from δ by marking her true favourite anyway. Formally: If $s \in S_N$, $M \subseteq N$, $a \in A_f$, $i \in M$, and δ is a deviation scheme for M such that $p_{s/\delta} \not\sim_M p_s$, $(s/\delta)(a) > 0$, and $f_a(i) \neq f_0(i)$, then $p_{s/\delta/\delta'} \succ_i p_{s/\delta}$, where δ' is the deviation scheme for $\{i\}$ defined by $\delta'\big((x, f_0(i))|(x,y)\big) = 1$ for all $x \in X$.

Proof Let $a = (x, y)$, $a' = (x, f_0(i))$, and $s' = s/\delta_{aa'}$. By definition, for all $s \in S_N$ we have $p_{s'}(f_0(i)) \geqslant p_s(f_0(i))$, $p_{s'}(y) \leqslant p_s(y)$, and $p_{s'}(z) = p_s(z)$ for all $z \in X \setminus \{y, f_0(i)\}$, hence $p_{s'} \succcurlyeq_i p_s$ because of (2). Also, if $s(b) > 0$ for some $b \in A_f$, then $p_{s'}(f_0(i)) > p_s(f_0(i))$ and $p_{s'}(y) < p_s(y)$, and thus $p_{s'} \succ_i p_s$ because of (2). Since such an s exists because of $n > 1$, a is dominated by a'. The second claim follows analogously. □

In other words, under mild rationality assumptions, one can expect that the fall-back lottery p_f equals the benchmark lottery p_0. Because of this result, we restrict our further analysis to the case where all voters indeed mark their true favourite. That

[2] This concept was called a "feasible deviation" in Moreno and Wooders (1996).

is, we study a second game G_2 where player i's action set is now reduced to the set $A_i = \{(x, f_0(i)) : x \in X\} \subset A_0$ of *sincere actions*. Consequently, all sets A_x are now singletons containing only one consensus-x action which we will now denote by a_x.

3.6 Solution concept and main result

We call a grand coalition strategy $s \in S_N$ a *very strong correlated equilibrium* if no coalition has incentives to deviate from it, that is, if no $M \subseteq N$ has any deviation scheme δ with $p_{s/\delta} \succ_M p_s$. Note that a very strong correlated equilibrium does not even allow for deviations in which only one member of the coalition has a strict preference, whereas most common concepts of equilibrium, including that in Moreno and Wooders (1996), only care about *improving* deviations in which all members of the coalition *strictly* prefer the new outcome lottery. Also, we will see that we do not have to weaken this concept to a "coalition-proof correlated equilibrium" in order to get nice existence results, as was done in Moreno and Wooders (1996) by considering only *self-enforcing* deviation schemes in which no sub-coalition has incentives to further deviate. In other words, our solution concept singles out equilibria that are stable even when coalitions can plan for deviations that are not self-enforcing but are somehow else made "binding".[3]

Our main result makes use of the fact that for all $\ell \in \Delta(X)$ there is a strategy $s_\ell \in S_N$ such that $p_{s_\ell} = \ell$, defined by $s_\ell(a_x) = \ell(x)$ for $x \in X$ and $s_\ell(a) = 0$ for all $a \in A_f$. In other words, the grand coalition can bring about any outcome lottery by using a totally correlated strategy which consists of the corresponding mixture of consensus actions. Nicely, it turns out that all very strong correlated equilibria are of this form:

Theorem 1 *Assume that agents are restricted to sincere actions and at least one potential consensus option exists, i. e. $y \succ_N p_0$ for some $y \in X$. Then an $s \in S_N$ is a very strong correlated equilibrium if and only if it is a weakly Pareto-undominated mixture of consensus-x actions with options x that no-one strictly prefers the benchmark lottery to.*

Formally: $s \in S_N$ is a very strong correlated equilibrium of the second game G_2 iff $\ell \succ_N p_s$ for no $\ell \in \Delta(X)$, and, for all $a \in A_N$ with $s(a) > 0$, there is some $x \in X$ with $a = a_x$ and such that $p_0 \succ_i x$ for no $i \in N$.

Proof For the grand coalition, there is a deviation scheme δ with $p_{s/\delta} \succ_N p_s$ iff $\ell \succ_N p_s$ for any $\ell \in \Delta(X)$, because for all $b \in A_N$ we can put $\delta(a_x|b) = \ell(x)$ for all $x \in X$ and $\delta(a|b) = 0$ for all $a \in A_f$, so that $p_{s/\delta} = \ell$.

For a proper coalition $M \subset N$, we first assume that s is of the given form and that there is a deviation scheme δ with $p_{s/\delta} \succ_M p_s$, and choose some $i \in M$ with $p_{s/\delta} \succ_i p_s$. Since M is proper, δ can only shift probability from some consensus-x actions to some fall-back actions, where, for all such x, we have $p_0 \succ_i x$ for no $i \in N$.

[3] A deviating coalition will often be smaller, more homogeneous, and/or close-knit than the grand coalition, so that socially binding agreements can not be ruled out in deviating coalitions even when legally binding voting agreements are impossible. This is a situation somewhat similar to that in Moulin and Vial (1978).

Consequently, δ shifts outcome probability from a mixture p' of such options x to the fall-back lottery p_0, that is, $s = \lambda p' + (1-\lambda)\ell$ and $s/\delta = \lambda p_0 + (1-\lambda)\ell$ for some $\ell \in \Delta(X)$ and $\lambda \in (0,1]$. But then $p_{s/\delta} \succ_i p_s$ and (4) imply that $p_0 \succ_i x$ for one of those x, a contradiction. Hence s is a very strong correlated equilibrium after all.

On the other hand, assume that s is a very strong correlated equilibrium and let $a \in A_N$ with $s(a) > 0$. If a were in A_f, we could choose some $y \in X$ with $y \succ_N p_0$, and put $\delta(a_y|a) = \delta(b|b) = 1$ for all $b \in A_N \setminus \{a\}$, so that $p_{s/\delta} \succ_N p_0$ because of (3), in contradiction to the equilibrium assumption. In other words, the grand coalition would deviate from a by replacing it with the consensus-y action. Hence $a = a_x$ for some $x \in X$. Assume $p_0 \succ_i x$ for some $i \in N$, choose some $x' \in X \setminus \{x\}$ and put $b = (x, f_0(i)), b' = (x', f_0(i)) \in A_i$. Then $\delta_{b\to b'}$ shifts probability from the consensus-x action a to a fall-back action. Consequently, it shifts outcome probability from x to the fall-back lottery p_0, that is, $s = \lambda x + (1-\lambda)\ell$ and $s/\delta_{b\to b'} = \lambda p_0 + (1-\lambda)\ell$ for some $\ell \in \Delta(X)$ and $\lambda \in (0,1]$. Hence $p_{s/\delta_{b\to b'}} \succ_i p_s$ because of (3), again a contradiction to the equilibrium assumption. In other words, i would deviate from a by blocking. Thus $p_0 \succ_i x$ for no $i \in N$ after all. $\square$

Note that Lemma 1 and Theorem 1 imply that in case of vNM utilities, those strategies characterised here are *strong correlated equilibria* (in the sense of Moreno and Wooders 1996). In particular, the pure consensus-x actions with $x \succ_N p_0$ are *strong Nash equilibria* of game G_2 (in the sense of Aumann 1959), and, by virtue of Lemma 1 (b), they are *coalition-proof Nash equilibria* of game G_1 (in the sense of Bernheim et al. 1987).

If there is a "natural" consensus option x which is not weakly Pareto-dominated and which everyone prefers to the benchmark lottery, then it seems quite likely that all agents will indeed support this consensus. For example, in situations where we have at least partially *transferable utility,* the majority option can often be combined with side-payments ensuring that also the minorities prefer this combined option to the benchmark.[4]

4 Variants

Despite these positive theoretical results, our basic method has a lot of room for practical improvements. We will sketch some suggestive examples of practical improvement, without attempting to exhaust the possibilities. First we will see how to modify the basic method to improve its efficiency in choosing from among several potential consensus options. Then we will consider the possibility of partial consensus in the case when there is no realistic possibility of full consensus, i. e., when it is obvious

[4] If one assumes linearly transferable von Neumann–Morgenstern utilities $u_i(\ell)$ for all $i \in N$ and $\ell \in \Delta(X)$ and analyses the second game as a *coalitional* game in minimax, defensive-equilibrium, or rational-threats representation (following Myerson 1991 again), its characteristic function v is $v(M) = u_M(p_0)$ for $M \subset N$ and $v(N) = \max_{x\in X} u_N(x)$ with $u_M(x) = \sum_{i\in M} u_i(x)$, its core is $\{a \in [0,\infty)^N : \sum_{i\in N} a(i) = v(N)$ and $a(i) \geqslant u_i(p_0)\}$, and its Shapley value and nucleolus is $\phi_i(v) = u_i(p_0) + \big(v(N) - u_N(p_0)\big)/n$. In other words, the core allocations are those which adopt the option that maximizes total utility and redistribute the latter so that all are no worse off than with the benchmark, and the focal allocation distributes this excess utility equally.

from the outset that there is not even one alternative that is preferred unanimously over the fall-back option, but there still might be a near unanimous consensus option.

One way to automatically (i. e., without negotiations on the side) decide on a unanimous consensus from among several viable alternatives is to supplement the consensus and fall-back ballot with a third, *ratings ballot* on which each voter i assigns each option x a real-valued rating $r_i(x)$. This is then used as follows:

Voting method 2 *Each voter submits a consensus, a fall-back, and a ratings ballot. Let x be the marked option on a randomly drawn consensus ballot. If on each ratings ballot, x is rated at or above the expected rating of the fall-back lottery, $\sum_y p_f(y) r_i(y)$, then x wins. Otherwise, a fall-back ballot is drawn to decide the outcome.*

This method has two nice strategic properties. First of all, for each voter with vNM utilities it is optimal to specify these as her ratings, i. e., $r_i(x) = u_i(x)$. This is because the ratings do not influence the choice of x or the probabilities in p_f but are only used to decide between a given x and p_f. Most other known methods with such a "revelation" property (e. g. the famous one in Tideman and Tullock (1976) based on the demand-revealing process by Vickrey, Clarke, and Groves) involve some kind of side payments.

Second, for each voter knowing the set C of options preferred by everyone to p_f, it is also optimal to mark the most preferred member of C on the consensus ballot. In particular, with full information and non-empty C, the result will be a lottery among the members of C with winning probabilities proportional to the number of voters preferring the respective option.[5]

In the next subsection, we will also present an alternative that avoids the somewhat cumbersome ratings ballots, in which voters may mark a set of "agreeable consensuses".

Another simple but iterative variant that could be used in assemblies was suggested by one anonymous referee: going through all options in random order, ask voters for their agreement to each option. The first option all voters agree to wins; when no such option exists, Random Ballot is used. Assuming rational vNM voters, an analysis of this stage game considering subgame-perfect pure strategies reveals that when a potential consensus exists, the outcome of the whole process including the initial random ordering is a certain mixture of all weakly Pareto-undominated potential consensus options. If the random initial ordering is determined by drawing fall-back ballots, this variant even becomes clone-proof (see the next subsection).

Next, consider the case where there may be a good possibility for a near consensus option, but where unanimous consensus appears unlikely. No doubt the alert reader has already anticipated some of the possibilities of using thresholds to cope with this difficulty. For example, one can slightly relax the requirement of unanimity on

[5] However, the method is not fully strategy-free since the fall-back ballot is now strategic, e. g., incentives can arise to exaggerate and report more extreme favourites than the true favourites in order to raise the approval for a wanted potential consensus. Still, such strategic behaviour would at least be detectable by comparing the reported "favourite" with the reported ratings if both are required to be submitted together on a combined ballot.

the above methods. Similarly, a low threshold of support can be applied to the fall-back method to filter out options considered to be dangerous, if there is no other precautionary filter up front.

Below, we suggest a method that allows each voter to specify an individual threshold for how much partial consensus an option must have before they are willing to support that option as a potential consensus. This method avoids the potential conflict engendered by the arbitrariness of a choice of threshold for the whole electorate and retains the property of equal power of probability allocation.

4.1 TAPF voting

The following is a method designed to be applicable in situations where none, one, or several potential consensus options of varyingly broad appeal exist:

On a *TAPF ballot,* a voter $i \in N$ specifies a percentage $t(i) \in (50, 100]$ as her "threshold for consensus", marks one or more options as "agreeable consensus" (which we code as a subset $C(i) \subseteq X$), marks one of these options as "preferred consensus" (coded as $c(i) \in C(i)$), and marks one option $f(i) \in X$ as "favourite".

Given a coalition $M \subseteq N$ and some member $i \in M$, M is called *feasible for i* iff the number of members $j \in M$ whose "preferred consensus" is among those marked as "agreeable" by voter i is at least $t(i)$ percent of all n voters, that is, if $|\{j \in M : c(j) \in C(i)\}| \geqslant t(i)n/100$. Denoting this condition by $\varphi(M,i)$, coalition M is then called *feasible* iff it is feasible for each of its members, i. e., if $\varphi(M,i)$ for all $i \in M$. Note that since any union of feasible coalitions again fulfils this requirement, there is a unique (but maybe empty) largest feasible coalition $M_{\max}$ which contains all others: $M_{\max} = \bigcup\{M \subseteq N : \varphi(M,i) \text{ for all } i \in M\}$.

Voting method 3 (TAPF voting) *Each voter submits a TAPF ballot, and one of them is drawn at random. If it belongs to the largest feasible coalition $M_{\max}$, its "preferred consensus" wins, otherwise its "favourite" wins. Formally, the winning probabilities are $p_{\mathrm{TAPF}}(x) = \big(|\{i \in M_{\max} : c(i) = x\}| + |\{i \in N \setminus M_{\max} : f(i) = x\}|\big)/n$.*

Example 2 *A body of ten must choose between options A (the status quo), B (the motion), C (an amended motion), and D (the status quo plus some monetary compensation), where the first six have a ranking of $A \succ D \succ C \succ B$, the other four have $B \succ C \succ D \succ A$, and all but the last voter prefer both C and D to the benchmark lottery of 60% A and 40% B. With TAPF voting, in order to make sure their preferred potential consensus gets a fair winning chance, the first nine specify a threshold of 90%, and mark D or C as their "preferred consensus" and both as "agreeable consensuses", while marking A or B as "favourite",respectively. In this way, the result is 60% D, 30% C, and 10% B.*

Let us finally remark that the last method is not only anonymous, neutral, (in a suitable sense) monotonic, and gives equal power of probability allocation, but is also *clone-proof* in the following sense: Suppose some option x is replaced by a set of nearly indistinguishable options $x_1, \ldots, x_m$, and all ballots are changed so that (i) when x was marked as "agreeable consensus", all x_i are now thus marked, and (ii) when x was marked "preferred consensus" or "favourite", then one of the x_i is now thus marked. Then these alterations leave the winning probabilities of all options except x unchanged (a similar property called "composition consistency" is studied in Laffond et al. 1996).

5 Performance in simulations

To assess the typical performance of the suggested kind of voting method both from a more egalitarian and a more utilitarian perspective, we performed Monte Carlo simulations using several common spatial models of utility (see, e. g., Enelow and Hinich (1984) for the underlying theory).

A varying number k of options x were assumed to occupy points $\hat{x}$ in some metric space (X,d) (the *policy space*), and a varying number n of voters i to possess *ideal points* $\hat{\imath} \in X$ and to assign vNM utilities $u_i(x)$ equaling either the negative distance $-d(\hat{\imath},\hat{x})$ ("linear", risk-neutral model), or the negative quadratic distance $-d(\hat{\imath},\hat{x})^2$ ("quadratic", risk-averse model), or the reciprocal distance $1/d(\hat{\imath},\hat{x})$ (risk-acceptant model). Points $\hat{x},\hat{\imath}$ were drawn independently at random from several distributions: the one-dimensional standard uniform, standard Cauchy, standard normal, or standard log-normal distribution, a three-dimensional standard normal ("multi-normal"), and a symmetric one-dimensional normal mixture with density $\varphi(y) \propto \exp(-(y-2)^2/2) + \exp(-(y+2)^2/2)$.

In each simulated situation, the benchmark lottery p_0 and two options were determined: the *utilitarian solution* x_u that maximized total utility $t(x) = \sum_i u_i(x)$, and the broadest potential consensus option x_c which the largest subset of voters preferred to p_0: $x_c = \operatorname{argmax}_x |M_x|$ with $M_x = \{i \in N : u_i(x) > u_i(p_0)\}$. It was then assumed that the share of the winning probability of this *cooperating coalition* M_{x_c} was assigned to x_c while the share of the remaining voters i was assigned to their respective favourites $f(i)$. This results in the *partial consensus lottery* $p_{\text{sim.}}(x_c) = |M_{x_c}|/n$, $p_{\text{sim.}}(x) = |\{i \in N \setminus M_{x_c} : f(i) = x\}|/n$ for $x \neq x_c$. This would for example arise under TAPF voting when all voters in M_{x_c} mark x_c as their sole "agreeable consensus" and specify $100|M_{x_c}|/n$ as their threshold.

Overall, in about 57 percent of all situations, x_c and x_u were the same, i. e., the utilitarian solution also was the broadest potential consensus option. Also, in only about 44 percent of all situations, the Condorcet winner, i. e., the option that was preferred to each other option by some majority, was equal to x_c, while in more than half of all cases, a different option turned out to be a broader potential consensus. Moreover, in about three out of four situations, the value $N - |M_{x_2}|$ for the *second* broadest potential consensus option x_2 was at most twice as larges as for x_c, meaning that very often more than one good consensus option existed.

In Table 1, we report several performance measures for different combinations of k options, n voters, spatial distribution, and utility model, each from 1000 independent simulations: (i) the average relative size $C = |M_{x_c}|/n$ of the cooperating coalition; (ii) the average proportions $P_c = |\{i \in N : u_i(p_{\text{sim.}}) > u_i(p_0)\}|/n$, $P_u = |\{i \in N : u_i(x_u) > u_i(p_0)\}|/n$ of voters "profiting" from the partial consensus lottery or the utilitarian solution, respectively, as compared to the benchmark lottery; and (iii) the

distribution		standard normal		multi-normal		mixed normal		standard normal		multi-normal		mixed normal	
utility model		lin.	qu.	lin.	qu.	lin.	qu.	lin.	qu.	lin.	qu.	lin.	qu.
		relative size S of the cooperating coaltion						**utilitarian efficiency E of the partial consensus lottery**					
k	n												
3	10	69	76	70	72	58	64	*21*	*33*	*25*	*29*	*16*	*29*
	100	71	77	70	73	58	64	*25*	*40*	*37*	*43*	*28*	*37*
	1000	71	77	70	73	58	64	*27*	*40*	*36*	*44*	*30*	*39*
6	10	77	90	74	80	65	79	*45*	73	52	60	*34*	63
	100	78	92	75	81	64	78	52	80	62	70	50	68
	1000	79	92	75	81	65	79	53	80	62	71	52	70
12	10	83	98	81	88	71	92	54	92	66	77	*48*	84
	100	86	98	83	89	71	91	65	94	76	83	57	86
	1000	86	98	83	90	72	91	67	95	76	85	58	86
24	10	88	100	87	95	78	98	61	96	73	86	56	94
	100	92	100	90	95	79	98	75	99	85	92	63	95
	1000	92	100	91	96	79	98	76	99	86	93	64	96
		proportion P_c of voters profiting from the partial consensus lottery						**proportion P_u of voters profiting from the utilitarian solution**					
k	n												
3	10	*41*	53	*46*	*47*	*33*	*46*	**68**	**73**	**69**	**70**	**58**	**63**
	100	64	78	73	76	60	**73**	**70**	75	70	73	58	64
	1000	69	**84**	72	76	**67**	**81**	70	75	70	73	58	63
6	10	71	**92**	75	80	66	**90**	74	87	73	78	63	79
	100	76	**98**	**79**	85	**75**	**99**	76	90	74	81	64	78
	1000	77	**98**	**79**	85	**75**	**99**	77	89	74	81	65	79
12	10	78	99	**84**	**90**	**70**	**98**	77	96	78	85	65	92
	100	82	100	**88**	92	**75**	**100**	84	97	82	89	70	91
	1000	84	100	**88**	93	74	**100**	86	98	83	90	72	91
24	10	81	100	**89**	96	72	100	80	99	83	92	68	98
	100	87	100	94	97	76	100	89	100	90	95	77	98
	1000	88	100	94	97	77	100	91	100	91	95	79	98

Table 1 Performance measures from simulations, as defined in the text (rounded average percentages). Values below 50 are in italics, proportions P_c or P_u at least 10 percent larger than their respective comparison value P_u or P_c are in boldface.

average *utilitarian efficiency* $E = (t(p_{\text{sim.}}) - t(p_0))/(t(x_u) - t(p_0))$ of the partial consensus lottery, i. e., the ratio between the total utility of the consensus lottery and the utilitarian solution after subtracting that of the benchmark lottery.

If the number of options is very small, the cooperating coalition size S is only around two thirds on average, and the partial consensus lottery does not perform much better than the benchmark lottery when judging by our utilitarian efficiency measure E. For larger k, however, both S and E grow and seem to approach one for most spatial distributions, and also did not change substantially for $n \geqslant 10,000$. From the more egalitarian perspective, we see that, on the other hand, even for small n, the proportion of voters profiting from the partial consensus lottery is larger than that profiting from the utilitarian solution (or, likewise, from the Condorcet solution) on average. The overall picture is basically the same for all studied spatial distributions (the uniform, Cauchy, and log-normal have been omitted in the table since the results were quite similar to those for the standard normal) and all three utility models (the reciprocal being omitted in the table). Both the partial consensus lottery and the utilitarian solution perform worst in the model using reciprocal distances and best in the model using quadratic distances, which can be expected because in the latter model, for large k and n, it can easily be seen that the option nearest to the mean voter position not only maximizes total utility but must also be preferred to the benchmark lottery by most voters.

6 Conclusion

This paper's title succinctly summarizes our main idea: Some opportunities for consensus can be exploited by methods that make essential, judicious use of randomness. Random Ballot, as a benchmark of minimal fairness, provides a means of conceptual definition (if not outright detection) of potential consensus options, namely those options that are preferred by all members of the consensus seeking community over choice by Random Ballot lottery. Its use as a fall-back method provides rational incentive for the community members to adhere unanimously to a consensus option without essential reliance on appeals to community spirit, guilt, exhaustion of patience, or other psychological manipulations.

For the domain of applicability of the basic method, we demonstrated formally that no strategy, whether pure, mixed, or correlated, individual or factional, can deter a set of rational voters from electing the consensus option with certainty under conditions of perfect information about all preferences.

Moreover we hope to have convinced the reader that sufficiently careful use of chance in the design of a voting method makes it possible for voters to choose a consensus option from among several possibilities by secret, sincere ballot without agreeing ahead of time which of the options is to be "the" consensus option. In this context, "sincere" means that voters specify their true favourites as "favourite" and they specify as "agreeable consensus" only those options that they genuinely prefer to the benchmark lottery.

We have also endeavored to demonstrate how to adapt the basic method to settings where there is little potential for full consensus, while preserving the essential properties of anonymity, neutrality, monotonicity, and the newly introduced property of equal power of probability allocation.

It may offend the sensibilities of some voters that we resort to a potentially high entropy lottery like Random Ballot for more than a tie breaking role in these methods. However, when there is a real chance for consensus, the sure result of the method will be the zero entropy lottery that elects a consensus candidate with certainty. On the other hand, when there is no potential even for partial consensus, one can argue that Random Ballot may well be the fairest alternative. In any case, we cannot usually rely on deterministic methods to choose "consensus candidates" except through intimidation or other external incentives.

In a subsequent paper we intend to further examine the theme of low entropy lotteries that give equal power of probability allocation like our basic method does. The goal will then be to maximally exploit partial consensus among sub-factions of voters to choose a lottery that maximizes an appropriate, natural amalgamation of their expected ratings of the lottery winner.

Acknowledgements We would like to thank the interdisciplinary community of the *Election Methods mailing list* (http://lists.electorama.com/listinfo.cgi/election-methods-electorama.com) for the uncounted fruitful discussions which shaped many of the ideas in this paper, as well as two anonymous referees for their numerous valuable suggestions.

References

Aumann R (1959) Acceptable points in general cooperative n-person games. In: Contributions to the theory of games IV, Princeton Univ Press, Princeton

Aumann R (1974) Subjectivity and correlation in randomized strategies. J Math Econ 1: 67–96

Bernheim B, Peleg B, Whinston M (1987) Coalition-proof Nash equilibria I: Concepts. J Econ Theory 42: 1–12

Bloch F, Dutta Bh (2009) Correlated equilibria, incomplete information and coalitional deviations. Games Econ Behav 66,2: 721–728

Einy E, Peleg B (1995) Coalition-proof communication equilibria. In: Barnett WA (ed.) Social choice, welfare, and ethics. Cambridge University Press, Cambridge, MA

Enelow JM, Hinich MJ (1984) The spatial theory of voting: an introduction. Cambridge University Press, Cambridge, MA

Gibbard A (1977) Manipulation of schemes that mix voting with chance. Econometrica 45,3: 665–681

Hansson SO (1996) Social choice with procedural preferences. Soc Choice Welfare 13: 215–230

Laffond G, Lainé J, Laslier J-F (1996) Composition-consistent tournament solutions and social choice functions. Soc Choice Welfare 13,1: 75–93

Machina M (1987) Choice under uncertainty: problems solved and unsolved. J Econ Perspectives 1,1: 121–154

Mandler M (2005) Incomplete preferences and rational intransitivity of choice. Games Econ Behav 50,2: 255–277

Milgrom P, Roberts J (1996) Coalition-proofness and correlation with arbitrary communication possibilities. Games Econ Behav 17: 113–128

Moreno D, Wooders J (1996) Coalition-proof equilibrium. Games Econ Behav 17: 80–112

Moulin H, Vial JP (1978) Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon. Int J Game Theory 7,3–4: 201–221

Myerson RB (1991) Game theory: analysis of conflict. Harvard Univ Press, Cambridge, MA

Ray I (1996) Coalition-proof correlated equilibrium: a definition. Games Econ Behav 17,2: 56–79

Ray I (1998) Correlated equilibrium as a stable standard of behavior. Rev Econ Design 3: 257–269

Saint S, Lawson JR (1994) Rules for reaching consensus: a modern approach to decision making. Pfeiffer, San Diego

Tideman N, Tullock G (1976) A New and Superior Process for Making Social Choices. J Pol Econ 84,6: 1145–1159