Making most voting systems meet the Condorcet criterion reduces their manipulability

François Durand, Fabien Mathieu, Ludovic Noirie

MAKING MOST VOTING SYSTEMS MEET THE CONDORCET CRITERION REDUCES THEIR MANIPULABILITY*

François Durand
Fabien Mathieu
Ludovic Noirie

June 17, 2014

Correspondence: François Durand
23 avenue d'Italie, 75 013 Paris, France
francois.durand@inria.fr, +33 6 82 92 95 00

Word count: 10,770.

Since any non-trivial voting system is susceptible to manipulation, we investigate how it is possible to reduce the set of situations where it is manipulable, that is, such that a coalition of voters, by casting an insincere ballot, may secure an outcome that is better from their point of view. We prove that, for a large class of voting systems, a simple modification allows to reduce manipulability. This modification is *Condorcification*: when there is a Condorcet winner, designate her; otherwise, use the original rule. Our very general framework allows to do this for any voting system, whatever the form of the original ballots. Hence, when searching for a voting system whose manipulability is minimal, one can restrict to those that meet the Condorcet criterion.

JEL code: D71.

*The work presented in this paper has been carried out at LINCS (www.lincs.fr).

I Introduction

I.A Motivation

A voting system is said to be *manipulable* in a given situation if and only if a coalition of voters, by misrepresenting their preferences, may secure an outcome that they all prefer to the result of sincere voting.

For example, the French presidential election of 2002 used the two-round system. In the first round, Jacques Chirac (right) received 19.9% votes, Jean-Marie Le Pen (far right) 16.9%, Lionel Jospin (left) 16.2% and 13 miscellaneous candidates shared the rest. In the second round, Chirac won by 82.2% against Le Pen. However, according to some opinion surveys, Jospin would have won the second round against any contender.

So, there may have been a possibility of manipulation: if all voters who preferred Jospin to Chirac had voted for Jospin in the first round, then the second round might have been held between Jospin and Chirac, leading to a possible election of Jospin.

Note that the term *manipulation*, widely employed in the academic community of social choice, must be taken in a neutral, technical sense, disregarding its negative moral connotation. We will argue that this way of playing the voting game, also known as *tactical voting*, is not intrinsically better or worse than *sincere voting*; but that *manipulability*, which is the discrepancy between the results of the two, certainly is an undesirable property for a voting system.

Firstly, it challenges the outcome of the election. If we estimate that the result of sincere voting best represents the opinions of the voters, then manipulability is bad because it may lead to another outcome. On the contrary, if we estimate that the manipulated outcome may be better in terms of collective welfare, then manipulation itself is not bad[1], but manipulability still is: indeed, it makes this "better" outcome not straightforward to identify and produce. For example, if the whole population votes sincerely, it will not be achieved.

Secondly, manipulability leads to several problems for voters. Before the election, they face a dilemma: vote sincerely or try to vote tactically? In the later case, they need information about what the others will vote, which gives a questionable power to polling organizations. After the election, some sincere voters may experience regrets about the choice of their ballots, and also a feeling of injustice: since insincere ballots would have better defended their views, they may estimate that their sincere ballots did not have the impact they deserved.

Finally, it may be argued that resistance to manipulation is a prerequisite for the other desirable properties of a voting system. Indeed, most of such classical properties[2] relate ballots and candidates; but if ballots do not reflect the true opinions of voters, then these properties become hard to interpret. On the

1. About the French presidential election of 2002, proponents of Condorcet efficiency would argue that manipulation would have had a desirable effect, the victory of the presumed Condorcet winner.

2. Tideman (2006) provides a very complete overview of such classical desirable properties, such as independence of irrelevant alternatives, consistency, etc.

opposite, when there is no possibility of manipulation, the practical relevance of other properties is perfectly clear.

Unfortunately, Gibbard (1973) proved that any non-dictatorial voting system with three eligible candidates or more is manipulable. Although this result is frequently cited under the form of Gibbard-Satterthwaite theorem (Satterthwaite, 1975), which deals only with voting systems whose ballots are orders of preferences, it is worth remembering that Gibbard's fundamental theorem applies to any *game form*, where available strategies may be objects of any kind.

Once this negative result is known, the only hope is to limit the damage, by investigating in what extent classical voting systems are manipulable, and by identifying processes to design less manipulable voting systems.

To quantify the degree of manipulability of a voting system, several indicators have been defined and studied, for example by Lepelley and Mbih (1987), Saari (1990), Smith (1999), Slinko (2004) and Tideman (2006). A very common one is the manipulability rate, which is the probability that a situation is coalitionaly manipulable, under a given assumption on the probabilistic structure of the population (or *culture*). It is an important indicator because it is an upper bound for most of the others: if we could identify voting systems with close-to-zero manipulability rates in realistic cultures, then the impact of manipulability would be tolerable.

Several authors have used a theoretical approach (Lepelley and Mbih, 1987, 1994; Lepelley and Valognes, 1999; Smith, 1999; Favardin et al., 2002; Lepelley and Valognes, 2003; Favardin and Lepelley, 2006; Lepelley et al., 2008), computer simulations (Lepelley and Mbih, 1987; Pritchard and Wilson, 2007; Green-Armytage, 2011, 2014; Green-Armytage et al., 2014) or experimental results (Chamberlin et al., 1984; Tideman, 2006; Green-Armytage et al., 2014) to evaluate the manipulability rates of several voting systems, according to various assumptions about the structure of the population.

Among the studies above, some, like those of Chamberlin et al. (1984), Lepelley and Mbih (1994), Lepelley and Valognes (2003) or Green-Armytage (2011, 2014) suggest that a voting system like Instant-Runoff Voting (IRV) is one of the least manipulable voting systems known. On the other hand, authors like Chamberlin et al. (1984), Smith (1999), Favardin et al. (2002), Lepelley and Valognes (2003), Favardin and Lepelley (2006) or Tideman (2006) emit the intuition that voting systems that meet the Condorcet criterion have a general trend to be less manipulable than others.

Combining both ideas, Green-Armytage et al. (2014) introduce an alteration of IRV that meets the Condorcet criterion. Then they prove, independently of us (Durand et al., 2012), that for a large class of voting systems, making them meet the Condorcet criterion cannot improve their manipulability. The main difference between their approach and ours is that Green-Armytage et al. (2014) focus on the case where each voter has a strict total order of preferences between candidates, while we provide the result in a wider framework that encompasses all population cultures. We also prove that most of time, making a voting system meet the Condorcet criterion *strictly* reduces its manipulability.

I.B Contributions and plan

In some studies about voting, each voter's opinions are only represented by a weak or total order of preferences over the candidates. However, in such a framework, it is difficult to study voting systems like Range voting or Approval voting. Some authors circumvent this problem, for example by studying Approval voting with a fixed number of approvals (Chamberlin et al., 1984), but doing so does not take into account the very specificities of such a voting system. In section II, we introduce a general framework where preferences and ballots can embed not only orders of preference but also grades or more general objects.

In section III, we recall the Condorcet criterion and the majority favorite criterion, then we define the *ignorant majority coalition criterion* and the *informed majority coalition criterion*[3]. We prove that each of these criteria implies the next one (the converse being false), and that the last one, informed majority coalition criterion, is met by a large class of classical voting systems from literature.

For any voting system f, we define f^c, its *Condorcification*[4]: if there is a Condorcet winner, f^c designates her; otherwise, f^c has the same output as f. Then we prove that, if f meets the informed majority coalition criterion, its Condorcification f^c is *at most as manipulable* as f.

In section IV, we define the notion of *resistant Condorcet winner*, which is proved to characterize an immunity to manipulation for all voting systems that meet the Condorcet criterion. Exploiting this property, we prove that for all classical voting systems that meet the informed majority coalition criterion but not the Condorcet criterion, their Condorcification is *strictly less manipulable*.

To the best of our knowledge, most previous theoretical results on the manipulability rate are applicable only to some particular values of the number of voters, the number of candidates, or both. On the opposite, the results of this paper allow to decrease the manipulability rate for any value of the parameters and any type of culture.

In appendix A, the reader will find a glossary of the main notations used in this paper. Other appendices contains some developments that can be skipped at first reading: appendix B shows how the framework from section II can be extended further, and appendices C, D, E and F detail some proofs from the main sections.

II Framework

II.A Electoral space

For $n \in \mathbb{N} \setminus \{0\}$, we note $\mathcal{V}_n = \{1, \dots, n\}$ the set of the indexes of *voters*, called simply *voters* in the following. For $m \in \mathbb{N} \setminus \{0\}$, we note $\mathcal{C}_m = \{1, \dots, m\}$

3. These two last notions are based on the same ideas as Peleg's second and third simple games associated to a voting system (Peleg, 1984).

4. Such a process has already been used to define the Black method from the Borda method (Black, 1958).

the set of the indexes of the *candidates*[5].

In order to model preferences, we need to recall a few definitions about binary relations. Most of the ones in this paper represent the strict preferences of a voter and are generally supposed to be antisymmetric, so we will use the "strict" notation $\succ$, even when relaxing the assumption of antisymmetry for the sake of mathematical generality.

Let E be a set, $r \in \mathcal{P}(E^2)$ a binary relation and $(c,d) \in E^2$. We note:

- $c \succ_r d$ if and only if (iff) $(c,d) \in r$;
- $c \gtrsim_r d$ iff $c \succ_r d$ and $d \succ_r c$;
- $c \not\gtrsim_r d$ iff not$(c \succ_r d)$ and not$(d \succ_r c)$;
- $c \ggg_r d$ iff $c \succ_r d$ and not$(d \succ_r c)$.

Let us remark that if relation r is antisymmetric, which will be assumed in most models, then there are only three mutually exclusive possibilities: $c \succ_r d$ (which is equivalent to $c \ggg_r d$ in this case), $d \succ_r c$ and $c \not\gtrsim_r d$.

We say that relation r is:

- *Irreflexive* iff $\forall c \in E, \text{not}(c \succ_r c)$;
- *Antisymmetric* iff $\forall (c,d) \in E^2, c \neq d$ and $c \succ_r d \Rightarrow \text{not}(d \succ_r c)$;
- *Complete*[6] iff $\forall (c,d) \in E^2, c \neq d$ and $\text{not}(d \succ_r c) \Rightarrow c \succ_r d$;
- *Transitive* iff $\forall (c,d,e) \in E^3, c \succ_r d$ and $d \succ_r e \Rightarrow c \succ_r e$;
- *Negatively transitive* iff $\forall (c,d,e) \in E^3, \text{not}(c \succ_r d)$ and $\text{not}(d \succ_r e) \Rightarrow \text{not}(c \succ_r e)$;
- A *strict weak order* iff it is negatively transitive, irreflexive and antisymmetric;
- A *strict total order* iff it is transitive, irreflexive (hence antisymmetric) and complete.

For $m \in \mathbb{N} \setminus \{0\}$, we note $\mathcal{R}_m = \mathcal{P}({\mathcal{C}_m}^2)$ the set of the binary relations over $\mathcal{C}_m$: an element of this set allows to represent a voter's binary relation of preference over the candidates.

Throughout this paper, we will illustrate our results with an example of model that allows to study most classical voting systems. In this specific model, each voter i is capable of mentally establishing:

- A strict weak order of preference r_i over the candidates,
- A vector $u_i \in [0,1]^m$ of grades over the candidates[7],
- A vector $a_i \in \{0,1\}^m$ of approval values over the candidates.

The triple $\omega_i = (r_i, u_i, a_i)$ will be called her *state* and we will note R_i the function that extracts the first element of this triple: $R_i(\omega_i) = R_i(r_i, u_i, a_i) = r_i$.

We could assume that a voter's state has some form of inner coherence: for example, if voter i strictly prefers candidate c to d, i.e. if $c \succ_{r_i} d$, we could demand that her grade (resp. her approval value) for c is no lower than her one for d, i.e. $u_i(c) \geq u_i(d)$ (resp. $a_i(c) \geq a_i(d)$). But in fact, our results hold true with or without such an additional assumption.

5. Candidates can be or not be voters themselves, without impact on our results.
6. We could say *weakly complete* because this definition does not require $c \succ_r c$.
7. We could demand that grades belong to a finite subset of $[0,1]$ for practical applications, but this has no impact on our results.

To be more general, we now define a class of models, called *electoral spaces*, that allow to study any kind of voting system, with virtually any assumptions about the structure of each voter's opinions.

Definition II.1 (ES, electoral space)**.** An *electoral space*, or *ES*, is given by:

- A number of voters $n \in \mathbb{N} \setminus \{0\}$ and a number of candidates $m \in \mathbb{N} \setminus \{0\}$,
- For each voter $i \in \mathcal{V}_n$, a non-empty set Ω_i of her possible states,
- For each voter $i \in \mathcal{V}_n$, a function $R_i : \Omega_i \to \mathcal{R}_m$, which allows to know the binary relation of preference associated to her state.

We note $\Omega = \prod_{i \in \mathcal{V}_n} \Omega_i$: it is the set of possible states for the whole population. We note $R = (R_1, \ldots, R_n)$ the n-uple of all functions R_i.

Such an electoral space is denoted (n, m, Ω, R), or just Ω when there is no ambiguity.

If we want our model to embed the inner coherence assumption of our example, we simply have to define Ω_i as the set of triples (r_i, u_i, a_i) that meet this assumption.

Here is another example: in some studies about voting, a voter's opinion is only represented by a strict total order over the candidates. This model simply corresponds to the choice of the *electoral space of strict total orders (for n and m)*, where each Ω_i is the set $\mathfrak{S}_m$ of strict total orders and each R_i is the identity function.

In definition II.1, binary relations of preference $R_i(\omega_i)$ are not supposed to be antisymmetric: we will discuss this later. If the reader feels uncomfortable with this, she may read the following with an additional assumption of antisymmetry in mind: in this case, $c \ggcurly_{R_i(\omega_i)} d$ means the same as $c \succ_{R_i(\omega_i)} d$.

In most reasonable models, it is assumed that any voter might prefer any candidate to all the others and, more generally, that any strict total order is a possible relation of preference for any voter. Since these properties will be useful for some of our results, we formally define them now.

Definition II.2 (richness of an electoral space)**.** Let Ω be an ES. We say that:

1. Ω *comprises all strict total orders* iff for any voter, the set of her possible relations of preference comprises all strict total orders; that is, $\forall i \in \mathcal{V}_n, \mathfrak{S}_m \subset R_i(\Omega_i)$;
2. Ω *allows any candidate as most liked* iff any voter may strictly prefer any candidate to all the others; that is, $\forall (i, c) \in \mathcal{V}_n \times \mathcal{C}_m, \exists r \in R_i(\Omega_i)$ s.t. $\forall d \in \mathcal{C}_m \setminus \{c\}, c \ggcurly_r d$.

We trivially have the implication 1 $\Rightarrow$ 2.

In our reference model, where $\omega_i = (r_i, u_i, a_i)$, relation r_i may be any strict weak order, so it is allowed to be any strict total order. Hence, this electoral space meets the above properties 1 and 2.

II.B State-based voting systems and manipulability

Now, we present *state-based voting systems*, a framework that allows to study most voting systems. In appendix B, we define an even more general model and we prove that, for the purpose of diminishing manipulability, we can always restrict our study to state-based voting systems[8].

Definition II.3 (state-based voting system, SBVS)**.** Let Ω be an ES.

A *state-based voting system* over Ω, or *SBVS*, is a function:

$$f : \left| \begin{array}{ccc} \Omega & \to & \mathcal{C}_m \\ (\omega_1, \ldots, \omega_n) & \to & f(\omega_1, \ldots, \omega_n). \end{array} \right.$$

For example, let us consider one of the possible variants for the voting system called *Range voting*, in our reference electoral space where $\omega_i = (r_i, u_i, a_i)$.

- Each voter i communicates a state belonging to Ω_i;
- We say that she votes *sincerely* iff she communicates her true state ω_i;
- Function f takes into account only the vectors of grades communicated by the voters, then returns the candidate with highest total grade (and resolves ties in an arbitrary deterministic way).

Definition II.4 (manipulability of an SBVS)**.** As is usually done in the literature, we say that a voting system is *manipulable* in a given situation iff there exists a coalition of voters who, by misrepresenting their states, may secure an outcome that is different from the sincere winner and that they all prefer to her, while assuming that other voters still vote sincerely.

Formally, let Ω be an ES and f an SBVS.

For $(\omega, \psi) \in \Omega^2$, we say that f is *manipulable in situation ω towards ψ* iff:

$$\begin{cases} f(\psi) \neq f(\omega), \\ \forall i \in \mathcal{V}_n, \psi_i \neq \omega_i \Rightarrow f(\psi) \succ_{R_i(\omega_i)} f(\omega). \end{cases}$$

For $\omega \in \Omega$, we say that f is *manipulable in situation ω* iff $\exists \psi \in \Omega$ s.t. f is manipulable in ω towards ψ.

In most reasonable models we could think of, relations $R_i(\omega_i)$ are supposed to be antisymmetric. If we remove this assumption and if we have $c \precsim\succsim_{R_i(\omega_i)} d$, how can it be interpreted? We could read that voter i "strictly prefers" candidate c to d and d to c in the same time, but it is hard to conceive.

In fact, relation $R_i(\omega_i)$ has only been used in definition II.4 so far. Hence, assertion $c \precsim\succsim_{R_i(\omega_i)} d$ should be interpreted in terms of manipulation: it means that if c is the sincere winner, then voter i is susceptible to manipulate for d, and vice versa. For example, she may be indifferent between c and d; but she is susceptible to accept a bribery to manipulate for d against c, and vice versa. With this interpretation, if $c \not\precsim\succsim_{R_i(\omega_i)} d$, it means that voter i is also indifferent between c and d, but incorruptible.

8. This notion of *state-based voting system* is a generalization of what is called *elementary voting procedure* ("procédure de vote élémentaire") by Moulin (1978, chapter II, definition 2). The author already remarks that considering such a procedure only reduces the strategical possibilities of the voters.

In a model that embeds such possibilities, the *manipulability* of situation ω (as formally defined above) should be interpreted as a vulnerability to the combined effect of the *manipulation* by voters who really prefer the new winner to the sincere one and the *bribery* of some corruptible voters. Again, we insist that for readers who might be disturbed by such a model, all the following can be read with an assumption of antisymmetry and the usual interpretation of manipulability.

When we consider two SBVS f and g, we say that f is *at most as manipulable as* g (in the sense of inclusion[9]) iff:

$$\{\omega \in \Omega \text{ s.t. } f \text{ is manipulable in } \omega\} \subseteq \{\omega \in \Omega \text{ s.t. } g \text{ is manipulable in } \omega\}.$$

If the electoral space Ω is endowed with a probability measure $\mathbb{P}$, we will call *manipulability rate of* f *for* $\mathbb{P}$:

$$\rho_{\mathbb{P}}(f) = \mathbb{P}(f \text{ is manipulable in } \omega).$$

For f, being at most as manipulable as g (in the sense of inclusion) is a very interesting property because it implies[10] that for any probability measure $\mathbb{P}$ over Ω, we have $\rho_{\mathbb{P}}(f) \leq \rho_{\mathbb{P}}(g)$: the manipulability rate of voting system f is lower or equal to the one of g.

III Condorcification and manipulability

In this section, we define the *Condorcification* of a voting system and we prove that, for most voting systems, the Condorcification is at most as manipulable as the original system.

III.A Some criteria for a voting system

First, we recall the Condorcet criterion and we present three other criteria, which are successive relaxations of this criterion.

When $A(i)$ is an assertion that depends on voter i, we note $|A(i)| = \text{card}\{i \in \mathcal{V}_n \text{ s.t. } A(i)\}$. For example, $D_{cd}(\omega) = |c \succ_{R_i(\omega_i)} d| = \text{card}\{i \in \mathcal{V}_n \text{ s.t. } c \succ_{R_i(\omega_i)} d\}$ denotes the number of voters who prefer c to d in situation ω. The matrix $D(\omega)$ is called the *matrix of duels of* ω.

We call *victory relation in* ω, and we note $V(\omega) \in \mathcal{P}({\mathcal{C}_m}^2)$, the binary relation defined by: $c \succ_{V(\omega)} d$ iff $D_{cd}(\omega) > \frac{n}{2}$. When this relation is met, we indifferently say that c has a *victory* against d in ω, or that d has a *defeat* against c in ω.

Usually in literature, a victory for c against d is defined as a situation where more voters prefer c to d than the opposite: it is a notion of *relative* victory. But in this paper, we will always be interested in *absolute* victories, i.e. when

9. This notion is defined by Lepelley and Mbih (1994).

10. If Ω is endowed with a sigma-algebra such that any singleton is measurable, it is not only an implication but an equivalence.

more than half of all the voters prefer c to d. For conciseness, we just call these situations *victories*.

For example, with 100 voters, if one voter strictly prefers c to d and all the others have no preference between c and d, then we do not say that c has a *victory* against d. For that, we request that 51 voters or more prefer c to d.

In models where relations $R_i(\omega_i)$ may not be antisymmetric, let us note that c and d may simultaneously have victories against each other. For this reason, we say that c has a *strict victory* (resp. *strict defeat*) against d in ω iff she has a victory and no defeat (resp. a defeat and no victory) against d, i.e. iff $c \ggcurly_{V(\omega)} d$ (resp. $d \ggcurly_{V(\omega)} c$).

Proposition III.1 (total number of points in a duel)**.** *Let Ω be an ES, $\omega \in \Omega$ and $(c,d) \in {\mathcal{C}_m}^2$. We have:*

$$D_{cd}(\omega) + D_{dc}(\omega) = n + |c \gtrsim_{R_i(\omega_i)} d| - |c \not\gtrsim_{R_i(\omega_i)} d|.$$

Proof. Immediate consequence of partitioning the voters in four classes, according to the value of their binary relation of preference on (c,d) and (d,c). $\square$

From this, we deduce that if all preference relations $R_i(\omega_i)$ are antisymmetric, which is the case in most models, then relation $V(\omega)$ is antisymmetric: two distinct candidates c and d cannot have mutual victories (in other words, any victory is strict).

On the other hand, if all relations $R_i(\omega_i)$ are complete and if the number of voters n is odd, then relation $V(\omega)$ is complete: between two distinct candidates c and d, there cannot be an absence of victory.

Definition III.2 (Condorcet winner, Condorcet-admissible candidate, majority favorite)**.** Let Ω be an ES, $\omega \in \Omega$ and $z \in \mathcal{C}_m$.

We say that z is *Condorcet winner*[11] in ω iff z has a strict victory against any other candidate c; that is:

$$\forall c \in \mathcal{C}_m \setminus \{z\}, z \ggcurly_{V(\omega)} c, \text{i.e.} \begin{cases} |z \succ_{R_i(\omega_i)} c| > \dfrac{n}{2}, & (1) \\ |c \succ_{R_i(\omega_i)} z| \leq \dfrac{n}{2}. & (2) \end{cases}$$

We say that z is *Condorcet-admissible*[12] in ω iff z has no defeat; that is:

$$\forall c \in \mathcal{C}_m \setminus \{z\}, \text{not}(c \succ_{V(\omega)} z), \text{i.e. } |c \succ_{R_i(\omega_i)} z| \leq \frac{n}{2}.$$

We say that z is a *majority favorite* in ω iff a strict majority of voters strictly prefers z to any other candidate: $|\forall c \in \mathcal{C}_m \setminus \{z\}, z \ggcurly_{R_i(\omega_i)} c| > \frac{n}{2}$.

11. Since we consider absolute victories and not relative victories, our definition coincides with the one commonly found in the literature when preferences are strict total orders, but may differ otherwise.

12. When relations of preference are strict total orders, this notion coincides with the *weak Condorcet winner*, as defined for example by Taylor (2005).

If relations $R_i(\omega_i)$ are antisymmetric, which is the case in most models, then we have remarked that any victory is strict. So z is Condorcet winner in ω iff:

$$\forall c \in \mathcal{C}_m \setminus \{z\}, |z \succ_{R_i(\omega_i)} c| > \frac{n}{2}.$$

If z is a majority favorite in ω, then obviously z is a Condorcet winner in ω.

We also have the following trivial but important property.

Proposition III.3 (unicity of the Condorcet winner)**.** *If a candidate is Condorcet winner in ω, then she is Condorcet-admissible and no other candidate is Condorcet-admissible; in particular, she is the unique Condorcet winner.*

If all binary relations $R_i(\omega_i)$ are complete and if the number of voters n is odd, then we have remarked that relation $V(\omega)$ is complete. In that case, a Condorcet-admissible is always a Condorcet winner and these two notions become equivalent.

Definition III.4 (some criteria for an SBVS)**.** Let Ω be an ES and f an SBVS.

We say that f *meets the Condorcet criterion* iff, for any $\omega \in \Omega$ and $z \in \mathcal{C}_m$, if z is a Condorcet winner in ω, then $f(\omega) = z$.

We say that f *meets the majority favorite criterion* iff, for any $\omega \in \Omega$ and $z \in \mathcal{C}_m$, if z is a majority favorite in ω, then $f(\omega) = z$.

We say that f *meets the ignorant majority coalition criterion* iff any majority coalition may decide of the outcome, whatever the other voters do. That is, $\forall \mathcal{M} \in \mathcal{P}(\mathcal{V}_n)$, if $\mathrm{card}(\mathcal{M}) > \frac{n}{2}$ then $\forall c \in \mathcal{C}_m, \exists \omega \in \Omega$ s.t. $f(\omega) = c$ and $[\forall \psi \in \Omega, (\forall i \in \mathcal{M}, \psi_i = \omega_i) \Rightarrow f(\psi) = c]$.

We say that f *meets the informed majority coalition criterion* iff any majority coalition, that is informed of what the other voters do, may decide of the outcome. That is, $\forall \omega \in \Omega, \forall \mathcal{M} \in \mathcal{P}(\mathcal{V}_n)$, if $\mathrm{card}(\mathcal{M}) > \frac{n}{2}$ then $\forall c \in \mathcal{C}_m$:

$$\exists \psi \in \Omega \text{ s.t. } \begin{cases} f(\psi) = c, \\ \forall i \in \mathcal{V}_n \setminus \mathcal{M}, \psi_i = \omega_i. \end{cases}$$

Proposition III.5 (implications between the criteria)**.** *Let Ω be an ES and f an SBVS. We consider the following conditions.*

1. *f meets the Condorcet criterion.*
2. *f meets the majority favorite criterion.*
3. *f meets the ignorant majority coalition criterion.*
4. *f meets the informed majority coalition criterion.*

We have the following implications: 1 $\Rightarrow$ 2 and 3 $\Rightarrow$ 4.

If the electoral space allows any candidate as most liked (which is a common assumption, cf. definition II.2), then we also have 2 $\Rightarrow$ 3.

Each converse is false.

Proof. 1 $\Rightarrow$ 2: If z is a majority favorite, then she is Condorcet winner so she gets elected.

2 ⇒ 3: If the members of an ignorant majority coalition wish to get c elected, they can just pretend that they strictly prefer c to any other candidate, which is possible since the electoral space allows any candidate as most liked.

3 ⇒ 4: It is obvious.

The following proposition proves that each converse is false. □

Proposition III.6 (criteria met by some classical voting systems)**.**

1. *The following voting systems meet the Condorcet criterion: maximin*[13]*, Tideman method (also called Ranked pairs, Tideman, 1987), Schulze method (Schulze, 2011), Kemeny-Young method (Kemeny, 1959; Young and Levenglick, 1978), Dodgson method, Baldwin method, Nanson method.*

We now consider electoral spaces that comprise all strict total orders.

2. *The following voting systems meet the majority favorite criterion but not the Condorcet criterion: plurality, two-round system, instant-runoff voting (IRV), Bucklin method*[14]*.*
3. *The following voting systems meet the ignorant majority coalition criterion but not the majority favorite criterion: approval voting (Brams and Fishburn, 1978), range voting with average, range voting with median*[15]*, Coombs method.*
4. *The Borda method meets the informed majority coalition criterion but not the ignorant majority coalition criterion.*
5. *Veto (also called antiplurality) does not meet the informed majority coalition criterion.*

Proof. 1 and 2 are classical and easy results.

3. For approval voting, range voting with average and range voting with median, it is sufficient that the members of the coalitions assign the maximal grade to c and the minimal grade to all other candidates.

For the Coombs method, it is sufficient that the coalition members give an order with c on top and all other candidates in the same arbitrary order: this ensures that these candidates will be eliminated one after another, from the last to the second.

However, it is classical and easy to prove that these voting systems do not meet the majority favorite criterion.

4. See appendix C.

5. Let us consider $m = 4$ candidates and $n = 3$ voters, among which there are 2 manipulators. For a candidate to be elected, she must receive no veto at

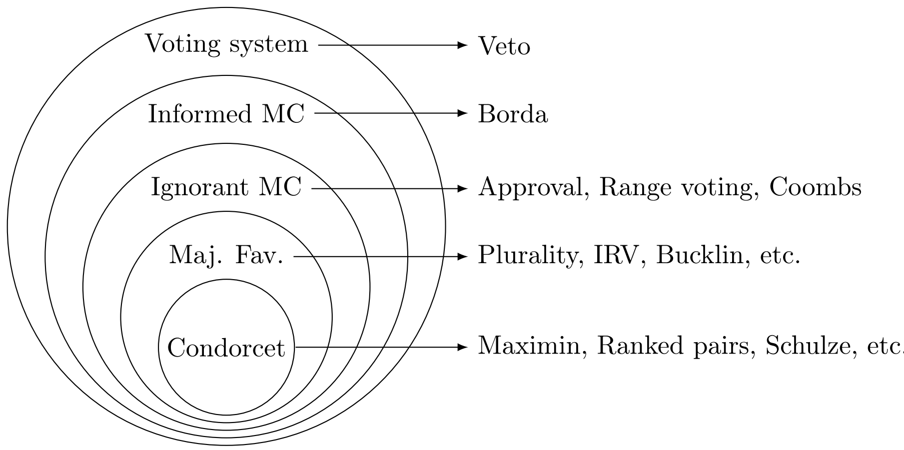
13. Unless specific reference is mentioned, all voting systems used in this paper are described in details by Tideman (2006). Whatever tie-breaking rule is considered has no impact for this proposition.

14. Among different variants of the *Bucklin method*, here is the one we consider (on the electoral space of strict total orders). At counting step k, the score of each candidate is the number of voters who placed her between the first and the k-th position. If at least one candidate gets at least $\frac{n}{2}$ points, the candidate with the greatest score gets elected. Otherwise, the counting step $k + 1$ is done.

15. *Range voting with median* is defined like *range voting with average*, but using the median range for each candidate. This voting system is similar to Majority Judgment (Balinski and Laraki, 2010).

all. If the sincere voter votes against candidate c, then whatever the members of the informed majority coalition do, they cannot secure a victory for c. $\square$

In most models, the electoral space allows any candidate as most liked, so we can sum up the previous results in the following inclusion diagram. It is emphasized that all the voting systems we mentioned except veto meet the informed majority coalition criterion.



III.B Condorcification theorem

Now, we define the Condorcification of an SBVS. Then we prove that, if an SBVS meets the informed majority coalition criterion, then its Condorcification is at most as manipulable as the original system.

Definition III.7 (Condorcification)**.** Let Ω be an ES and f an SBVS.

We call *Condorcification of f* the state-based voting system:

$$f^c : \left| \begin{array}{lcl} \Omega & \to & \mathcal{C}_m \\ \omega & \to & \left| \begin{array}{l} \text{if } \omega \text{ has a Condorcet winner } z, \text{then } z, \\ \text{otherwise, } f(\omega). \end{array} \right. \end{array} \right.$$

For example, let us consider the Condorcification of range voting, in our reference electoral space where $\omega_i = (r_i, u_i, a_i)$.

- Each voter gives a full state.
- If there appears to be a Condorcet winner (computed with relations r_i that are communicated), then she is elected.
- Otherwise, the candidate with highest total grade (computed with vectors of grades u_i that are communicated) is elected.

By design, the Condorcification f^c meets the Condorcet criterion.

Lemma III.8 (no manipulation from Condorcet-admissible towards Condorcet)**.** *Let Ω be an ES, f an SBVS and $(\omega, \psi) \in \Omega^2$. We note $z = f(\omega)$ and $c = f(\psi)$.*

If z is Condorcet-admissible in ω and if f is manipulable in ω towards ψ, then c cannot have a victory against z in ψ; in particular, c is not a Condorcet winner in ψ.

Proof. Since z is Condorcet-admissible in ω, we have: $|c \succ_{R_i(\omega_i)} z| \leq \frac{n}{2}$. Since only voters who prefer c to z may have changed their ballots in ψ, we also have $|c \succ_{R_i(\psi_i)} z| \leq \frac{n}{2}$. It should be noticed that, for this lemma, it is not necessary to assume that f meets the informed majority coalition criterion. □

This lemma implies a result already shown by Moulin (1978, chapter I, theorem 1): if the voting system meets the Condorcet criterion, then it cannot be manipulable in a situation with a Condorcet winner towards another situation with a Condorcet winner.

Lemma III.9 (non Condorcet-admissible implies manipulable)**.** *Let Ω be an ES, f an SBVS and $\omega \in \Omega$. We assume that f meets the informed majority coalition criterion.*

If $f(\omega)$ is not Condorcet-admissible in ω, then f is manipulable in ω.

Proof. Let us note $z = f(\omega)$. Since z is not Condorcet-admissible, there is another candidate c who has a victory against her: the subset $\mathcal{M}$ of voters who prefer c to z is a strict majority.

Using informed majority coalition criterion, $\exists \psi \in \Omega$ s.t. $f(\psi) = c$ and $\forall i \in \mathcal{V}_n \setminus \mathcal{M}, \psi_i = \omega_i$.

Hence, f is manipulable in ω towards ψ, in favor of candidate c. □

As an immediate consequence, if ω has no Condorcet-admissible candidate, then any f meeting the informed majority coalition criterion is manipulable in ω.

Theorem III.10 (Condorcification)**.** *Let Ω be an ES and f an SBVS. We assume that f meets the informed majority coalition criterion.*

Then its Condorcification f^c is at most as manipulable as f:

$$\{\omega \in \Omega \text{ s.t. } f^c \text{ is manipulable in } \omega\} \subseteq \{\omega \in \Omega \text{ s.t. } f \text{ is manipulable in } \omega\}.$$

Proof. Let us suppose that f^c is manipulable in ω towards ψ, but f is not manipulable in ω.

Since f is not manipulable in ω, lemma III.9 ensures that $f(\omega)$ is Condorcet-admissible in ω. If she is a Condorcet winner in ω, then $f^c(\omega) = f(\omega)$; otherwise, there is no Condorcet winner in ω (proposition III.3) so, by definition of f^c, we have also $f^c(\omega) = f(\omega)$.

So, let us note $z = f^c(\omega) = f(\omega)$ and $c = f^c(\psi)$. Since z is Condorcet-admissible in ω, lemma III.8 (applied to f^c) ensures that c is not a Condorcet winner in ψ. So, by definition of f^c, we have $f^c(\psi) = f(\psi)$.

Hence, we have $f(\omega) = f^c(\omega)$ and $f(\psi) = f^c(\psi)$ so f is manipulable in ω: contradiction! □

For example, we have just proved that if f is Range voting, then its Condorcification f^c is at most as manipulable as f (in the sense of inclusion). In particular, for any probability measure $\mathbb{P}$ over Ω, we have $\rho_{\mathbb{P}}(f^c) \leq \rho_{\mathbb{P}}(f)$.

In appendix D, we show that this theorem does not generalize when using Condorcet-admissible candidates.

IV Resistant Condorcet winner

In this section, we define the notion of *resistant Condorcet winner*, which is characterized by an immunity to manipulation. Then, using this property, we give a sufficient condition for the Condorcification to be *strictly* less manipulable than the original system.

If the voting system meets the Condorcet criterion and if there is a Condorcet winner z, manipulators in favor of c need to prevent z from appearing as a Condorcet winner. So, they need to prevent a strict victory for z against some candidate $d \neq z$. However, this plan is doomed to failure if sincere voters (those who do not prefer c to z) ensure on their own: (1) a victory for z against d; and (2) a non-victory for d against z. This observation leads to the following definition.

Definition IV.1 (resistant Condorcet winner)**.** Let Ω be an ES, $\omega \in \Omega$ and $z \in \mathcal{C}_m$.

We say that z is *resistant Condorcet winner* in ω iff $\forall (c, d) \in (\mathcal{C}_m \setminus \{z\})^2$:

$$\begin{cases} |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \succ_{R_i(\omega_i)} d| > \dfrac{n}{2}, & (3) \\ |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and not}(d \succ_{R_i(\omega_i)} z)| \geq \dfrac{n}{2}, & (4) \end{cases}$$

If all binary relations $R_i(\omega_i)$ are antisymmetric, which is the case in most models, then z is resistant Condorcet winner iff $\forall (c, d) \in (\mathcal{C}_m \setminus \{z\})^2$:

$$|\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \succ_{R_i(\omega_i)} d| > \frac{n}{2}.$$

If, moreover, all binary relations $R_i(\omega_i)$ are complete, then the definition is even simpler, since it becomes symmetrical between c and d. Indeed, with these assumptions, candidate z is a resistant Condorcet winner if and only $\forall (c, d) \in (\mathcal{C}_m \setminus \{z\})^2$:

$$|z \succ_{R_i(\omega_i)} c \text{ and } z \succ_{R_i(\omega_i)} d| > \frac{n}{2}.$$

It is clear that for a candidate z, being a majority favorite implies being a resistant Condorcet winner, which implies being a Condorcet winner. As a consequence of proposition III.3, if there is a resistant Condorcet winner, then she is unique.

Proposition IV.2 (characterization of the resistant Condorcet winner)**.** *Let Ω be an ES, $\omega \in \Omega$ and $z \in \mathcal{C}_m$.*

Let us consider the following conditions.

1. *z is a resistant Condorcet winner in ω.*
2. *z is a Condorcet winner in ω and, for any state-based voting system f meeting the Condorcet criterion, f is not manipulable in ω.*

We have the implication 1 $\Rightarrow$ 2.

If the electoral space comprises all strict total orders (which is a common assumption, cf. definition II.2), then the converse 2 $\Rightarrow$ 1 is true.

Proof. 1 $\Rightarrow$ 2: since z is a resistant Condorcet winner, even after an attempt of manipulation for c, sincere voters guarantee that z still has a strict victory against any other candidate $d \neq z$; so, candidate z still appears as a Condorcet winner and gets elected.

For the converse 2 $\Rightarrow$ 1, see appendix E. □

Definition IV.3 (resistant-Condorcet criterion)**.** Let Ω be an ES and f an SBVS.

We say that f *meets the resistant-Condorcet criterion* iff, for any $\omega \in \Omega$ and $z \in \mathcal{C}_m$, if z is a resistant Condorcet winner in ω, then $f(\omega) = z$.

It is clear that meeting the Condorcet criterion implies meeting the resistant-Condorcet criterion, which implies meeting the majority favorite criterion.

To the best of our knowledge, there is no classical voting system from literature that meets the resistant-Condorcet criterion but not the Condorcet criterion.

Theorem IV.4 (strict improvement by Condorcification)**.** *Let Ω be an ES and f an SBVS. We assume that f meets the informed majority coalition criterion but not the resistant-Condorcet criterion.*

Then its Condorcification f^c is strictly less manipulable than f:

$$\{\omega \in \Omega \text{ s.t. } f^c \text{ is manipulable in } \omega\} \subsetneq \{\omega \in \Omega \text{ s.t. } f \text{ is manipulable in } \omega\}.$$

Proof. Condorcification theorem III.10 ensures the inclusion.

Since f does not meet the resistant-Condorcet criterion, there exists $\omega \in \Omega$ and $z \in \mathcal{C}_m$, resistant Condorcet winner in ω, such that $f(\omega) \neq z$. Thanks to proposition III.3, we know that $f(\omega)$ is not Condorcet-admissible in ω, so lemma III.9 ensures that f is manipulable in ω. But proposition IV.2 ensures that f^c is not manipulable in ω. Thus, the inclusion is strict. □

Corollary IV.5 (Condorcification of classical voting systems)**.** *We consider electoral spaces that comprise all strict total orders. For each voting system f in the following list, there are values of n and m such that the Condorcification f^c is strictly less manipulable than f: approval voting, Borda method, Bucklin method, Coombs method, range voting with median, range voting with average, instant-runoff voting (IRV), plurality, two-round system.*

Proof. It is sufficient to prove that none of these voting systems meet the resistant-Condorcet criterion. See appendix F for counterexamples. □

Conclusion

We have proved that for a large class of voting systems (those meeting the informed majority coalition criterion), the Condorcification is at most as manipulable as the original system, and often strictly less. It does not mean that Condorcet systems are "generally better": a given non-Condorcet system might be less manipulable than a given Condorcet system.

However, this shows that Condorcet criterion is a crucial asset in the search for voting systems whose manipulability is minimal. More precisely, let us consider the function $M : f \to \mathcal{P}(\Omega)$ that, to each voting system f, associates the set of situations where f is manipulable. If we want to find the minima of M (for the inclusion) among the voting systems meeting the informed majority coalition criterion, then we must restrict our search to those meeting the resistant-Condorcet criterion (theorem IV.4), and we can restrict our search to those meeting the Condorcet criterion: indeed, it follows from theorem III.10 that any minimal value of M is reached by a Condorcet voting system.

François Durand: Inria, Paris, France.
Fabien Mathieu and Ludovic Noirie: Alcatel-Lucent Bell Labs France, Nozay, France.

A Main notations

Binary relations	
$c \succ_r d$	Pair (c, d) belongs to binary relation r.
$c \ggcurly_r d$	Pair (c, d) belongs to binary relation r, but not pair (d, c).
$c \succsim_r d$	Both pairs (c, d) and (d, c) belong to binary relation r.
$c \not\succsim_r d$	Elements c and d are not comparable in the sense of binary relation r.
Electoral space	
$n \in \mathbb{N} \setminus \{0\}$	Number of voters.
$\mathcal{V}_n$	Set $\{1, \ldots, n\}$ of the indexes of the voters.
$m \in \mathbb{N} \setminus \{0\}$	Number of candidates.
$\mathcal{C}_m$	Set $\{1, \ldots, m\}$ of the indexes of the candidates.
$\mathcal{R}_m$	Set of binary relations over $\mathcal{C}_m$.
Ω_i	Set of possible states for voter i.
$\omega_i \in \Omega_i$	State of voter i.
R_i	Function $\Omega_i \to \mathcal{R}_m$ that extracts from voter i's state ω_i her binary relation of preference r_i.
Ω	Set $\prod_{i \in \mathcal{V}_n} \Omega_i$ of possible states for the whole population of voters. Also used as a shortcut of notation for electoral space (n, m, Ω, R).
$D(\omega)$	Matrix of duels for ω.
$V(\omega)$	Binary relation of victory in ω.
Voting sytem	
f	A state-based voting system (SBVS), i.e. a function $\Omega \to \mathcal{C}_m$.
f^c	Condorcification of state-based voting system f.

B General voting systems

Here, we present a general framework that allows to represent any kind of voting system. Then we prove that, for the purpose of diminishing manipulability, investigation can be restricted to *state-based voting systems*, which justifies that we did so throughout the paper.

Let us consider one of the possible variants for the voting protocol called *Range voting*. Each voter i has at her disposal a set of strategies $\mathcal{S}_i = [0, 1]^m$: she has to attribute a grade to each candidate. Once these grades are communicated, we use a processing rule f that returns the candidate with highest total grade.

More generally, a *game form*[16] (for n and m) is given by:

- For each voter $i \in \mathcal{V}_n$, a set $\mathcal{S}_i$ whose elements are called *strategies*[17];
- A function $f : \mathcal{S}_1 \times \ldots \times \mathcal{S}_n \to \mathcal{C}_m$ that is called *processing rule*.

Now, the question is: when voter i is in state $\omega_i \in \Omega_i$, which strategy $S_i \in \mathcal{S}_i$ should be called a sincere ballot? As noticed by Gibbard (1973), there is no obvious general way to define sincere voting on the only basis of the game form itself[18]. So, to finish defining a voting system, we need to add functions $s_i : \Omega_i \to \mathcal{S}_i$ that reflect our idea of sincere voting.

For example, in our study of Range voting, we might estimate that when voter i is in state (r_i, u_i, a_i), her sincere ballot is u_i. In other words, we transmit this message to the voters: "Your instruction is to communicate your vector u_i"[19].

More generally, a *general voting system* (over an ES Ω) is given by:

- A game form $((\mathcal{S}_i)_{i \in \mathcal{V}_n}, f)$;
- For each voter $i \in \mathcal{V}_n$, a function $s_i : \omega_i \to \mathcal{S}_i$ that is called *sincerity function*.

Let Ω be an ES, $F = ((\mathcal{S}_i)_{i \in \mathcal{V}_n}, f, (s_i)_{i \in \mathcal{V}_n})$ a general voting system and $\omega \in \Omega$.

For $S = (S_1, \ldots, S_n) \in \mathcal{S}_1 \times \ldots \times \mathcal{S}_n$, we say that F is *manipulable in ω towards S* iff:

$$\begin{cases} f(S_1, \ldots, S_n) \neq f(s_1(\omega_1), \ldots, s_n(\omega_n)), \\ \forall i \in \mathcal{V}_n, S_i \neq s_i(\omega_i) \Rightarrow f(S_1, \ldots, S_n) \succ_{R_i(\omega_i)} f(s_1(\omega_1), \ldots, s_n(\omega_n)). \end{cases}$$

We say that F *manipulable in ω* iff there exists $S \in \mathcal{S}_1 \times \ldots \times \mathcal{S}_n$ such that F is manipulable in ω towards S.

Now, we shall prove that, for the purpose of diminishing manipulability, we can restrict our study to *state-based voting systems*. For example, let us consider the electoral space of strict total orders: the state ω_i of each voter i is a strict total order of preference over the candidates. In this model, let us examine the *two-round system*:

- During the first round, each voter i votes for a candidate; the two candidates with most votes are selected for the second round. Then each voter votes for one of them two and the one with most votes is elected. Let us note that any element of $\mathcal{S}_i$ is a decision tree: i chooses one candidate for the first round and, for all possible results of the first round, one of the two selected candidates for the second round.

16. The terminology *game form* is taken from Gibbard (1973).

17. Here, strategies are what is called "pure strategies" in game theory. They can simply be ballots, decision trees in a multi-stage process or, more generally, any kind of objects.

18. Gibbard (1973) provides an illuminating example, the *non-alcoholic party*.

19. Other choices are possible; for example, we could give the instruction that each voter should scale her vector of grades so that her minimum is 0 and her maximum is 1. This is just another model, that leads to study another *voting system* based on the same *game form* (Gibbard, 1973).

- We choose the most intuitive sincerity function: vote for the candidate one prefers among the available candidates.

In this system, let us imagine a voter who votes for candidate A during the first round and then, in a second round between A and B, who votes for B. Obviously, she can not be sincere.

To prevent this kind of behavior, we can modify the voting system: voters communicate their orders of preference; then, two-round system is emulated with the corresponding sincere strategies.

We are going to generalize this idea.

Proposition B.1 (state-based version)**.** *Let Ω be an ES and $F = ((\mathcal{S}_i)_{i \in \mathcal{V}_n}, f, (s_i)_{i \in \mathcal{V}_n})$ a general voting system. Let us consider the voting system $F' = ((\mathcal{S}'_i)_{i \in \mathcal{V}_n}, f', (s'_i)_{i \in \mathcal{V}_n})$, defined as follows.*

1. *Each voter communicates a state: $\forall i \in \mathcal{V}_n, \mathcal{S}'_i = \Omega_i$.*
2. *Sincerity consists of giving one's true state: $\forall i \in \mathcal{V}_n, s'_i = Id$.*
3. *To get the result, the former rule f is used, considering that each voter uses the sincere strategy corresponding to the state she communicated: $f'(\omega) = f(s_1(\omega_1), \ldots, s_n(\omega_n))$.*

Then voting system F', called the state-based version *of F, is at most as manipulable as F:*

$$\{\omega \in \Omega \text{ s.t. } F' \text{ is manipulable in } \omega\} \subseteq \{\omega \in \Omega \text{ s.t. } F \text{ is manipulable in } \omega\}.$$

Proof. In F', sincere voting leads to the same result as in F, but manipulators have access to at most the same alternative ballots. □

A voting system F is called a *state-based voting system* iff it is equal to its state-based version; that is, iff for any voter $i \in \mathcal{V}_n$, $\mathcal{S}_i = \Omega_i$ and $s_i = \text{Id}$. As a shortcut of notation, an SBVS is denoted as its processing rule f.

Considering the state-based version of a voting system has two advantages:

- It may prevent from using strategies that are obviously not sincere, like in the example of the two-round system.
- In an SBVS, the ballot embeds the binary relation of preferences, even if it was not the case in the original voting system. For voting systems like Range voting, this step is necessary prior to defining Condorcification.

C Properties of the Borda method

Here, we prove that the Borda method meets the informed majority coalition criterion, but not the ignorant majority coalition criterion, as claimed in proposition III.6.

Since preferences may not be strict orders, let us precise how we define the Borda score of a candidate: she gets one point whenever a voter strictly prefers her to another candidate, and half a point whenever a voter judges her incomparable or mutually preferable to another candidate. The Borda method designates the candidate with highest Borda score.

Let us prove that the Borda method meets the informed majority criterion. If there are α sincere voters and β manipulators, with $\beta > \alpha$, it is sufficient that the first α manipulators put themselves in bijection with the α sincere voters, each voting the inverse binary relation of her counterpart: the manipulator pretends to prefer j to k iff her counterpart prefers k to j. This disposition allows to equal the scores of all candidates. Then, the remaining $\beta - \alpha$ manipulators simply say that they strictly prefer c to all other candidates, which ensures that c gets elected.

To prove that the Borda method does not meet the ignorant majority coalition criterion, let us consider $m = 5$ candidates and $n = 5$ voters, among which there are 3 manipulators in favor of candidate A. Clearly, they should pretend that they strictly prefer A to all other candidates, which gives $3 \times 4 = 12$ points to A. They also allocate $3 \times 6 = 18$ points to other candidates, so an average of 4.5 points: hence, one of them, say B, has at least 4.5 points.

If the two remaining voters give a strict total order with B on top and A last, then B has at least $4.5 + 8 = 12.5$ points and A still has 12 points, so A cannot be elected.

It may be noticed that the same proofs hold if it is required that voters have strict weak orders (or strict total orders) of preference.

D Condorcification theorem does not generalize to Condorcet-admissible candidates

Following the ideas of Condorcification theorem III.10, we could think of a more advanced form of Condorcification: if there is at least one Condorcet-admissible candidate, choose one arbitrarily; otherwise, use the original rule. We could hope that this new voting system would be at most as manipulable as the original one, as in Condorcification theorem III.10; but it is not true.

Indeed, let us consider $n = 6$ voters, $m = 4$ candidates called A,B,C,V and the electoral space of strict total orders. We define the voting system *Condorcet-dean*: if a candidate is a Condorcet winner, she is elected; otherwise, the dean V is elected. This system meets the informed majority coalition criterion, since it meets the Condorcet criterion.

Let us consider situation ω as follows (where orders of preferences are represented with the most liked candidate on top) and its matrix of duels.

Preferences:

ω	Voter					
	1	2	3	4	5	6
Preferences	A	B	V	A	C	C
	B	V	A	C	V	V
	V	A	B	V	A	A
	C	C	C	B	B	B

Matrix of duels:

ω	V	A	B	C
V	–	4	4	3
A	2	–	5	4
B	2	1	–	3
C	3	2	3	–

So, V is the only Condorcet-admissible candidate. It is easy to see that Condorcet-dean is not manipulable in situation ω: if it were, that would be towards a situation without a Condorcet winner (lemma III.8). But all such situations elect V!

Now, let us consider situation ψ as follows, where only voters 4 and 5 have been modified, trying to make C win instead of V.

Preferences:

ψ	Voter					
	1	2	3	4	5	6
Preferences	A	B	V	C	C	C
	B	V	A	A	B	V
	V	A	B	B	V	A
	C	C	C	V	A	B

Matrix of duels:

ψ	V	A	B	C
V	–	4	2	3
A	2	–	4	3
B	4	2	–	3
C	3	3	3	–

So, C is the only Condorcet-admissible candidate.

If the voting system is modified so as to elect a Condorcet-admissible candidate when possible, then V stays elected in ω but C gets designated in ψ. Hence, this modified voting system is manipulable in ω towards ψ, while the original system is not manipulable in ω.

This example shows that generalizing Condorcification to Condorcet-admissible candidates can worsen manipulability, even for situations with only one Condorcet-admissible candidate. That is why the Condorcification we defined takes only (strict) Condorcet winners into account.

E Characterization of the resistant Condorcet winner

In this appendix, we finish the proof of proposition IV.2. We will prove: not 1 $\Rightarrow$ not 2. So, let us assume that z is not a resistant Condorcet winner. We assume however that z is a Condorcet winner, otherwise it is trivial. We will prove that there exists a state-based voting system f, meeting the Condorcet criterion, such that f is manipulable in ω.

Since z is not a resistant Condorcet winner, at least one of conditions (3) or (4) from definition IV.1 is not met. We distinguish three cases: condition (3) is not met for some $d = c$; condition (4) is not met; condition (3) is not met with $d \neq c$.

In all three cases, the idea of the proof is the same: we exhibit a situation ψ that has no Condorcet winner and that differs from ω only by voters preferring c to z (in case 2, up to exchanging the roles of c and d). So, it is possible to choose a state-based voting system f that meets the Condorcet criterion and such that $f(\psi) = c$. From this, we deduce that f is manipulable in situation ω towards ψ, in favor of c.

Case 1 If condition (3) is not met for some $d = c$, let r a strict total order where c is preferred to z, who is preferred to all other candidates. For any voter $i \in \mathcal{V}_n$ such that $c \succ_{R_i(\omega_i)} z$, we take $\psi_i \in R_i^{-1}(r)$, which is possible because $\mathfrak{S}_m \subset R_i(\Omega_i)$. For the other voters, we take $\psi_i = \omega_i$. Candidate z is not a Condorcet winner in ψ because $|z \succ_{R_i(\psi_i)} c| = |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \succ_{R_i(\omega_i)} c| \leq \frac{n}{2}$. Candidate c cannot be a Condorcet winner (lemma III.8) and neither can other candidates, because the number of voters who claim preferring z to them has not diminished.

Case 2 If condition (4) is not met, let us notice that $d \neq c$, otherwise z would not be a Condorcet winner. Up to switching roles between c and d, we can assume that d has no strict victory against c. Let r be a strict total order where c is preferred to d, who is preferred to z, who is preferred to all other candidates. For any voter $i \in \mathcal{V}_n$ such that $c \succ_{R_i(\omega_i)} z$, we take $\psi_i \in R_i^{-1}(r)$, which is possible because $\mathfrak{S}_m \subset R_i(\Omega_i)$. For all other voters, we take $\psi_i = \omega_i$. Candidate z is not a Condorcet winner in ψ because she has a defeat against d: indeed, $|\text{not}(d \succ_{R_i(\psi_i)} z)| = |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and not}(d \succ_{R_i(\omega_i)} z)| < \frac{n}{2}$. Candidate c cannot be a Condorcet winner (lemma III.8), neither can candidate d because she still has no strict victory against c and neither can other candidates, because the number of voters who claim preferring z to them has not diminished.

Case 3 Remains the case where condition (3) is not met for some $d \neq c$. Using the previous case, we may assume that however, condition (4) is met. In final situation ψ, we will ensure that there is neither victory for z against d, nor for d against z.

Let r be a strict total order with c first, then d, then z, then all other candidates. Let r' be a strict total order with c first, then z, then d, then all other candidates.

Let us note $B_{cd} = |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \succ_{R_i(\omega_i)} d|$.

Since z is a Condorcet winner, we have:

$$|z \succ_{R_i(\omega_i)} d| > \frac{n}{2},$$

$$|c \succ_{R_i(\omega_i)} z \text{ and } z \succ_{R_i(\omega_i)} d| > \frac{n}{2} - B_{cd}.$$

And, since condition (3) is not met for this pair (c, d), this last term is nonnegative. As a consequence, we may choose $\lfloor \frac{n}{2} \rfloor - B_{cd}$ voters among those who prefer

c to z. For each i of them, let us take $\psi_i \in R_i^{-1}(r')$, which is possible because $\mathfrak{S}_m \subset R_i(\Omega_i)$. For other voters i who prefer c to z, let us take $\psi_i \in R_i^{-1}(r)$. Lastly, for all other voters (who do not prefer c to z), we take $\psi_i = \omega_i$.

Then, we have:

$$D_{vd}(\psi) = B_{cd} + \left(\lfloor\frac{n}{2}\rfloor - B_{cd}\right) = \lfloor\frac{n}{2}\rfloor,$$

hence z has no victory against d.

By the way, (3) is not met for this pair (c, d) but (4) is met. Hence we have:

$$\begin{cases} |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \ggcurly_{R_i(\omega_i)} d| \\ \quad + |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \gtrsim\!\!\!\!\lessapprox_{R_i(\omega_i)} d| \leq \lfloor\frac{n}{2}\rfloor, \\ |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \ggcurly_{R_i(\omega_i)} d| \\ \quad + |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \not\gtrless_{R_i(\omega_i)} d| \geq \lceil\frac{n}{2}\rceil, \end{cases}$$

so, by subtraction:

$$\begin{aligned} &|\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \gtrless_{R_i(\omega_i)} d| \\ &\quad - |\text{not}(c \succ_{R_i(\omega_i)} z) \text{ and } z \not\gtrless_{R_i(\omega_i)} d| \leq \lfloor\frac{n}{2}\rfloor - \lceil\frac{n}{2}\rceil. \end{aligned}$$

Using proposition III.1, we deduce:

$$\begin{aligned} D_{dv}(\psi) &= n + |z \gtrless_{R_i(\psi)} d| - |z \not\gtrless_{R_i(\psi)} d| - D_{vd}(\psi), \\ &\leq n + \lfloor\frac{n}{2}\rfloor - \lceil\frac{n}{2}\rceil - \lfloor\frac{n}{2}\rfloor, \\ &\leq \lfloor\frac{n}{2}\rfloor, \end{aligned}$$

hence d has no victory against z.

To sum up, neither z nor d can be a Condorcet winner. And, for the same reasons as in previous cases, neither can c nor other candidates.

The reader may notice that, if relations $R_i(\omega_i)$ are antisymmetric and complete, conditions (3) or (4) from definition IV.1 are equivalent, so the proof is reduced to case 2 only.

F Classical voting systems and the resistant-Condorcet criterion

In this appendix, we finish the proof of corollary IV.5 by showing that the voting systems mentioned do not respect the resistant-Condorcet criterion (for some values of n and m). In the following examples, preferences are strict total orders, with the most liked candidate on top.

Counterexample 1 In the following table, each group has approximatively the same number of voters (not exactly, to avoid questions about tie-breaking).

	Group of voters				
	1	2	3	4	5
Preferences	A	B	C	D	E
	V	V	V	V	V
	Others	Others	Others	Others	Others
	⋮	⋮	⋮	⋮	⋮

Candidate V is preferred to any other pair of candidates by about 60% of the voters, so V is a resistant Condorcet winner.

However, she is elected neither by IRV, nor by plurality, nor by the two-round system.

For voting systems with grades (including approval voting), we consider the case where the sincere opinion of the voters is to put the maximum grade to their most liked candidate and the minimal grade to all others. Then, we see that V is neither elected by approval voting, nor by range voting (with median or average).

Counterexample 2 In the following table, each group has approximatively the same number of voters, but group 0 is slightly larger than the others.

	Group of voters						
	0	1	2	3	4	5	6
Preferences	A	A	B	C	D	E	F
	B	V	V	V	V	V	V
	C	Others	A	A	A	A	A
	D		Others	Others	Others	Others	Others
	E	⋮	⋮	⋮	⋮	⋮	⋮
	F						
	V	B	C	D	E	F	A

Candidate V is preferred to any other pair of candidates by about 4/7 of the population, so V is a resistant Condorcet winner.

The average ranking of candidate A is $(1 \times 2 + 3 \times 5)/7 = 17/7$ and the one of V is $(7 \times 1 + 2 \times 6)/7 = 25/7$ so V is not elected by the Borda method.

Using the Coombs method, candidate V is eliminated in the first round therefore she is not elected.

Counterexample 3 The candidates are $\{\mathrm{A}, \mathrm{B}_1, \mathrm{B}_2, \mathrm{C}_1, \dots, \mathrm{C}_8\}$. The size of each group of voters is given in proportion of the whole population.

	Size of the group of voters						
	10%	10%	40%	10%	10%	10%	10%
Preferences	A	A	V	C_1	C_3	C_5	C_7
	B_1	B_2	A	C_2	C_4	C_6	C_8
	V	V	Others	V	V	V	V
	Others	Others		Others	Others	Others	Others

Candidate V is preferred to any pair $\{A, B_i\}$ by 80% of the voters, to any pair $\{A, C_i\}$ by 70% of the voters and to any pair $\{B_i, C_j\}$ by 80% of the voters, so V is a resistant Condorcet winner.

If we use the Bucklin method, processing stops in the second round: at this stage, candidate A has 60% of ballots, V has 40% and each other candidate has 10% (it is normal that the total is 200%). So A is elected.

References

Michel Balinski and Rida Laraki.
Majority Judgment: Measuring, Ranking, and Electing.
MIT Press, 2010.

Duncan Black.
The theory of committees and elections.
University Press, Cambridge, 1958.

Steven Brams and Peter Fishburn.
Approval voting.
American Political Science Review, 72:831–847, 1978.

John R. Chamberlin, Jerry L. Cohen, and Clyde H. Coombs.
Social choice observed: Five presidential elections of the american psychological association.
The Journal of Politics, 46:479–502, 1984.

François Durand, Fabien Mathieu, and Ludovic Noirie.
Manipulability of voting systems.
2012.
URL http://www.liafa.univ-paris-diderot.fr/~displexity/docpub/6mois/votes.pdf.

Pierre Favardin and Dominique Lepelley.
Some further results on the manipulability of social choice rules.
Social Choice and Welfare, 26:485–509, 2006.

Pierre Favardin, Dominique Lepelley, and Jérôme Serais.
Borda rule, copeland method and strategic manipulation.
Review of Economic Design, 7:213–228, 2002.

Allan Gibbard.
Manipulation of voting schemes: A general result.
Econometrica, 41(4):pp. 587–601, 1973.

James Green-Armytage.
Four condorcet-hare hybrid methods for single-winner elections.
Voting matters, 29:1–14, 2011.

James Green-Armytage.
Strategic voting and nomination.
Social Choice and Welfare, 42(1):111–138, 2014.

James Green-Armytage, Nicolaus Tideman, and Rafael Cosman.
Statistical evaluation of voting rules, 2014.

John G. Kemeny.
Mathematics without numbers.
Daedalus, 88:575–591, 1959.

Dominique Lepelley and Boniface Mbih.
The proportion of coalitionally unstable situations under the plurality rule.
Economics Letters, 24(4):311 – 315, 1987.

Dominique Lepelley and Boniface Mbih.
The vulnerability of four social choice functions to coalitional manipulation of preferences.
Social Choice and Welfare, 11:253–265, 1994.

Dominique Lepelley and Fabrice Valognes.
On the Kim and Roush voting procedure.
Group Decision and Negotiation, 8:109–123, 1999.

Dominique Lepelley and Fabrice Valognes.
Voting rules, manipulability and social homogeneity.
Public Choice, 116:165–184, 2003.

Dominique Lepelley, Ahmed Louichi, and Hatem Smaoui.
On ehrhart polynomials and probability calculations in voting theory.
Social Choice and Welfare, 30:363–383, 2008.

Hervé Moulin.
La Stratégie du vote.
Cahiers du CEREMADE. Université Paris IX-Dauphine Centre de recherche de mathématiques de la décision, 1978.

Bezalel Peleg.
Game theoretic analysis of voting in committees.
Number 7 in Econometric Society monographs in pure theory. Cambridge Univ. Press, Cambridge [u.a.], 1984.

Geoffrey Pritchard and Mark Wilson.
Exact results on manipulability of positional voting rules.
Social Choice and Welfare, 29:487–513, 2007.

Donald G. Saari.
Susceptibility to manipulation.
Public Choice, 64(1):pp. 21–41, 1990.

Mark A. Satterthwaite.
Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions.
Journal of Economic Theory, 10(2):187–217, 1975.

Markus Schulze.
A new monotonic, clone-independent, reversal symmetric, and Condorcet-consistent single-winner election method.
Social Choice and Welfare, 36:267–303, 2011.

Arkadii Slinko.
How large should a coalition be to manipulate an election?
Mathematical Social Sciences, 47(3):289 – 293, 2004.

David A. Smith.
Manipulability measures of common social choice functions.
Social Choice and Welfare, 16:639–661, 1999.

Alan D. Taylor.
Social choice and the mathematics of manipulation.
Outlooks Series. Cambridge University Press, 2005.

Nicolaus Tideman.
Independence of clones as a criterion for voting rules.
Social Choice and Welfare, 4:185–206, 1987.

Nicolaus Tideman.
Collective Decisions And Voting: The Potential for Public Choice.
Ashgate, 2006.

H. P. Young and A. Levenglick.
A Consistent Extension of Condorcet's Election Principle.
SIAM Journal on Applied Mathematics, 35(2), 1978.