Rousseau's General Will as a Bound for Strategic Self-Interested Voting

GUSTAVO ADRIÁN BODANZA
Universidad Nacional del Sur
Consejo Nacional de Investigaciones Científicas y Tecnológicas

SEBASTIÁN LINARES LEJARRAGA
Universidad Nacional del Sur
Consejo Nacional de Investigaciones Científicas y Tecnológicas

DOI: 10.36446/rlf2022263

Abstract: Condorcet's voting method, as its extension known as the Kemeny-Young rule, is often seen as the incarnation of Rousseau's general will. But the Kemeny-Young rule has problems when cycles in the social ranking arise. In particular, it can lead to choose a "covered" alternative, i.e. one for which there are candidates with better performances in pairwise comparisons. On the other hand, the *uncovered set*, the set of all the alternatives that are not covered, puts limits to insincere votes of sophisticated voters under certain conditions. Basically, voting insincerelly for a non-preferred alternative in order to favor the actually preferred candidate would lead to the choice of an uncovered alternative. We argue that Rousseau's general will could be embodied in a different kind of rule than

Kemeny-Young's, with both epistemic credentials and whose outcomes are within the uncovered set altogether. In this work we begin to explore that possibility.

Key-words: Rousseau, Condorcet, social choice, epistemic voting, strategic voting.

La voluntad general de Rousseau como un límite al voto estratégico autointeresado

Resumen: El método de votación de Condorcet, así como su extensión conocida como la regla Kemeny-Young, suele verse como la encarnación de la voluntad general de Rousseau. Pero la regla Kemeny-Young tiene problemas cuando surgen ciclos en el ranking social. En particular, puede llevar a elegir una alternativa "cubierta", es decir, una para la que hay candidatos con mejores resultados en las comparaciones por pares. Por otro lado, el *uncovered set*, el conjunto de todas las alternativas que no están cubiertas, pone límites a los votos estratégicos sofisticados bajo ciertas condiciones. Básicamente, votar de manera no sincera por una alternativa no preferida para favorecer al candidato realmente preferido no puede llevar a la elección de una alternativa cubierta. Argumentamos que la voluntad general de Rousseau puede implementarse en un tipo de regla diferente a la de Kemeny-Young, con credenciales epistémicas y cuyos resultados se encuentran dentro del *uncovered set*. En este trabajo comenzamos a explorar esa posibilidad.

Palabras clave: Rousseau, Condorcet, elección social, voto epistémico, voto estratégico.

1. Introduction

In a famous passage of the *Social Contract*, Rousseau expresses that majority is legitimate because it expresses the *general will*:

> When a law is proposed in the people's assembly, what is asked of them is not precisely whether they approve of the proposition or reject it, but whether it is in conformity with the general will (…) each by giving his vote gives his opinion on this question, and the counting of votes yields a declaration of the general will. When, therefore, the opinion contrary to my own prevails,

> this proves only that I have made a mistake, and that what I believed to be the general will was not so (Rousseau 1962: 153).

In a widely celebrated paper, Peyton Young (1986) claimed that this vague idea was given expression twenty years later by Condorcet in his work *Essai sur l'application de l'analyse a la probabilité des decisions rendues a la pluralité de voix* (Condorcet 1785). Condorcet argued that voters will honestly attempt to judge what decision will best serve society. They may judge wrongly, but assuming that more often right than wrong, the majority opinion will very likely be "correct". Indeed, if we keep on increasing the number of voters, the probability that a majority will select the right alternative tends to infallibility.[1]

The winner alternative under Condorcet's method is the one which beats all the other alternatives in pairwise comparisons. If voters have more than one half of probability of being right, then the Condorcet winner can claim epistemic credentials. But the Condorcet winner is not a well-defined concept, meaning that some judgment profiles could not yield one, as happens in situations of cyclic majorities. Condorcet attempted to show how his method could be extended to the case of more than two alternatives, and he was the first to discover the possibility of cyclical majorities. Young showed that Condorcet's goal can be translated –with more than two alternatives– into a statistical procedure for estimating *the ranking of the candidates that is most likely to be correct*, which is equal to Kemeny's rule (this equivalence leading to call it the "Kemeny-Young method").

Young is correct in the interpretation of Condorcet's method with more than two alternatives. However, could its extension, the Kemeny-Young method, be characterized as a materialization of Rousseau's thought? We are going to argue that it does not fit well with Rousseau's dictum that voters in the minority should recognize that they had been wrong, and therefore either follow the majoritarian opinion or suspend judgment. We would like to show that, if followed literally, Rousseau's thought would be better embodied in a different kind of method or algorithm for estimating the most likely correct ranking or alternative. Minority voters appear to be in the wrong about the general will and recognize majority voters as better judges

[1] Roughly, the Condorcet Jury Theorem says two things. First, the majority vote among a group of (independent, competent, sincere) voters, each of whom is more likely to be right than not, is itself more likely to be right than are individual voters separately. Second, as the number of such voters approaches infinity, the probability that the majority among them is correct approaches one (see Goodin and Spiekermann 2018).

when handing over judgment to the majority. The majority is right and the minority is wrong. Should this be interpreted literally? If so, it seems legitimate to be consequent with that deference commitment of the minority in favor of the majority, and their votes should be dismissed. But then, Young's *maximum likelihood estimation* (and equivalently, Kemeny's rule) is not a suitable characterization of Rousseau's thought, since it assigns voters on the "wrong" side a probability greater than 0 of being right, which leads to a very different result than that obtained by ignoring their votes. In sum, to extend Condorcet's approach to more than two alternatives, one can think of a more satisfying method in accordance with Rousseau's idea of the minority deference commitment.

The possibility of majority cycles is also a point to consider. Social choice theorists have widely studied methods that, in situations of majority cycles, could select single alternatives or transitive rankings. But then other problems arise. Some methods are not Condorcetian (i.e. there exists a Condorcet winner but is not the chosen alternative). Others are manipulable by strategic voters who advance a non-preferred alternative that improves the chance of getting a preferable outcome. Under certain conditions, with full information about voter preferences, and voters with sincere and self-interested preferences, an agenda setter might obtain any outcome she wants by manipulating the agenda (McKelvey 1976, 1979). However, Miller (1980) and later McKelvey (1986) showed that, if we drop sincerity and assume sophisticated (i.e. strategic) voters with complete information on the alternatives, the set of sophisticated equilibrium outcomes –through voting by successive elimination– must lie inside a central bunch of alternatives, labeled the *uncovered set*. Roughly, alternative A covers alternative B if A beats (i.e. is preferred by the majority to) B and every alternative beaten by B. The uncovered set is the set of all uncovered alternatives. The results show that the uncovered set puts some bounds on the monopoly power of an agenda setter in the context of multidimensional spaces and amendment agendas, introducing a relative limit to the general instability of possible outcomes. Cycles may still persist, but indecision is limited to a reduced number of cases. And not less important, the Condorcet winner, if it exists, is among them.

Thus, the uncovered set seems a good candidate to evaluate Condorcetian methods. Now, are Rousseau's voters sophisticated? As Grofman and Feld (1988) pointed out, Rousseau recognized the possibility of individuals expressing their private will, but that could only be by agreement with the general will, an agreement that could not be lasting and constant. Anyway, it is worth exploring whether, departing from an epistemic model in accordance with Rousseau's spirit, it is possible or not to always guarantee the selection

of an alternative that, under certain preference profiles, are placed inside the uncovered set. Intuitively, there is no reason to think that any epistemic method, however designed, can assure the choice of an uncovered alternative. Epistemic methods could display outcomes divergent from those that could be predicted under strategic behavior by self-interested voters, which would confirm that the rationality of strategies is essentially different from the rationality of *judgments about the common good*. But an epistemic method that always yields an uncovered alternative as the social choice outcome would also set a limit on strategic voting, since sincere voters would not have strong incentives to misrepresent their true preferences. In that sense, such a method would be doubly virtuous.

The main point we defend in this work is to give an affirmative answer to the question in the title. In order to justify the point, will we explore a simple method that can be assumed to agree with Rousseau's theory in at least the following two aspects: i) it is epistemic: votes are judgments about what the general will is; and ii) minorities defer their votes in favor of the majority and, in consequence, of the general will. Moreover, iii) it is rational: on one hand, individuals are rational in the elementary sense that they can rank the alternatives in a transitive order; on the other hand, the chosen alternative must belong to the uncovered set, which puts rational bound to self-interested strategies.

The paper is organized as follows. We begin by reviewing the epistemic account on voting, arguing that, while Condorcet's method has been considered the materialization of Rousseau's ideals, the Kemeny-Young extension does not fit well. Looking for the best ranking is not the same task that finding the best alternative. Borda count can serve as an approach to the latter, but then we have to abandon the idea of always capturing the Condorcet winner. Next, we briefly review some passages of Rousseau's *Social Contract* to argue that voting is essentially an epistemic enterprise, and Kemeny-Young's is not a good method for tracking the general will insofar the outcomes often cannot be matched with rational voter's judgment profiles. Moreover, we argue that Rousseau's idea that minority should surrender its judgment in favor of the majority is not captured by the Kemeny-Young method, but can be modeled in a voting system that computes only majoritarian votes in pairwise comparisons. After those considerations, we explore a new method that consists in calculating the proportional majoritarian support of each alternative and picking that with the greatest result, and argue that the method could satisfy the expected features. Next we move from the epistemic view to the strategic one to argue that the proposed method also puts bounds to agenda setters and self-interested voters, via the uncovered set. Finally, we summarize our principal arguments and conclude that the

proposed method is capable of dealing with epistemic voting while giving stability to strategic voting.

2. The epistemic account on voting

For Condorcet, as for all epistemic approaches to democracy, the object of voting is not merely to balance subjective opinions or satisficing self-interested preferences; rather, it is a collective quest for truth, or knowledge (Young 1986, 1995; Cohen 1986; Schwartzberg 2015). Although it is controversial if truth aptness is attributable to *all* political decisions, we are going to put that question aside. For the present we are going to grant –in accordance with Young (1988)– Condorcet's hypothesis that the notion of a "correct" decision also applies to many political decisions, and that all voters have equal epistemic competence higher than 0.5. Thus, voters evaluate candidates in terms of their competence for the position.

The properties of social choice functions on two alternatives are quite well understood, and in general we may say that simple majority is the most natural rule in this case. In particular, May (1952) has given the standard axiomatization of simple majority. But there is no natural extension of simple majority rule to three or more alternatives, because the application of simply majority rule to all pairs of alternatives (tournaments or pairwise comparisons) may lead to cycles (i.e., A beats B, B beats C, C beats A: the famous voting paradox). However, Condorcet proposed that if there is some alternative that obtains a simple majority over every other, then that alternative should be chosen as the winner. This principle is known as the "Condorcet's rule", and any such alternative is known as a "Condorcet winner".

What voting rules are most likely to yield good outcomes in cases with more than two alternatives? Ideally, we would like to choose the alternative that has a simple majority over every other (i.e. the Condorcet winner), but such an alternative may not exist. Moreover, invoking any other method based on pairwise comparisons (v.g. Copeland, that selects as the winner(s) the alternative(s) that has/have more victories in pairwise comparisons) is not helpful in order to solve this problem from an epistemic point of view, just because majoritarian victories in pairwise comparisons are not epistemic sources in their own, and epistemic credentials are only derivative from the epistemic reliability of individuals.

Assuming an equal epistemic competence and an individual probability higher than 0.5 of getting a true vote in any situation, Young (1988, 1995) showed that the best way to interpret Condorcet's original purpose with more than two alternatives, is to try to find the *ranking* most likely to be

correct. With that goal in mind, the most likely best ranking turns out to be the one that has *maximum pairwise vote support,* which is the ranking selected by Kemeny's rule (Kemeny 1959).

Kemeny's rule avoids the problem of voting cycles in the social output (not because voting cycles disappear, but because by definition Kemeny's rule always brings about transitive rankings), although it may lead to ties between rankings. Inherent in that rule is the idea of being "close" to majority rule. Intuitively, ranking the alternatives according to Kemeny's rule can be seen as the best compromise in the sense that, on average, gives the "closest" social preference to all possible individual rankings given individual voter profiles, and also minimizes the inconsistencies between adjacent alternatives in the social ranking with individual rankings.[2] Young and Levenglick (1978) also proved that Kemeny's rule is the unique preference function that is, at the same time, neutral[3], consistent[4] and Condorcet efficient (it picks, as the top alternative in the best ranking, the Condorcet winner when there is one). It also shows certain stability when irrelevant alternatives are withdrawn or added (Young 1986): the ranking on a subset of alternatives is stable if we remove from consideration a block of bottom ranked alternatives (or a block of top ranked alternatives).

However, all these properties are met at certain price: the price of "weakening the crucial assumption about the individual rationality of the voters" (Saari and Merlin 2000: 404). It can be argued that Kemeny's rule misinterprets certain voters' real preferences, as coming from a group of voters that do not exist, or that they exist but support individual cyclical preferences.

Remember, from an epistemic point of view, individual preference orderings are the outcome of judgments *about the common good*. Though it might be disturbing to interpret a particular hierarchical relation between

---

[2] Kemeny's rule is equivalent to searching an average among rankings. The operation begins with a notion of *distance* between any two rankings. Define this distance as the number of pairs with regard to whose ranking they differ. The distance between (A, B, C) and (B, A, C) is 1: they differ only with regard to {A, B}. A suitable conceptualization for an average of rankings is their median relative to this metric, *the ranking minimizing the sum over the distances from* the rankings. The ranking minimizing the sum of distances "and" with most support, is the Kemeny Ranking.

[3] Which states that, given certain preference profile, the social choice is invariant to changing the identity of the alternatives.

[4] Which states that, given two sets of voters with preference profiles and 0, respectively, if the set of alternatives selected by a choice rule intersects on and 0, then the alternatives selected after the votes are combined must be this intersection.

adjacent alternatives inside the social ranking as coming from cyclical individual preferences, it should not cause too much worry if it is interpreted as coming from non-existent voters. Perhaps, although it remains to be studied, it might be argued that some judgmental relations between adjacent alternatives are the supervening judgmental output of the wisdom of crowds. Or, more weakly, it might be argued that if, as it happens, those particular "inversions" or "swaps" between adjacent alternatives in the social ranking –with respect to real individual majoritarian rankings– are always placed below the top, it is the epistemic price to pay for getting it right –about facts– at the top. Can these interpretations be traced back to Rousseau's insights? We have not found any reason to infer from Rousseau's writings that the general will steams from non-existent or irrational voters, so if we are seeking for a materialization of Rousseau's thought it is worth considering existent rational voters.

From an epistemic point of view, the problem with Kemeny's rule is not just that it could end in a draw between so many rankings, or that it could partially violate the rational consistency (transitivity) of individual rankings. The problem is that we can question the very goal of finding the best overall *ranking*. Why not just try to find the *best alternative*, irrespective of its relative position in a given ranking?

Consider Borda count. Since the Borda winner is not necessarily the alternative placed in the highest position in the Kemeny ranking, one is bound to infer that selecting the best ranking is a different task to that of selecting the best alternative (Balinski and Laraki 2010: xii). And what is more disturbing from an epistemic point of view, we don't have a single privileged method, since Kemeny and Borda are prone to bring about different social transitive rankings with the same profile of individual preferences. It is well known that the Borda count does not guarantee a Condorcet winner, if it exists. Conversely, when the Condorcet winner and loser are defined, they are, respectively, top and bottom ranked by Kemeny's rule (Saari and Merlin 2000: 418). However, the Borda count only ensures that the Condorcet winner is strictly ranked above the Condorcet loser. It is also proved that the Borda count always ranks the Kemeny top-ranked candidate strictly above the Kemeny bottom-ranked candidate. Conversely, Kemeny's rule ranks the Borda top-ranked candidate (or Borda winner) strictly above the Borda count bottom-ranked candidate or Borda loser (Saari and Merlin 2000: 418). The stability of both methods are also quite different: while Borda count is extremely vulnerable to the addition or withdrawal of irrelevant alternatives (sometimes bringing about the complete reversion of the social ranking after the withdrawal of an irrelevant alternative), it was proved that the withdrawal of a block of top ranked adjacent alternatives or a block of

bottom ranked adjacent alternatives –placed in the Kemeny ranking– leaves the order of the remaining alternatives untouched (Young 1988).

In addition to those formal relations between Borda count, on one side, and the methods of Condorcet and Kemeny, on the other, we also want to focus on a substantial, ontological difference. Condorcet's method uses votes as a mean to build a tournament from which counting victories is a derivative calculation. In fact, there exists a vast literature on tournament solutions as a basically qualitative problem totally independent of how tournaments are built (Brandt 2010).[5] In Borda count, instead, the number of votes is an essential quantitative information for the whole process that is not contained in a tournament. And that quantitative information can tell us in what extent or proportionality a candidate or ranking gets majoritarian support. Young's maximum likelihood estimation and, by equivalence, also Kemeny's rule, build a bridge between both ontologies. Still, doubts about whether Kemeny's rule extends Condorcet method in accordance with Rousseau's general will persist. Insofar as minority votes are counted, we obtain a different proportionality than if we counted only majority votes. So, the question we raise is: Is there any overcoming way to synthesize the epistemic aptness of the Condorcet method with a deference commitment to majority votes? Next, we reconsider Rousseau's theory of voting seeking support for an affirmative answer to that question and guidelines for a novel approach.

3. Reconsidering Rousseau's theory of voting

Rousseau has been declared by many scholars as the patron saint of epistemic democrats (Cohen 1986, Young 1988, Schwartzberg 2015, Waldron 1990, Estlund 2008, Pettit 2016). Although there is some controversy about him being really an epistemic democrat (who asks voters to follow judgments about facts on the best policy) or, instead, a non-epistemic, agonistic democrat (who ask voters simply to follow the general nude "will" on the preferred policy; see Tuck 2019), he is, in our opinion, a clear enthusiast of epistemic democracy, properly defined along the canonical lines settled by Cohen (1986).

Curiously, what seems to be undisputed in the literature is that Condorcet's theory of voting is, indeed, the best operative incarnation of

[5] Note, for instance, that any profile of transitive rankings can lead to a tournament via majoritarian pairwise comparisons, but there may exist no such a profile from which to build up a given tournament.

Rousseau's theory of the general will along the epistemic lines (Barry 1967, Cohen 1986, Grofman and Feld 1988, Young 1988, Schwartzberg 2015, Estlund 2008). On the contrary, we are going to argue that Condorcet's theory of voting must be considered fundamentally distinct from Rousseau's theory: the basic epistemological premises that Rousseau defended in the *Social Contract* are not properly captured by Condorcet's statistical approach (i.e. one that gives operative justification, as we explained, to Kemeny's rule). Instead, we believe that Rousseau's basic premises provide justification for a very different operative statistical estimation account, the mechanics of which we will explain in detail below.

Let's begin with the tenets of Rousseau's theory of the general will. In addition to the passage cited in the introduction, the following important paragraph comes from Book II, chapter 3. In outlining the difference between the will of all and the general will, Rousseau says:

> The latter [ie. the general will] looks only to the common interest, the former [i.e. the will of all] looks to private interest, and is nothing but a sum of particular wills; but if, *from these same wills, one takes away the pluses and the minuses which cancel each other out*, what is left as the sum of the differences is the general will (emphasis added).

There are many possible interpretations of these and other important paragraphs on the notion of general will from the *Social Contract*, and our aim is not to dismiss the validity of other readings. However, in our understanding, the general will of Rousseau is compatible with the will of an un-factionalized assembly where each voter searches for the truth about a matter of fact. That is why "when the opinion contrary to my own prevails, it proves nothing more than that I made a mistake and that what I took to be the general will was not". It can, therefore, be interpreted as an epistemic enterprise. Thus, the general will may be expected to track, through the aggregation of votes sustained in judgments (perhaps preceded by deliberation), the common interest of citizens, not the interests of this or that individual or subgroup: "it is concerned with their common preservation, and the general welfare", as it is said.

Now, voters may defend diverse judgments over what is best for the common interest, and some may vote for A, some others may find reasonable to vote for B, others for C, and so on. These disagreements may lead to contingent and variable splits between majorities and minorities on agenda voting issues, with no permanent identifiable minority on each and every issue (otherwise, that fact would be evidence of a factionalized assembly). Democracy involves some ruling others, *notwithstanding* the *disagreements* that might continue to exist between sequences of majorities and minorities (of

different size and people across issues). Roughly, on standard liberal accounts, the majority on any decision *rules over the minority*. However, Rousseau tries to show, to the contrary, that in a proper democracy, each "obeys no one but himself". How could this be achieved? Rousseau gives a provocative answer: by asking minority voters to surrender judgment to the majority judgment. A problem with Rousseau's idea of deference comes when we try to justify in what conditions minority voters should surrender judgment, and what does it mean. Does it mean only that they should give less credence to their beliefs (perhaps to the extent that they must suspend judgment), does it mean that they have epistemic reasons not only to suspend judgment but also to uphold the majority belief, or does it mean to follow *blindly* the majority belief?

From an epistemological point of view, in political matters where normative principles and values are always at stake, the idea of feeling obliged to surrender judgment, *just because* we are in a minority, is intriguing to say the least. If we are convinced that a certain person is clearly superior compared to us on many disagreement factors when it comes to answering the question "Is A good, or better, compared to B, C, or D?" then we'll probably say she is more likely to answer the question correctly. However, we would suspend judgment *not because she has more followers* than we have, but just because we acknowledge she is an expert on the dimensions that are relevant to decide the issue at stake. But in political matters where competent experts usually are in both sides, or where there are many dimensions involved pushing to divergent directions, the reasons for surrendering judgment are less clear. Citizens in a democracy have a *moral* duty to obey a binding rule coming from democratic voting. As Schwartzberg argues, the outcome under majority vote "possess moral force insofar as [it] induced outvoted minorities to recognize that they had likely erred" (Schwartzberg 2008: 406). But they do not have an *epistemic* duty to submit their judgment to it (cf. Estlund 2008: 103 ss.; Rawls 1950: 319).[6]

This is the problem of deference faced by Rousseau's epistemic approach to voting, a problem that –we argue– doesn't fit well with Condorcet's theory of voting and its byproduct, Kemeny's voting rule, as long as Kemeny's rule aggregates –while searching for the "average" ranking–

[6] Rousseau's deference conception seems to be defended in Nino (1991). However, Nino goes beyond Rousseau's dictum that the minority voter should only recognize that "she was wrong" on the alternative supported, but also asks the voter to "believe" that the alternative supported by the majority tends to be right. For example, "The democratic origin of a legal rule provides us with a reason to believe that there is good reason to accept its content and to act accordingly" (Nino 1991: 255).

not only majority votes, but also minority votes. Remember that when the voting agenda comprises more than two alternatives, Condorcet's theory leads to the application of Kemeny's rule, which stands as a reasonable rule for estimating the most probable correct ranking. But Kemeny's rule counts minority votes, not only majoritarian votes. This counting of minoritarian votes, we argue, is in stark contrast with Rousseau's theory of voting.

To be honest with Rousseau's thought, any statistical account on the probability of getting it right must portray in some way this theory of deference. And we think that that deference commitment could be perfectly introduced in an algorithm, by setting a simple condition: *minority votes on pairwise comparisons should not be counted*, insofar they are expressions of a *wrong* epistemic judgment.

Let us go back to Young (1988, 1995), whose interpretation of Rousseau's thought as the philosophical source of Condorcet's epistemic theory of voting has become canonical in the literature. Young argued that, when interpreting the phrase "*from these same wills, one takes away the pluses and the minuses which cancel each other out*, what is left as the sum of the differences is the general will" (emphasis ours), the only mathematical sense that we can make from that is to look up for the workings of something like the jury theorem elaborated by Condorcet. Young thinks that the way in which that theorem involves error (which is particular to each individual and randomly distributed among them) cancelling, and leaving truth (which is common to all) as the remainder, is thus the best operative embodiment of Rousseau's theory of voting. We want to bring up for discussion another interpretation of Rousseau's minority deference commitment. Error in Condorcet's theory comes from an a priori definition of individual competence. In contrast, error in Rousseau's theory comes from an a priori definition of competence *and* from the *ex post* awareness of being in the minority side. Sources of error are philosophical different, and that divergence should be consequential.

4. Proportional majoritarian support

How can we formulate an algorithm compatible with Rousseau's idea about the minorities' deference commitment? There surely exist many ways to carry out the task. But we also want the method to satisfy the other reasonable conditions that we have been discussing in accordance with Rousseau's thought. First, the epistemic virtues of the Condorcetian method should be respected, in the sense that if a Condorcet winner exists, then it should be chosen. Moreover, "taking away the pluses and minuses that cancel each other" supposes, in principle, that the choice should take

into account the *quantities* of the votes, and not just how the candidates are beaten in a tournament. This is an important difference with Condorcet's method (similarly, Copeland's, and others) where votes are only a mean to reach the tournament in which victories are counted. So, how votes are distributed in the preference profile is crucial. In addition, we should assume that voters are rational agents that always posit transitive rankings of alternatives. Once the majorities are identified in the pairwise comparison matrix, we calculate, for each alternative, the probability of that alternative being elected by the majority against *any* other alternative. Finally, the result should be such "what is left as the sum of the differences is the general will", i.e. the alternative with greatest proportional majoritarian support is chosen.

Let us see how our proposal works with an example. Consider a setting with eleven voters and alternatives A, B, C, D, E, F, and G. Adhesions to rankings are distributed as shown in Table 1, with alternatives ranked in top-down order:

Table 1

5v	5v	1v
A	B	B
C	A	G
D	C	F
E	D	E
F	E	D
G	F	C
B	G	A

This profile yields the following pairwise comparison matrix (the cells indicate the number of votes obtained by the alternative in the row heading against the alternative in the column heading):

Table 2

	A	B	C	D	E	F	G
A		5	**10**	**10**	**10**	**10**	**10**
B	**6**		**6**	**6**	**6**	**6**	**6**
C	1	5		**10**	**10**	**10**	**10**
D	1	5	1		**10**	**10**	**10**
E	1	5	1	1		**10**	**10**
F	1	5	1	1	1		**10**
G	1	1	1	1	1	1	

We proceed as follows. First, in each row we delete all the minority votes, which leaves us only with the numbers in bold. This intends to model the deference commitment of the minority in favor of the majority. Then, we calculate a coefficient for each alternative, call it the *proportional majoritarian support*, which is the sum of all the majoritarian votes in favor of the alternative (bold numbers in the alternative's row) divided by the sum of all the majoritarian votes both in favor and against that alternative (bold numbers in the alternative's row and column). This puts the majority support in relation to all majority opinions. For instance, the coefficient of alternative A is $10\times5/((10\times5)+6) = 50/56 = 0.892$. Doing the same with the remaining alternatives, we find that B is the chosen alternative, having the maximum proportional majoritarian support: $36/36 = 1$.

It is clear that the more majoritarian "pro" votes an alternative gets in the pairwise comparison (i.e. more bold numbers in its row) the less majoritarian "con" votes it gets (i.e. less bold numbers in its column). In consequence, the greater this difference between majority votes for and against, the closer the proportional majoritarian support will be to 1. This remark leads us to conjecture that our method guarantees that the winner alternative (or alternatives, since ties are possible) will never be *covered*[7] by other alternatives, provided that voters are rational agents[8]. It follows that the Condorcet winner, if it exists, will have the maximum proportional majoritarian support. In the above example, there is no Condorcet winner, but we still get a best alternative, B. This is also the alternative with the greatest difference between the number of victories and the number of defeats, which makes it also the Copeland winner (Copeland 1951). Even more important, our method could be more selective than Copeland's as the example of Tables 3 and 4 show.

[7] Define the dominion of alternative x as the set $D(x)$ of all the alternatives that x beats ("x beats y" meaning that x obtains more votes than y in the pairwise comparison). Then x *covers* y if and only if $y \in D(x)$ and $D(y) \subseteq D(x)$. In words, x *covers* y if and only if x beats y and everything y beats. (Miller 1980, 1983).

[8] The proof is left to further work. Nevertheless, our conjecture finds some inductive support from a simulation we run on 100 random voting scenarios, varying over 50 voters and 5 candidates. Moreover, considering only tournaments, every time we found a scenario where the alternative with maximum proportional majoritarian support was covered, the tournament cannot come from a rational (acyclic) preference profile.

Table 3. Individual rankings profile

18v	15v	5v	18v	19v
A	B	E	D	E
B	A	B	E	A
C	D	A	G	C
D	C	F	B	D
E	E	D	A	F
F	G	C	F	G
G	F	G	C	B
H	H	H	H	H

Table 4. Pairwise comparison matrix

	A	B	C	D	E	F	G	H
A versus		37	**75**	**57**	33	**75**	**57**	**75**
B versus	**38**		**56**	**38**	33	**56**	**38**	**75**
C versus	0	19		37	33	**52**	**57**	**75**
D versus	18	37	**38**		**51**	**70**	**75**	**75**
E versus	**42**	**42**	**42**	24		**75**	**75**	**75**
F versus	0	19	23	5	0		**42**	**75**
G versus	18	37	18	0	0	33		**75**
H versus	0	0	0	0	0	0	0	

There are two Copeland winners, B and E, both with a difference of 5 points between victories and defeats, while our method chooses B, whose proportional majoritarian support is 0.877. Once the Copeland's rule focuses on the resulting tournament to solve the outcome, leaving behind the voting tally, an important source of quantitative information is lost. And that information makes the difference to find more accurate results.

Another important relation among Copeland winners and alternatives with maximum proportional majoritarian support involves the *uncovered set* of alternatives (Miller 1980). It includes all the alternatives that are not covered by others. In formal terms, let $C=\{(x, y): x$ covers $y\}$ (see footnote 7). Then, the *uncovered set* of a set of alternatives X is $UC(X) = \{x\in X$: for all $y\in X, (y, x)\notin C\}$. If a Condorcet winner exists, it is the unique element of $UC(X)$. Otherwise, $UC(X)$ includes at least three alternatives (i.e. the set of all the alternatives that are not covered by others). Miller observed that Copeland winners always belong to the uncovered set. A fortiori, the

approach based on the proportional majoritarian support enables in some cases –as the last example shows– a refinement of the uncovered set.

In the next section, we make a case for the importance of the uncovered set as a limit to instability introduced by strategic voting.

5. Setting bounds to self-interested strategic voting

The view of voting as based not on judgment but on naked self-interest is attributable to Bentham, but was later assumed by most standard utilitarian approaches to voting. Bentham was a psychological egoist, and assumed that people always act to further their own satisfactions (Waldron 1990). Provided that people may vote pursuing their self-interested preferences, they may find appropriate –after being aware of other voter preferences– to express insincere individual rankings in their votes, if that strategy leads to a better satisfaction of their true –concealed– preferences. Considered from an utilitarian legislator's point of view, if all participants vote this way, the political choice tends to represent the *correct answer*, since by definition the outcome should maximize the aggregated satisfaction of individual utilities. This view of voting is *non-epistemic* by definition: voters do not vote following their judgments on the best choice, but their naked selfish preferences. Needless to say, the outcome of the procedure may be regarded as true or valid (in the sense that –according to utilitarianism– it might fit with a true or valid principle of justice: the utility principle), although the nature of voting as such, as satisfaction of selfish preferences, remains non epistemic. In this view, correct answers (given by the utility principle) can be derived from majority selfish preferences, and the "general consent" that is brought about by majority rule provides "the surest visible sign and immediate evidence of general utility" (Harrison 1983: 214; Cohen 1986: 28).[9] The important point about this model is that individual votes represent individual satisfactions, and majority vote-counting approximates a social welfare function with individual satisfactions as its non-epistemic support.

According to this model, voters may behave strategically to pursue their self-interests. That would imply that they might occasionally find in their interest to vote first for a candidate or alternative they do not place at the highest level in their preference order, if that strategy has a higher

[9] The available evidence is that Bentham took notice of Condorcet's 1785 essay in 1808 (Elster 2012: 157), so most of his work didn't address the challenges put forward by the epistemic approach of Condorcet.

probability of achieving –given other's exhibited preferences– his first true preference.

This theory is implicit in most positive approaches with which political economy and political science study the strategic behavior of voters and its outcomes, either in elections or legislatures. However, from a philosophical perspective, their difficulties are plain enough. Some are internal to the model (such as that voting cannot possibly be made to reflect the intensity of the satisfaction or dissatisfaction anticipated by individuals with respect to some law), some are external, this latter on which we would like to pay attention. As a predictive theory is almost certainly false. People, whether they are voters or politicians, do not make decisions purely on the basis of self-interest. They are occasionally (we think, often) motivated by their sympathies for others, their own judgments –based on imperfect knowledge– of what would be conducive to the general good, or adherence to some moral ideals.

Now, let's take for granted that people vote pursuing their self-interested preferences, and that the aggregation of votes "as mere preferences", trough majority rule, should be considered the maximization of aggregated utility or preference satisfaction, and therefore that the social outcome would be –according to utilitarian thought– the true or correct social welfare function. Well, it happens that there could be situations in which there is no single outcome, and in which the social outcome violates basic principles of consistency, stability of preferences, and rationality. These problems are summarized in the celebrated *Impossibility Theorem* demonstrated by Kenneth Arrow (Arrow 1951), the details of which are widely known and we have no space here to explain. According to Riker (1982), Arrow's Theorem undermines the very pretense of finding always a majoritarian "will" with more than two alternatives (and remember, in politics there are always more than two alternatives available, even if all of them are not submitted to a vote).

When a Condorcet winner exists, according to a certain distribution of preference profiles among voters, and the method picks the Condorcet winner, perhaps that outcome might be considered the true utilitarian outcome. But that very outcome may come at the price of introducing or withdrawing irrelevant alternatives (with the aim of manipulating the result), or at the price of restricting the menu of alternatives that are submitted to a vote (what is called "universal domain"). Arrow demonstrated that there is no single method that could meet the three conditions altogether: universal domain, independence from irrelevant alternatives, and transitivity (no cyclic majorities). But things don't end there. It was later shown that cyclic majorities are infrequent in unidimensional spaces with few alternatives, but

abound in multidimensional and continuous policy spaces[10], and that, when majority cycles exist, they might cover –under certain conditions elaborated by McKelvey (1976, 1979)– the entire space of political alternatives. A strategic agenda setter with full information of voter preferences can reach –through successive binary elections– to any point in the policy space, provided voters are sincere and shortsighted. Intuitively, the widespread possibility of cyclical majorities in these settings implies that the outcomes produced by voting may not have the consistency we require of a "common will", along utilitarian lines. Riker (1982) concludes that we should reject the goal of finding a common or collective preference "not because it is morally wrong, but merely because it is empty". With insincere, self-interested voters, the whole adventure of finding a utilitarian social function collapses in the maelstrom of McKelvey's chaos theorem.

This instability, however, has limits. Miller (1980) and later McKelvey (1986) showed that, if we drop sincerity and assume sophisticated voters (that is, voters that behave strategically and often vote insincerely for an alternative they don't prefer as the best, or against an alternative that they prefer in first place) with complete information on the alternatives, the set of sophisticated equilibrium outcomes –through voting by successive elimination– must lie inside a central bunch of alternatives, the uncovered set. Thus, the uncovered set puts some bounds on the manipulability of the agenda, while introduces a relative limit to the general instability of possible outcomes in multidimensional spaces. With sophisticated voters with complete information, voting cycles may emerge but the uncertainty is confined in a central area identified by the uncovered set. To put it clear: outcomes guaranteed within the uncovered set do not imply the strategy-proofness of the method, but manipulability can only replace one uncovered alternative with other uncovered alternative. In this sense, manipulability with the method proposed here is limited in such a way that an outcome obtained by manipulation cannot be worse (inasmuch it is not covered) that other alternatives.

The uncovered set might be the political equilibrium to which sophisticated voters concur with full information of other's preference profiles,

---

[10] Empirical studies have shown that cyclical preferences are infrequent, but they have been less clear about why. Single-peaked preferences, as defined by Black (1958), which were supposed to be displayed along single policy dimensions, appear to be constructed around strong feelings for or against candidates (Niemi and Wright 1987). That is, a relatively high proportion of the preference orders are consistent with single-peakedness, but often do not occur along partisan or left/right policy lines.

so while it might be unstable inside it, it could be regarded as the set of alternatives that loosely define –in an indecisive way– the social choice outcome that maximizes satisfaction of self-interested preferences under the utilitarian model. It's worth exploring whether, departing from the epistemic model, in accordance with Rousseau, it is possible or not to always guarantee the selection of an alternative that, under certain preference profiles, are placed inside the uncovered set. Intuitively, there is no reason to think that any epistemic method, however designed, can assure the choice of an uncovered alternative. For the meantime, it is enough to observe that epistemic methods could display divergent outcomes from the ones that could be predicted under strategic behavior by self-interested voters, an intuition that only would confirm that the rationality of the strategy is essentially different than the rationality of *judgment about the common good*.

In any case, the important point is that when an epistemic voter is expressing an opinion based on a judgment about what justice, or the common good, requires, on the overall utilitarian model, however, an individual's vote only expresses self-interested preferences. The utilitarian approach attempts to sum votes as utilitarianism sums satisfactions, but with strategic voters and more than three alternatives, the predicted outcome of that method is unstable and indecisive, however confined to the uncovered set. Instead, the epistemic model of voting tries to estimate the most probable correct alternative or ranking, assuming certain conditions (of which equal competence higher than 0.5 and sincerity are common) and while they are challenged from several angles, from the standpoint of judgmental rationality and stability, their potential outcomes go in predictable different directions than the utilitarian model, even if we put in brackets the internal quality (self-interested or judgmental) of preferences.

6. Conclusion

We can conclude that while it is difficult to conciliate epistemic and self-interested strategic voting, it is possible to find a common method to model at least some important aspects of both views. We have argued that a simple rule as that proposed here for calculating the maximum proportional majoritarian support satisfies the following features: 1) It preserves the epistemic credentials of the Condorcet's rule: the Condorcet winner, if it exists, is always the alternative with maximum proportional majoritarian support. Insofar as Condorcet's rule is an embodiment of Rousseau's theory of voting, we can claim that our method, at least in that aspect, also respects Rousseau's general will; 2) It provides a model for Rous-

seau's idea of the minority deference. Only majoritarian votes are taken into account when computing each alternative's coefficient; 3) It overcomes the Condorcetian guidelines to solve cycles. Young's maximum likelihood estimation and, equivalently, Kemeny's rule, also do that, but unlike those methods, ours ensures –we conjecture– that the winning alternative(s) belong to the uncovered set; 4) The previous mentioned feature also limits, though does not prevent, strategic voting. The uncovered set puts some bounds on the monopoly power of an agenda setter in the context of amendment agendas, while introduces a limit to the general instability of possible outcomes in multidimensional spaces.

The method we have explored is not proposed as a panacea for the epistemic vote, but simply as a possibility to satisfy the desired characteristics mentioned above. We also gave a justification for why those characteristics are desirable. Still, a more formal treatment is needed to develop the model and get a proof of our conjecture, what is left to further work.[11]

REFERENCES

Arrow, K. (1951), *Social Choice and Individual Values* (New York: John Wiley).

Balinski, M. and Laraki, R. (2010), *Majority Judgment: Measuring, Ranking, and Electing* (Cambridge: The MIT Press).

Barry, B. (1967), "The Public Interest", in A. M. Quinton (1967) (ed.), *Political Philosophy*, Oxford Readings in Philosophy (Oxford: Oxford University Press, 112-126).

Black, D. (1958), *The Theory of Committees and Elections* (Cambridge: Cambridge University Press).

Brandt, F. (2010), "Tournament Solutions and their Applications to Multiagent Decision Making", in J. Dix and C. Witteveen (2010) (eds.), *Multiagent System Technologies. MATES 2010*. Lecture Notes in Computer Science, vol. 6251 (Berlin: Springer Heidelberg, 1-1). DOI: 10.1007/978-3-642-16178-0_1

Cohen, J. (1986), "An Epistemic Conception of Democracy", *Ethics*, 9(1): 26-38.

Condorcet, J.-A.-N. de Caritat M. de (1785,) *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* (Paris: L'Imprimerie Royale)

Copeland, A. H. (1951), "A 'Reasonable' Social Welfare Function". Seminar on Applications of Mathematics to Social Sciences. Mimeographed Notes. (Ann Arbor: University of Michigan).

[11] We thank an anonymous reviewer whose comments helped improve the clarity of the article. This paper was supported by ANPCyT (PICT 2017-1702) and partially supported by Universidad Nacional del Sur (PGI 24/I265), Argentina.

Elster, J. (2012), "The Optimal Design of a Constituent Assembly", in H. Landemore and J. Elster (2012) (eds.), *Collective Wisdom: Principles and Mechanisms* (Cambridge: Cambridge University Press, 148-172). DOI: 10.1017/cbo9780511846427.008

Estlund, D. (2008), *Democratic Authority: A Philosophical Framework* (Princeton: Princeton University Press)

Goodin, R. and Spiekermann, K. (2018), *An Epistemic Theory of Democracy* (Oxford: Oxford University Press).

Grofman, B. and Feld, S. (1988), "Rousseau's General Will: A Condorcetian Perspective", *American Political Science Review*, 82: 567-576.

Harrison, R. (1983), *Bentham* (London: Routledge & Kegan Paul).

Kemeny, J. (1959), "Mathematics without Numbers", *Daedalus*, 88: 577-591.

May, K. O. (1952) "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decisions", *Econometrica*, 20(4): 680-684.

McKelvey, R. D. (1976), "Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control", *Journal of Economic Theory*, 12: 472- 482.

McKelvey, R. D. (1979), "General Conditions for Global Intransitivities in Formal Voting Models", *Econometrica*, 47: 1085-1112.

McKelvey, R. D. (1986), "Covering, Dominance and Institution-Free Properties of Social Choice", *American Journal of Political Science*, 30: 283-314.

Miller, N. (1980), "A New Solution Set for Tournament and Majority Voting: Further Graph-Theoretical Approaches to the Theory of Voting", *American Journal of Political Science,* 24: 68-96.

Miller, N. (1983), "The Covering Relation in Tournaments: Two Corrections", *American Journal of Political Science*, 27: 382-385.

Niemi, R. G. and Wright, J. R. (1987), "Voting cycles and the structure of individual preferences", *Social Choice and Welfare,* 4, 173-183. DOI: 10.1007/BF00433943

Nino, C. S. (1991), *The Ethics of Human Rights* (Cambridge: Cambridge University Press).

Pettit, P. (2016), "Rousseau's Dilemma", in A. Lifschitz (2016) (ed.), *Engaging with Rousseau: Reception and Interpretation from the Eighteenth Century to the Present* (Cambridge: Cambridge University Press, 168-88).

Rawls, J. (1950), *A Study in the Grounds of Ethical Knowledge: considered with Reference to Judgments on the Moral Worth of Character*, PhD Thesis (Princeton: Princeton University).

Riker, W. (1982), *Liberalism against Populism*: *A Confrontation between the Theory of Democracy and the Theory of Social Choice* (San Francisco: W. H. Freeman).

Rousseau, J. J. (1962), *The Social Contract* (London: Penguin Books).

Saari, D. and Merlin, V. R. (2000), "A Geometric Examination of Kemeny's Rule", *Social Choice and Welfare*, 17: 403-438.

Schwartzberg, M. (2008), "Voting the General Will: Rousseau on Decision Rules", *Political Theory*, 36(3): 403-423. DOI: 10.1177/0090591708315145

Schwartzberg, M. (2015), "Epistemic Democracy and its Challenges", *Annual Review of Political Science*, 18: 187-203.

Tuck, R. (2019), "Active and Passive Citizens. Lecture I: Rousseau and Sieyes". *The Tanner Lecture on Human Values* (Princeton: Princeton University).

Waldron, J. (1990) "Rights and Majorities: Rousseau Revisited". *Nomos*, 32 (Special Issue: Majorities and Minorities): 44-75.

Young, H. P. (1986), "Optimal Ranking and Choice from Pairwise Comparisons", in B. Grofman and G. Owen (1986) (eds.), *Information Pooling and Group Decision Making* (Greenwich: JAI Press, 113-122).

Young, H. P. (1988), "Condorcet's Theory of Voting", *American Political Science Review*, 82(4): 1231-1244.

Young, H. P. (1995), "Optimal Voting Rules", *Journal of Economic Perspectives*, 9(1): 51-64.

Young, H. P. and Levenglick, A. (1978), "A Consistent Extension of Condorcet's Election Principle", *SIAM Journal on Applied Mathematics*, 35: 285-300.

Recibido: 19-01-2021; aceptado: 20-04-2021