On Social Choice Theory

Jan van Eijck

An economist is visiting the project. The weather continues to be excellent, permitting the project members and their guest to eat their meals outside on the NIAS terrace again. This time the computer scientist has brought a laptop.

Economist: Your project on Games, Action and Social Software is intriguing, and this is certainly a splendid environment for carrying it out. But I wonder if what you guys intend to develop doesn't already exist. The field that is called *Social Choice Theory*, isn't that what Social Software is all about?

Philosopher: Oh, you mean the branch of welfare economics that was founded following the celebrated impossibility results that Kenneth Arrow proved in his thesis (later published in [1])? That work is certainly very relevant to us. You could say that Arrow, like Condorcet, is a founding father of our field of study.

Computer Scientist: Can anyone give me a brief sketch of what Arrow's result is about?

Logician: OK, I will give it a try. Suppose you have three voters 1, 2, 3 and three states x, y, z that represent the things they want to achieve in casting their votes. The states could represent preferences for which game to play: hide-and-seek, kick-the-can or I-spy. Or they could represent the range of choice of candidates for leader of the nation. It does not really matter. Suppose each voter has a ranking of the states. We do not allow ties, so there are six possible rankings:

$$\begin{array}{lll} x < y < z & x < z < y & y < x < z \\ y < z < x & z < x < y & z < y < x. \end{array}$$

Philosopher: I see that you list all the orderings of the set $\{x, y, z\}$ that are

linear and transitive. Linearity presumably reflects the condition that a voter has to make up her mind about how she values the outcomes. Fair enough. Transitivity imposes a kind of consistency requirement: if I prefer x over yand y over z then it is only natural that I prefer x over y. Why are ties not allowed?

Economist: Formally it does not matter much, but it makes sense to rule them out in the preferences. The preferences are established by voting. It is natural to assume that a valid vote expresses a definite preference one way or the other. The challenge is to combine the wishes of the voters in a single outcome, and in this outcome, ties are allowed. So in the outcome there are thirteen possible rankings: six rankings where the preferences are all different as before, six rankings where the voter is indifferent between two of the three options:

 $\begin{array}{lll} x,y < z & z < x,y & x < y,z \\ y,z < x & y < x,z & x,z < y; \end{array}$

and finally the don't care case x, y, z.

Logician: That is right. Now processing the votes boils down to mapping the preference orderings of the voters to an outcome. In our example case, there are thirteen possible outcomes. Arrow calls such an outcome a social ordering.

Computer Scientist: Arrow sets out to study social welfare functions by imposing reasonable conditions on them, isn't that right?

Logician: Yes, indeed. There are four conditions. In the first place, there is the condition of universal or unrestricted domain, call it U. What it says is that every possible set of individual voter preferences should be in the domain of the social welfare function.

Computer Scientist: Let us see. Taking the example case, there are six linear orderings and three voters, which means there are 6^3 sets of preference orderings for the three voters. That is $6 \times 36 = 216$ preference orderings. According to condition U, all of these should be in the domain of the welfare function. But this means that even in this simple example case the number of possible welfare functions is truly enormous: 13^{216} . (Consults his laptop.) This is larger than 10^{240} , so it is a number with more than 240 digits.

Logician: The second condition is what Arrow calls independence of irrelevant

alternatives. If x and y are social choices, and if voters are allowed to change their preferences about other choices than x and y, then this change should have no effect on what the social welfare function says about the relation between x and y. Let us call this condition I.

Computer Scientist: I suppose that this is a severe restriction.

Logician: The third condition is the so-called Pareto principle, call it P. If x and y are possible choices, and all voters prefer x over y, then the social welfare function should prefer x over y.

Computer Scientist: Why is this called the Pareto principle?

Economist: Because it has to do with a method of optimization proposed by the economist Vilfredo Pareto. According to Pareto, if a situation can be changed so as to make one individual better off without making anybody else worse off, then the change is an improvement. A situation is Pareto optimal if no such improvement is possible.

Computer Scientist: Clearly, if everyone prefers x over y then outcomes that rank x above y are Pareto optimal, with respect to x and y at least.

Logician: The final principle says that there should be no dictator. Call this ND, for 'not D'. There should be no voter such that for every set of orderings in the domain of the social welfare function and every pair of distinct social states x and y, if that particular voter strictly prefers x over y, then the social welfare function ranks x above y.

Computer Scientist: I suppose a social welfare function would be dictatorial if it is a projection function, a function that projects the preference vector to a particular component of the vector. So *ND* rules out that the social welfare function is a projection function?

Logician: Yes, that's another way of putting it.

Philosopher: Sounds reasonable enough, all of it.

Economist: Yes, one would think so. But here is the snag. Arrow's theorem states that no such social welfare function exists. In other words the four principles U, I, P and ND, taken together, are inconsistent.

Logician: Put otherwise, principles U, I, P together imply D. So in our example case, imposing U, I and P cuts down the number of possible social welfare functions from 13^{216} to just three: projection of the first input component,

projection of the second input component, and projection of the third input component.

Philosopher: I object to the word *dictator* for a voter who happens to have preferences that agree with the social welfare function.

Economist: You are missing the point. Let me try to explain this in a different way. A social welfare function would be democratic (in social choice theory this is called *anonymous*) if it assigns every individual vote the same weight. In other words, in the case of three voters with preference orderings L_1 , L_2 and L_3 , the value $F(L_1, L_2, L_3)$ should be identical to $F(L_2, L_1, L_3)$, which again should be identical to $F(L_2, L_3, L_1)$, and so on. Now the point is that not only is a social welfare function F satisfying U, I and P not democratic, but it is much worse than that ...

Philosopher: I see. I take it, then, that the only way to get around these results is by relaxing some conditions. Suppose we allow the input preferences to be weak orderings, with ties allowed?

Computer Scientist: This would give an initial domain of 13^3 possibilities (consults his laptop again) which gives 2197 possible inputs, and 13^{2197} possible welfare functions. Wow, that is a number with more than 2400 digits.

Economist: Yes, but Arrow's result still holds for this case.

Philosopher: How about relaxing the conditions on the input preferences still further? Many American voters may in retrospect prefer both Gore and Kerry to Bush, without feeling any need whatsoever to compare Gore to Kerry.

Economist: Well, a way to think about Arrow's theorem is that there exist situations where a conflict among the assumptions occurs. Note that the theorem does not assert that the assumptions always are in conflict. For instance, the plurality vote is included in his theorem, but there are many profiles where everything is perfectly fine.

Philosopher: What do you mean by a profile?

Economist: A vector giving the individual preferences over a set of options for a set of voters, the mathematically explicit version of "what the voters want". So if we require the conditions U, I and P to always hold for all possible profiles, then we need a dictator.

Philosopher: So when the preferences are partially ordered, are there demo-

cratic social welfare functions satisfying U, I and P?

Logician: Since partial orderings include linear orders, there exist settings where a conflict arises among his assumptions, so Arrow's result still applies.

Economist: One of the other cornerstones of social choice theory is a famous theorem of Gibbard and Satterthwaite [2; 6].

Philosopher: Isn't this a theorem about manipulability?

Economist: Well, I guess you could call it that. It has to do with the nonexistence of certain social choice functions. A social choice function is like a social welfare function, except that the outputs are social states. Recall that social states represent anything voters may want to achieve. So if x, y, z are social states, a function that picks one of these is a social choice. A social choice function is *strategy-proof* if no voter can improve the social choice by voting against his true preferences.

Logician: Suppose the social choice for a preference vector $L_1 \ldots L_N$ is x, and i changes his preference from L_i to L_i' . If the social choice for preference vector $L_1 \ldots L_i' \ldots L_N$ is y (different from x), then y should be ranked above x in L_i' .

Philosopher: So if a preference change for i has as a result that the social choice changes, then the change should reflect the new preference of i. But then it holds by symmetry that x should be ranked above y in L_i , isn't that right?

Logician: Yes, right indeed.

Economist: A choice function is dictatorial if there is a voter i such that it holds for every input vector $L_1 \ldots L_N$ that the social choice is x if and only if x is at the top of i's preference ranking L_i . What the theorem of Gibbard and Satterthwaite says is that if there are at least three social goods, then any social choice function that is strategy-proof and has the property that for each social good there should be a voting profile that results in the choice of that good, then the function is dictatorial.

Philosopher: In other words, if the function is strategy-proof and onto (or: surjective), then it is dictatorial.

Economist: That's what the theorem implies, indeed.

Logician: In a paper I have just read there is a claim that a single proof yields

both results [3]. In other words, the logical underpinnings of Arrow's theorem and the theorem of Gibbard-Satterthwaite are identical.

Economist: It is well known that there are close connections between the two theorems. They are pointed out in a textbook by Alan Taylor [7]. As a matter of fact, I discussed the matter once over a glass of wine with Don Saari, who filled me in on historical details. Gibbard and Satterthwaite proved the theorem at essentially the same time. But, Satterthwaite was a graduate student — and this result was part of his University of Wisconsin thesis — so there was a delay in his publishing it. By the time he submitted his paper, Gibbard's paper was in the works, which meant that Mark's paper was not publishable. Hugo Sonnenschein, however, suggested to Mark that he show the connection of his result to Arrow's result, so he did. As such, the real person to show that the logical underpinnings of Arrow's result and the Gibbard-Satterthwaite result are the same is Mark Satterthwaite.

Philosopher: (*To the logician*) Do you still remember the structure of the proof you just mentioned? Do you think you can present it to us?

Logician: If I am allowed to use pencil and paper, yes. As a matter of fact, I reread the paper yesterday, in preparation for our discussion. I will give you a proof of the fact that any Pareto efficient and monotonic social choice function is dictatorial. From this the Gilbert-Satterthwaite result easily follows.

Computer Scientist: But first you have to explain to us what it means for a social choice function to be Pareto efficient and monotonic. Pareto efficiency, I can guess: a social choice function f is Pareto efficient if whenever social good x is at the top of every voter's preference list, then f yields value x.

Logician: That's right. Monotonicity is also straightforward. If social choice function f yields choice x for preference vector $L_1 \ldots L_N$, then the choice does not change if we adjust the preferences of all the voters, provided in each new preference L_i' no social good y that was ranked below x in L_i is promoted to rank above x.

Philosopher: So only lowering the position of x in the voter preferences might effect a change from x to a different choice. This is the Gibbard-Satterthwaite counterpart of *independence of irrelevant alternatives*, I suppose.

Logician: Indeed, it is. Now here is the theorem: if there are at least three social goods, and f is a social choice function that is Pareto efficient and monotonic, then f is dictatorial.

Philosopher: Fine. Let's go for the proof.

Logician: Suppose x, y are distinct social goods. Assume a voter profile with x top of the list and y bottom of the list in every voter's ranking. What should the outcome of f be?

Philosopher: Well, x, of course. This follows from the fact that f is Pareto efficient.

Logician: That's right. Remember that y was bottom of the list for every voter. Now suppose that I take the preference list of the first voter, and start moving y upward on the list. What will happen?

Computer Scientist: As long as y stays below x, nothing I suppose.

Logician: And if I move y above x?

Computer Scientist: Either nothing, or the value changes to y. This follows by monotonicity of f, doesn't it?

Logician: Correct. Now suppose I am going through the voter list, and for each voter move y from the bottom position to the top position. What will happen?

Philosopher: Then for some voter i, at the point where y gets raised past x, the choice will change from x to y. For suppose it does not. Then we end up with a preference list where y is above x in every voter's preference, while the choice is still x. This contradicts Pareto efficiency.

Logician: That's right. So we get the following two pictures. Let's call these Figure 1 and Figure 2. (Draws two pictures for them to look at.)

L_1		L_{i-1}	L_i	L_{i+1}	•••	L_N		
y	• • •	y	x	x	•••	x		
x	• • •	x	y	•	•••	•		
•	• • •	•	•	•	•••	•	\mapsto	x
•	• • •	•	•	•	•••	•		
•	• • •	•	•	•	• • •	•		
•	• • •	•	•	y	• • •	y		

This first picture shows the situation just before the value flips from x to y.

 L_1 . . . L_{i-1} L_i L_{i+1} . . . L_N xy. . . \boldsymbol{y} y. . . xx. . . xx. \mapsto y. • . \boldsymbol{u} . . . y

This second picture shows the situation just after the value has flipped from x to y.

Philosopher: Fair enough. And now I suppose further on in the proof the patterns in these pictures get manipulated a bit more?

Logician: That's exactly right. Let us study what would happen if in the first picture and the second picture we were to move x down to the bottom for all voters below i, and move x down to the second last position for all voters above i.

Computer Scientist: Nothing, I suppose.

Logician: That's right, the situations would be as pictured in the following figures. Let us call these Figures 3 and 4.

L_1	• • •	L_{i-1}	L_i	L_{i+1}	• • •	L_N		
y	• • •	y	x	•	•••	•		
•	• • •	•	y	•	•••	•		
·	• • •	•	·	•	• • •	•	\mapsto	x
·	• • •	•	·	•	•••	•		
·	• • •	•	·	x	•••	x		
x	• • •	x	·	y	•••	y		

This is Figure 3. It is the result of taking Figure 1 and moving x down to the bottom for voters below i and moving x to the second last position for voters above i.

L_1	•••	L_{i-1}	L_i	L_{i+1}	•••	L_N		
y	•••	y	y	•	•••	•		
•	•••	•	x	•	•••	•		
•	•••	•	•	•	•••	•	\mapsto	y
•	•••	•	•	•	•••	•		
•	•••	•	•	x	•••	x		
x	•••	x	•	y	• • •	y		

This is Figure 4. It is the result of making similar changes to Figure 2.

Philosopher: For Figure 4, I can see why the value does not change. In Figure 2 the value was y, and it must remain y in Figure 4 by monotonicity.

Computer Scientist: OK, so Figure 4 has value y. But the Figures 3 and 4 differ only in the order of x, y in the ranking of i. It follows by monotonicity that the value in Figure 3 must be either y or x.

Philosopher: And it cannot be y, because then by monotonicity the value in Figure 1 would have to be y as well, and it is not. So the value in Figure 3 has to be x.

Logician: Just as I told you. Now suppose I take Figure 3 and move y down to the one but last position for all voters below i. This would not change the choice from x to a different value, would it?

Philosopher: I suppose it would not, by monotonicity again. The relative position of y with respect to x does not change.

Logician: So we get the following picture:

L_1	• • •	L_{i-1}	L_i	L_{i+1}	• • •	L_N		
·	• • •	•	x	•	• • •	•		
·	• • •	•	y	•	• • •	•		
	• • •	•	•	•	• • •	•	\mapsto	x
	• • •	•	•	•	• • •	•		
y	• • •	y	•	x	•••	x		
x	• • •	x	•	y	• • •	y		

Call this Figure 5. Now consider a social good z different from x and y. By moving z through the preference orderings without letting z move past x we can obtain the following situation without changing the value of the choice function:

L_1	•••	L_{i-1}	L_i	L_{i+1}	•••	L_N		
·	•••	•	x	•	• • •	•		
•	•••	•	z	•		•		
•	•••	•	y	•		•	\mapsto	x
z	•••	z	•	z		z		
y	• • •	y	•	x	• • •	x		
x	•••	x	•	y	•••	y		

Call this Figure 6.

Philosopher: I suppose monotonicity ensures that the value of the function does not change by the transition from 5 to 6?

Logician: That's correct. Now swap the rankings of x and y for all voters above i. By monotonicity, the choice value for the result must be either x or y.

Computer Scientist: But it cannot be y. For suppose it is, and consider the effect of moving z to the top in every preference. Since this would nowhere effect a swap with y, the value would have to remain y, by monotonicity. But then a profile with everywhere z on top would have value y, which contradicts Pareto efficiency.

Logician: So the value has to remain x, and we get the following picture:

L_1	• • •	L_{i-1}	L_i	L_{i+1}		L_N		
•	• • •	•	x	•	• • •	•		
•	• • •	•	z	•	• • •	•		
•	• • •	•	y	•	• • •	•	\mapsto	x
z	• • •	z	•	z	• • •	z		
y	• • •	y	•	y	• • •	y		
x		x	•	x		x		

Now we are done, for observe that monotonicity ensures that making changes in the preferences of i while making sure that x remains on top will have no effect on the outcome. This means that the social choice will be x whenever x is at the top of i's ranking.

Philosopher: So i is a dictator for social good x. But since x was arbitrary, there must also be a dictator j for social good z as well. Clearly if i dictates whether x is on top, and j whether z is on top, then, to paraphrase Henk

Wesseling, i and j have to be the same guy. Hence i must be a dictator for all alternatives.

Logician: Why do you quote Henk Wesseling?

Computer Scientist: (To the economist) Henk Wesseling is an honorary NIAS fellow. You met him yesterday at dinner.

Philosopher: In a column in a Dutch newspaper Wesseling once commented on the lack of historical knowledge of modern students. His juiciest example was the following anecdote. After an undergraduate history seminar, a student came up to him with bright eyes. "Professor, now I suddenly got it. This Jesus and this Christ that they are all talking about, that must be the same fellow."

Economist: Yes, I was introduced to Wesseling during yesterday's NIAS Banquet Dinner in Leiden. "May I introduce you to the teacher of Alexander"? I didn't get the joke, and nobody explained it to me.

Philosopher: Because it was no joke. Wesseling is professor emeritus of History from Leiden University, and he was the master's thesis supervisor of the Crown Prince of the Netherlands, Willem Alexander, or Alexander for short.

Logician: Not the same guy as the student from the seminar, I should hope.

Computer Scientist: I suppose it is shown in [3] that the proof of Arrow's theorem follows exactly the same pattern? And deriving the theorem of Gibbard and Satterthwaite from the above is just a matter of showing that any function that is strategy-proof and surjective has to be Pareto efficient and monotonic?

Logician: Right on both counts.

Economist: There is still an issue of how to interpret Arrow's results. Don Saari has written eloquently on that in two books that appeared in 2001 [4; 5]. Arrow's theorem hinges on the fact that the principle I of independence of irrelevant alternatives, or the principle of binary independence, as Saari calls it, allows one to hide the rationality of the voters.

Logician: That's right. In his investigation of positional voting procedures, Saari proposed a modification of I. His proposal is to replace I by what he calls the principle of *intensity of binary independence*. Let's call it II. This principle states that also the *intensity* of a voter's preference of one alternative over another should be taken into account.

Economist: In particular, it matters not only that x is preferred over y, but also how many candidates there are between x and y.

Philosopher: Aha, I can see how that would break steps in the reasoning above. The manipulations of the preference vectors in the proof rely heavily on monotonicity. But what is a positional voting method?

Economist: Positional voting methods are methods that score candidates by allotting numbers of points to them to reflect their position in the preference ordering of a voter. The paradigm of this is the so-called Borda count. This was proposed in 1770 by Jean-Charles de Borda. Suppose there are n candidates. Then the Borda count assigns n - j points to a voter's j-th ranking candidate.

Philosopher: So in the case of three candidates, my first choice gets 2 points, my second candidate 1 point, and my least preferred candidate 0 points?

Economist: That's right. In the case of three candidates ordered x < y < z, the Borda count has the form x : 2, y : 1, z : 0. When this was proposed as voting method for the *Académie Française*, of which Borda was a member, another member, the mathematician Laplace, proposed to compare this to other ways of assigning points to a candidate depending on position in the preference ordering.

Logician: The mind of a true mathematician at work.

Economist: Such methods are positional methods. Plurality voting, where each voter votes for one candidate assigns points (1, 0, ..., 0). Antiplurality voting, where each voter object to one candidate assigns points (1, ..., 1, 0). The Borda count assigns points (n - 1, n - 2, ..., 1, 0). Saari has a theorem stating that the only positional social welfare function satisfying U, P and II is the Borda count. All other positional methods fail.

Philosopher: It looks like Saari turns his analysis of Arrow's result into a plea for adopting voting procedures for rational voters that reflect transitivity of preferences.

Economist: That's right. The Borda count voting procedure does so. All kinds of pairwise comparison procedures are dangerous, is what he claims.

Philosopher: But wait. I seem to remember that Condorcet, also a French

Academy member, made his proposal for pairwise run-off voting procedures precisely because he did not agree with the Borda method. We have talked about Condorcet before (see page ??). Condorcet objected to positional methods generally because they do not always select the candidate that would be victorious in a pairwise voting contest against *any* of the other candidates.

Economist: Surely, the Borda count does not always pick what has come to be known as the Condorcet winner. But the point Saari is trying to make is that this may not be as bad as Condorcet thought it was. Saari analyzes Condorcet's original example of a selection procedure with three candidates.

Philosopher: You have me intrigued. Why don't you look it up?

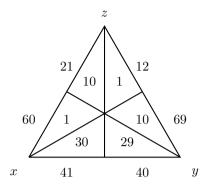
Economist: (Leafs through Saari's 'Chaotic Elections') Right, here it is. In the case of three candidates x, y, z there are six voting profiles: x < y < z, x < z < y, y < x < z, y < z < x, z < x < y, z < y < x. Call these profiles 1 through 6. Condorcet's example was as follows: (Draws a table on a sheet of paper)

1 30 x < y < z2x < z < y1 3 y < x < z294 y < z < x105z < x < y106 z < y < x1

Condorcet reasoned that a positional voting scheme will elect y: 39 voters put y in first place, 30 put y in second place, 31 voters put x in first place and 39 put x in second place, and 11 voters put z in first place, and 11 put z in second place. So any reasonable positional voting scheme would yield profile 3, the profile with y < x < z, as outcome of the voting procedure. According to Condorcet, this is counterintuitive: x would have beaten y with 41 to 40 votes, and x would have beaten z with 60 to 21 votes.

Philosopher: Wait, wait, not so quick. No-one will be able to work out these numbers on the fly.

Economist: Well, after reading Saari's books you will be. Saari presents beautiful geometric representations. Let me draw the one for the Condorcet example. (*Draws on the paper*)



Philosopher: Let me figure this out ... The vertices in the equilateral triangle represent the three candidates x, y, z, right? Presumably closeness to a vertex indicates preference. Then each region in the triangle corresponds to a profile. Yes, that's right. The region with 30 written in it corresponds to profile x < y < z. And the numbers to the side of the triangle indicate the results of pairwise run-offs. Now I can see how you can say so quickly that x beats z with 60 votes to 21.

Economist: You are quick. (With a smile) I will never underestimate a philosopher again. Saari, by the way, draws a completely different conclusion from the example than Condorcet did. He argues in favor of y as winner, as follows. He is looking for profiles that cancel out. For instance, let me ask you the following question. Is it reasonable to assume that opposite profiles cancel out, in the sense that if one voter with preference x < y < z and one voter with preference z < y < x stay home, this should not affect the voting result? Or put otherwise, can we tally ballots by counting the votes of these two voters as a tie?

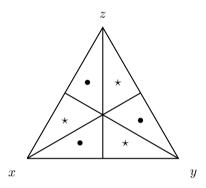
Philosopher: Are you asking me? Well, I think it should make a difference. After all, the two voters agree that y is not so bad, and that information gets lost if they don't vote.

Computer Scientist: Yes, I agree with that. But what if three voters, one with profile x < y < z, one with profile y < z < x, and one with profile z < x < y, all stay home?

Philosopher: Then I suppose that should make no difference to the outcome, for the three profiles together create a cycle, and no preferential information

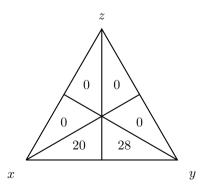
can possibly be extracted from that. Yes, we should be able to combine these three profiles to form a tie.

Economist: That's exactly right. What this means is that we can proceed by counting ties first, and then see what remains. So the above picture can be simplified. For there are two of those preferential cycles. First there is the one you just mentioned. Let's mark it with \bullet . Then there is the one that runs in the opposite direction: z < y < x, y < x < z, and x < z < y. Let's mark this with \star :



Philosopher: And now you are going to simplify the picture by subtracting the largest possible fixed numbers from regions with the same mark?

Economist: That's right. Here is the result of counting all triples of voters whose profiles cancel out as ties (*Draws a new picture*):



Philosopher: Wow, a clear win for candidate y.

Logician: All this theorem proving and analyzing voting profiles makes one crave a refreshment. The NIAS restaurant boasts an excellent espresso machine. Shall we go inside for some coffee or cappuccino?

Economist: Good idea.

References

- K. Arrow. Social Choice and Individual Values. Wiley, New York, 1951, second edition: 1963.
- [2] A. Gibbard. Manipulation of voting schemes: A general result. *Econo*metrica, 41:587–601, 1973.
- [3] P.J. Reny. Arrow's theorem and the Gibbard-Satterthwaite theorem: A unified approach. *Economics Letters*, 70(1):99–105, 2001.
- [4] Donald G. Saari. *Chaotic Elections*. American Mathematical Society, 2001.
- [5] Donald G. Saari. Decisions and Elections: Explaining the Unexpected. Cambridge University Press, 2001.
- [6] M.A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.
- [7] Alan D. Taylor. Social Choice and the Mathematics of Manipulation. Mathematical Association of America and Cambridge University Press, 2005.