Synthese (2011) 179:35–56
Knowledge, Rationality & Action 35–56
DOI 10.1007/s11229-010-9854-8

Measuring voting power for dependent voters through causal models

Luc Bovens · **Claus Beisbart**

Received: 11 November 2010 / Accepted: 12 November 2010 / Published online: 30 November 2010

Abstract We construct a new measure of voting power that yields reasonable measurements even if the individual votes are not cast independently. Our measure hinges on probabilities of counterfactuals, such as the probability that the outcome of a collective decision would have been yes, had a voter voted yes rather than no as she did in the real world. The probabilities of such counterfactuals are calculated on the basis of causal information, following the approach by Balke and Pearl. Opinion leaders whose votes have causal influence on other voters' votes can have significantly more voting power under our measure. But the new measure of voting power is also sensitive to the voting rule. We show that our measure can be regarded as an average treatment effect, we provide examples in which it yields intuitively plausible results and we prove that it reduces to Banzhaf voting power in the limiting case of independent and equiprobable votes.

Keywords Voting power · Banzhaf measure · Counterfactuals · Causal models · Average treatment effect

1 Introduction

Recent years have seen an intense interest in the measurement of *voting power*. Here, voting power is the *extent to which a political agent can affect the outcome of a*

L. Bovens
Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK
e-mail: L.Bovens@lse.ac.uk

C. Beisbart (✉)
Institute for Philosophy and Political Science, TU Dortmund, 44221 Dortmund, Germany
e-mail: Claus.Beisbart@udo.edu

collective decision by casting her vote (Felsenthal and Machover 1998, p. 36). A very popular measure of voting power is the *Banzhaf measure*. It equals the chance of pivotality in a binary decision under the assumptions that (i) the voters cast their votes independently and (ii) each voter is equally likely to vote yes or no. These probabilistic assumptions define the so-called Bernoulli model (Felsenthal and Machover 1998, Def. 3.1.1, p. 37).

However, one might argue, if sufficient empirical data on past voting patterns are available, then one should base the measurement of voting power on a probability model that fits the data, and of course the Bernoulli model will seldom be a fitting model (see Gelman et al. 2002, 2004 for a related discussion). A *simple suggestion* to generalize the Banzhaf measure for such cases would be to calculate the probability that a voter is pivotal under the appropriate probability model—whatever it is like. This suggestion has been put forward by Morriss (1987, p. 169). But unfortunately this suggestion is problematic if the independence assumption does not hold, since it leads to counterintuitive assessments of the voting powers of the voters. The following example is a simplification of what has come to be known as Wilmers' example in Machover (2007). Five voters take decisions under simple majority vote. The 5–0 split and the 0–5 split have a probability of one half each. The other voting profiles have probability zero. Clearly, the probability of being pivotal is zero for each voter. Accordingly, under the simple suggestion, no voter would have voting power in the example. But, intuitively, it seems odd to say that no voter can ever influence the outcome of the collective decision (ibid., p. 3). A different measure of voting power is needed.[1]

How can we construct such a new measure? In the example above, the obvious question to ask is: How did the probability distribution arise? Suppose, first, that voter

[1] One may want to deal with Wilmers' example by generalizing the Banzhaf measure in a different way. Under the Bernoulli model, Banzhaf voting power is also a measure of success—it is a linear and monotonically increasing function of the probability of success (see Theorem 3.2.16 on p. 45 in Felsenthal and Machover 1998). Here a voter is successful if and only if her vote coincides with the outcome of the collective decision (i.e. if and only if she gets what she wants in this sense). One may therefore suggest to generalize the Banzhaf measure by calculating that linear function of the probability of success under the most realistic probability model. Under this suggestion, we would obtain non-zero values for voting power in Wilmers' example, and the puzzle has seemingly disappeared. The suggestion does not work though. Success and power in the sense of decisiveness are simply different notions (Laruelle and Valenciano 2005). Under general conditions, the linear transform of the probability of success does not provide a measure of decisiveness. Now in this study we started from the idea that somebody can make a difference to what the outcome of a collective decision is. This is power in the sense of decisiveness and not success. We are therefore not interested in success in this paper. It is even arguable that success is simply not a variety of power at all. An example from Machover (2007, p. 4) makes this plain. If a dummy (for a definition see Felsenthal and Machover 1998, p. 24) always votes the way a dictator votes, the dummy is very successful—her vote always reflects the outcome. But intuitively the dummy does not have any power. The reason seems to be that power is a modal notion—it concerns what we can do—whereas success is not a modal notion.

A slightly different suggestion regarding Wilmers' example is that our intuitions about the example are misguided, since decisiveness and success are confused: We tend to find it strange that no voter has power, because we think of power in terms of success. But if we were to focus on power as decisiveness, we would not find it strange that no voter has zero power whatsoever. We disagree once more. Even if we focus on the power to make a difference to the outcome of a collective decision, the result is puzzling. Is it really never the case that the outcome would have been different, had a particular voter cast a different vote? This is not necessarily so, e.g. if there is an opinion leader.

A is an opinion leader, i.e. if A votes yes (or no), then all other voters simply copy her vote. What one would say intuitively is that A has more voting power than the other voters—if A had voted differently, then the others would have followed suit and the outcome would have been different. But we cannot say the same for the other voters. However, if, second, a different voter B is an opinion leader, clearly B should have more voting power than the other voters. Or, third, if the voting pattern comes about only due to political views shared by the voters, then we expect equal voting power in the example. Thus, in order to provide an assessment of voting power, one should take into account the *causal relations* between the votes that bring about the distribution over profiles.

We will go beyond the simple suggestion to generalize the Banzhaf measure. We will develop a measure of voting power that does not clash with our intuitions because it is sensitive to causal information. The measure relies on causal models and probabilities for counterfactual queries. We define the measure in Sect. 2. It is also shown that the measure can be regarded as an average treatment effect well-known in causal analysis. In Sect. 3, we develop a simple example to show how the measure deals with opinion leaders, shared political views and dictatorial voting procedures. In Sect. 4, we show that the measure reduces to Banzhaf voting power under the Bernoulli model. Section 5 provides an analysis of Wilmers' example (Machover 2007). Section 6 concludes, and technical details are provided in the Appendix.

Since we work with causal models and probabilities, one may distinguish between causal dependence and stochastic dependence. If not stated differently, "(in)dependence" means stochastic (in)dependence. In principle, two random variables may be stochastically independent, although one having a particular value may have a causal bearing on the value of the other. This is so if, e.g., the positive causal bearing from A on D via B is balanced out by the negative causal bearing from A on D via C (cf. lack of faithfulness, Spirtes et al. 2000, p. 13; see also Perl 2000, p. 48). Models that display this feature will not be considered in our paper though.

The relation of our work to the existing literature is as follows. Felsenthal and Machover (1998, pp. 35–36) draw a distinction between I- and P-power. I-power is the extent of the influence that a political agent can have on the outcome of a collective decision. P-power is the share that a voter receives of a fixed prize, where the prize is thought to belong to those voters whose vote reflects the final outcome. We are clearly interested in I-power in this work. In the terminology of Laruelle and Valenciano (2005), we are interested in power in the sense of decisiveness rather than of success. Now Felsenthal and Machover (1998, p. 36) suggest, that the I-power of a voter should be measured by the probability that her vote is pivotal. They further claim that the Bernoulli model is the most natural model to work with, if one is interested in I-power (cf. ibid., pp. 37–8 for a justification of this choice). This leads to the Banzhaf measure of voting power. Dubey and Shapley (1979, pp. 122–123) consider probability models in which the equiprobability assumption is dropped. As a straightforward generalization of the Banzhaf measure, they consider the probability of being pivotal under such models. Very recently, there has been interest in giving up the independence assumption as well. Beisbart and Bovens (2008) quantify the voting power of U.S. citizens in presidential elections by calculating the probability of being pivotal under a model with dependent votes. Kaniovski and Leech (2009) propose a slightly different measure of voting power. For any probability model over the votes, they calculate a weighted sum of the conditional probabilities of (i) being pivotal given that one has voted yes and (ii) being pivotal given that one has voted no. As Machover (2007) argues, measuring power as the probability of being pivotal under models with dependent votes leads to counterintuitive assignments of voting power, as is evidenced in Wilmers' example. The approach of Kaniovski and Leech runs into the same problem. The measure that we propose, on the contrary, avoids the counterintuitive assignments of voting power. The problem is also avoided by Beisbart (2010) who offers a hierarchy of measures in order to quantify overall voting power. This suggestion and the proposal in our paper are completely unrelated though. For a general framework to study voting power see Laruelle and Valenciano (2005).

2 Generalized voting power

The voting power of a voter i is the extent to which she is able to make a difference as to what the result of a collective decision is, whether a proposal is accepted or rejected, say (Felsenthal and Machover 1998, p. 36). How can that extent be measured? One way to measure it is to calculate the probability that voter's vote is pivotal (cf. ibid.). But this proposal runs into the problem that is made plain by Wilmers' example. So an alternative is required.

Suppose that with person i voting no on a proposal in the actual world, the chance of acceptance is quite low. But if i were to vote yes instead, then the chance of acceptance would be much greater. Similarly, suppose, with i voting yes on a proposal in the actual world, the chance of rejection is quite low. But if i were to vote no instead, then the chance of rejection would be much greater. Under these assumptions, intuitively, i has more voting power than a person j for whom these respective effects would be notably smaller. The idea is thus to take differences of chances in order to measure voting power. Let us make this idea precise by constructing a measure.

For each voter i, let V_i be a random variable that equals one if i votes yes, and zero otherwise. Let V be a random variable that equals one if the outcome is yes, and zero otherwise. We assume that we have a full probability model for the votes. The model provides us with probabilities for each possible voting profile (i.e. the joint probabilities of the V_i). We assume for now that all conditional probabilities such as $P(V_j = 0|V_i = 1)$ are defined. (In Sect. 5, we will consider a case in which this assumption is violated.)

Collective decisions are taken by amalgamating the individual votes following a voting rule. The voting rule maps voting profiles—vectors of the type $(V_1, V_2, \ldots, V_n)$—into $\{0, 1\}$. Once we know the voting rule, we can calculate conditional probabilities for acceptance given some single vote or given a combination of votes, such as $P(V = 1|V_i = 1)$ or $P(V = 1|V_i = 1, V_j = 0)$. It is not our concern in this paper how to obtain a realistic probability model from empirical data.

We will also assume that we have causal information about the votes. In our models a person's vote can be influenced by other people's votes and by their political views (for discussion see Sect. 6). This is not inconsistent with taking voting to be an instance of free agency. How the related causal information can be obtained will not be discussed in this paper.[2]

For calculating our measure of voting power, we first assess the chance that a proposal is accepted given that i voted no, i.e. $P(V = 1|V_i = 0)$ and the chance that a proposal is rejected given that i voted yes, i.e. $P(V = 0|V_i = 1)$. Subsequently, we construct probability distributions over certain counterfactuals. We ask what the chance is that a proposal would have been accepted, had i voted yes (rather than no, as i did in the actual world), call it $Q_i^0(V = 1|V_i = 1)$. And we ask what the chance is that a proposal would have been rejected, had i voted no (rather than yes, as i did in the actual world), i.e. $Q_i^1(V = 0|V_i = 0)$.

These chances are calculated following the approach by Balke and Pearl (1994). The basic idea is very simple. For calculating $Q_i^0(V = 1|V_i = 1)$, e.g., we first assume that i votes no, as she does in the actual world. We infer the probabilities of the other votes being one way or another. Then i's vote is switched to yes. We trace the causal effects of i's voting yes and recalculate the probabilities that the votes that are causally affected by i's vote are one way or another.[3] Finally, the probability of acceptance is calculated on this basis. A general algorithm for calculating the probabilities for counterfactuals is given in the Appendix.

Let D_i^0 be the difference between the chance that the proposal would have been accepted had i voted yes and the chance that the proposal is accepted conditional on i having voted no:

$$D_i^0 = Q_i^0 (V = 1|V_i = 1) - P (V = 1|V_i = 0) . \tag{1}$$

Let D_i^1 be the difference between the chance that the proposal would have been rejected had i voted no and the chance that the proposal is rejected conditional on i having voted yes:

$$D_i^1 = Q_i^1 (V = 0|V_i = 0) - P (V = 0|V_i = 1) . \tag{2}$$

Because $Q_i^1(V = 0|V_i = 0) = 1 - Q_i^1(V = 1|V_i = 0)$ and $P(V = 0|V_i = 1) = 1 - P(V = 1|V_i = 1)$, D_i^1 can also be written as follows:

$$D_i^1 = P(V = 1|V_i = 1) - Q_i^1(V = 1|V_i = 0). \tag{3}$$

[2] Note that there are two varieties of influence that are important for our measure. First, there is the influence a voter can have on the outcome of a collective decision. The extent of that influence is the voting power of that voter and this is what we want to measure. Second, there is the influence that voters can have on the votes of the other voters. That influence enters our measure. Both times, we are using the same notion of influence, but the influences are *on different things* (on the outcome vs. on the other votes).

[3] Following Lewis (1986), we take it to be the case that truth-value assignments to counterfactuals under a default interpretation do not permit backtracking.

We can now construct the measure:

$$D_i = D_i^0 \times P(V_i = 0) + D_i^1 \times P(V_i = 1). \tag{4}$$

It can be shown that D_i is always in the interval $[-1, 1]$.

In order to motivate our definition of the D-measure further, we show that it can be thought of as an *average (or mean) treatment effect* as known from causal analysis (see King et al. 1994, pp. 76–82 and Morgan and Winship 1999, pp. 661–666 for the following). To see this, consider an analogy first. Suppose that there is a particular treatment for a particular disease. We want to quantify the causal impact of that treatment on survival in the subpopulation of people with that disease. Let N be the size of this subpopulation. As a matter of fact, some people chose to take the treatment, whereas others did not. Set $T_\alpha = 1$, if person α chose to take the treatment and 0 otherwise. Likewise, set $S_\alpha = 1$, if person α survives and 0 otherwise.

For measuring the causal impact of the treatment, it is no use comparing the survival rates between people who chose the treatment and people who did not, since choosing the treatment may be connected with social strata membership, which, in turn, may influence the chances of survival. Instead, for each person, we have to compare the outcome—the value of S_α—in the actual world with the outcome that would have occurred had person α chosen differently. For a mathematical description, we assign each individual α two numbers, $S_\alpha^{T=1}$ and $S_\alpha^{T=0}$ that describe the *potential* outcomes under treatment (top index '$T = 1$') and non-treatment (top index '$T = 0$', see Morgan and Winship 1999, p. 662). Since we are talking about potential outcomes, $S_\alpha^{T=1}$ is even defined in case $T_\alpha = 0$ and vice versa. $S_\alpha^{T=1}$ and $S_\alpha^{T=0}$ each are 1 in case of survival, and 0 otherwise. $S_\alpha^{T=0}$ is the "unobservable counterfactual outcome" if α chose the treatment; $S_\alpha^{T=1}$ is the "unobservable counterfactual outcome" if α did not choose the treatment (ibid.).

The treatment effect for an individual is the difference δ_α between $S_\alpha^{T=1}$ and $S_\alpha^{T=0}$. The average treatment effect is then defined as the average over the δ_α, taken in the subpopulation. The average treatment effect can be written as follows:

$$\overline{\delta} = \frac{1}{N}\left(\sum_{\alpha \text{ with } T_\alpha=0}\left(S_\alpha^{T=1} - S_\alpha^{T=0}\right) + \sum_{\alpha \text{ with } T_\alpha=1}\left(S_\alpha^{T=1} - S_\alpha^{T=0}\right)\right). \tag{5}$$

Since $S_\alpha^{T=1}$ in the first addend is unobserved and $S_\alpha^{T=0}$ in the second addend is unobserved, care is needed in estimating the average treatment effect from data. For literature on this problem see Morgan and Winship (1999) and Winship and Morgan (2007).

But we are not interested here in how to estimate the average treatment effect from data. Rather, we will assume a causal model from which the average treatment effect can be calculated. For this, it is useful to write $\overline{\delta}$ in slightly different terms. Let $|A|$ the cardinality of some set A. Since $S_\alpha^{T=0/1}$ can only take the values 0 or 1, it follows that $\overline{\delta}$ equals

$$\left(\frac{\left|\left\{\alpha|T_\alpha = 0 \wedge S_\alpha^{T=1} = 1\right\}\right|}{|\{\alpha|T_\alpha = 0\}|} - \frac{\left|\left\{\alpha|T_\alpha = 0 \wedge S_\alpha^{T=0} = 1\right\}\right|}{|\{\alpha|T_\alpha = 0\}|}\right)\frac{|\{\alpha|T_\alpha = 0\}|}{N}$$
$$+\left(\frac{\left|\left\{\alpha|T_\alpha = 1 \wedge S_\alpha^{T=1} = 1\right\}\right|}{|\{\alpha|T_\alpha = 1\}|} - \frac{\left|\left\{\alpha|T_\alpha = 1 \wedge S_\alpha^{T=0} = 1\right\}\right|}{|\{\alpha|T_\alpha = 1\}|}\right)\frac{|\{\alpha|T_\alpha = 1\}|}{N} \tag{6}$$

This is a weighted sum of two differences. Suppose now that the fractions that figure in this expression approach probabilities, as the size of the sample under scrutiny goes to infinity.[4] In this limit, the weights in the formula approach the probability that a person chose no treatment, call that $P(T = 0)$, and the probability that a person chose treatment, $P(T = 1)$, respectively. The second term in the first difference approaches the conditional probability that a person survives given that she chose no treatment, $P(S = 1|T = 0)$. Likewise, the first term in the second difference approaches the conditional probability of survival given treatment, $P(S = 1|T = 1)$. The other terms in the differences concern counterfactual outcomes. The first term in the first difference approaches the probability of the counterfactual that a person would have survived had she chosen treatment rather than choosing no treatment as she did in the real world, call that $Q^0(S = 1|T = 1)$. In the same way, the second term in the second difference approaches $Q^1(S = 1|T = 0)$. Altogether the average treatment affect approaches

$$\begin{aligned}\overline{\delta} &= \Big(Q^0(S = 1|T = 1) - P(S = 1|T = 0)\Big) \times P(T = 0)\\ &\quad + \Big(P(S = 1|T = 1) - Q^1(S = 1|T = 0)\Big) \times P(T = 1).\end{aligned} \tag{7}$$

A little algebra yields

$$\begin{aligned}\overline{\delta} &= \Big(Q^0(S = 1|T = 1) - P(S = 1|T = 0)\Big) \times P(T = 0)\\ &\quad + \Big(Q^1(S = 0|T = 0) - P(S = 0|T = 1)\Big) \times P(T = 1).\end{aligned} \tag{8}$$

Let us now carry over the analogy to votes. An analogy is established, if we let the treatment correspond to i voting yes, and survival correspond to a collective yes vote. In terms of our mathematical notation, S corresponds to V, and T corresponds to V_i. Consider now the average treatment effect that i's voting yes has on the outcome of a collective decision. In order to measure that average treatment effect, we only need replace S by V, and T by V_i in Eq. 8. If we do this replacement, we arrive exactly at our D-measure. Our *D-measure thus equals the average treatment effect of i's vote on the outcome of the collective decision*. It has a clear probabilistic meaning.

Equation 7 was arrived at by looking at frequencies and taking an appropriate limit. Likewise, our D-measure may be arrived at by looking at frequencies. This time,

[4] According to the law of large numbers, the probability that a series of independent trials of a random experiment reproduces the probability characteristic of that random experiment converges to one, as the number of trials goes to infinity.

however, the right frequencies to look at do not concern a population—every voter is supposed to have her own D-measure. Rather, we have to look at a series of collective votes and to consider appropriate frequencies for each voter within this series. In any case, the frequency story is not essential for the definition of average treatment effect. The frequency story is a nice motivation for defining the average treatment effect. But once we have arrived at Eq. 7, we may think of the probabilities in different ways. In particular, they may be single-case probabilities. In fact, King et al. (1994, pp. 76–82) define the average treatment effect (in their terms the mean causal effect) for a particular individual i without considering actual frequencies at all.

Our finding that the D-measure equals an average treatment effect may seem to give rise to an objection against the D-measure. The D-measure is supposed to be a measure of voting power—it is supposed to quantify the extent to which a voter *can* influence the outcome of a collective decision. But the average treatment effect is about the influence that the treatment *does* indeed have.

The objection can be rebutted. In voting theory, a voter actually having influence on the outcome of a collective decision is equivalent to her being able to have that influence. The equivalence holds in the following way: A voter *does* make a difference to the outcome of a collective decision in some particular case if and only if she *can* do so. The left to right direction of this equivalence is trivial. The right to left direction goes as follows: Suppose you *can* make a difference to what the outcome of a collective decision is. Then, if you vote yes, the outcome will be one way, and if you vote no, the outcome will be the other way. Now whatever your vote will be, you *will* make a difference in this case. The consequence of the equivalence is that, in being sensitive to the influence that a voter has, our D-measure is at the same time sensitive to the influence that a voter can have.

The D-measure is different from the Banzhaf measure in a crucial respect though. The reason is that two kinds of influence may be distinguished. A voter has positive influence in case had she voted a instead of b, the outcome of the collective decision would have been a instead of b. Here, a and b are variables that can take the values "yes" and "no". It is assumed that, if a equals "yes", then b equals "no" and the other way round. A voter has negative influence in case, had she voted a instead of b, the outcome of the collective decision would have been b instead of a. Negative influence is unexpected; commonly, if a voter switches from "no" to "yes", it is not to be expected that the outcome switches from "yes" to "no". But it is possible that a voter's voting yes induces some other voters to vote no. Note that, whenever a voter *can* have positive influence, she *does* have positive influence, and the same with negative influence.

Now the assumptions that underlie the Banzhaf measure exclude negative influence. The question how negative influence should be counted, simply does not arise for the Banzhaf measure. But the question does arise for our D-measure. As it were, for calculating the D-measure, positive influence is counted positively, and negative influence is counted negatively. As a consequence, the D-measure may be negative. An example later in this paper will make this plain. Also, the D-measure may average to zero not because a voter cannot influence the outcome of the votes at all, but rather because positive and negative influences that she can have can cancel each other out. All this cannot happen under the Banzhaf measure.

One might want to object that the D-measure is not an appropriate measure of voting power, because it is sensitive to the direction of influence. This very feature, the argument goes, is not compatible with the standard definition of voting power. Now voting power is supposed to measure the extent to which a vote can make a difference to the outcome of a collective vote. But "the extent to which i can make a difference to y" is itself ambiguous. Suppose that there is a pot of money. If I regularly add 10 units to the pot of money then clearly, the value of the measure of the difference an average transaction makes to the pot equals 10. But what if I regularly take out 10 units? If we are not sensitive to the direction of the difference, then we set the value of the measure at 10. If we are sensitive to the direction of the difference, then we set the value of the measure at minus 10. What if I add 10 units on Mondays and take out 6 units on Tuesdays? If we are not sensitive to the direction of the difference, then we set the value of the measure of the difference an average transaction makes to the pot at $(10 + 6)/2$. If we are sensitive of the direction of the difference, then we set the value of the measure at $(10 - 6)/2$. Both measures would be reasonable measures of the difference of an average transaction to the pot of money. In this paper we chose to measure voting power by being sensitive to the direction of the difference that one's vote makes to the collective outcome. One could choose to define a measure that is not sensitive to this direction. The D-measure could then be readily adapted to fit this interpretation.

We need not necessarily motivate the definition of the D-measure by showing it to be an average treatment effect. There is also an ex post justification for quantifying voting power in terms of D: D yields intuitively plausible results if applied to a range of examples. We now turn to such examples.

3 Simple examples

We start with a simple three-person example featuring a Supreme Court with Scalia, Thomas and Ginsburg as justices.[5] This example will illustrate how the D-values are calculated. We will first assume that proposals are decided by simple majority voting. Rather than using the terminology that is fitting for the Supreme Court, we will conduct our presentation in terms of *voters* and *proposals*.

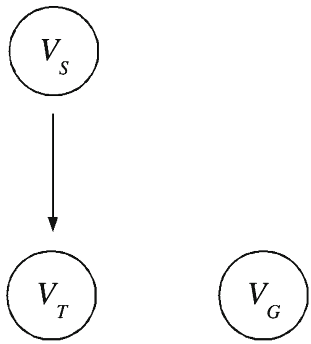
We consider different models for the votes. Under every model, each voter votes yes with a probability of .5.

3.1 Opinion leader

Ginsburg's vote is stochastically independent from the other votes and there are no causal relations between her and the others' votes. Thomas, however, keeps a close eye on Scalia and there is .9 chance that he will vote yes, given that Scalia votes yes, and there is a .9 chance that he will vote no, given that Scalia votes no. There is thus not only a stochastic dependency between the votes; Scalia's vote has also a causal

[5] These names are just mnemonic aids. For actual correlation coefficients between the justices' votes in the Supreme Courts, see Kaniovski and Leech (2009, Table 3).

Fig. 1 The causal network for the opinion leader model



bearing on Thomas's vote. The causal influences in the model are represented in Fig. 1. Let us now assess the voting power of each voter by calculating her D-value.

We start with the D-value of Thomas. We consider the first addend, viz. $D_T^0 \times P(V_T = 0)$. Clearly, $P(V_T = 0) = .5$. For calculating D_T^0 we have to assume that Thomas votes no in the real world. We first turn to $P(V = 1|V_T = 0)$. If Thomas votes no, then the conditional chance that Scalia voted no is .9. Given that Thomas votes no, we only get acceptance in the real world, if both Scalia and Ginsburg vote yes. And so the conditional chance that the proposal is accepted is $P(V = 1|V_T = 0) = .1 \times .5 = .05$. We now turn to $Q_T^0(V = 1|V_T = 1) = 1 - Q_T^0(V = 0|V_T = 1)$. If Thomas were to vote for the proposal, this would not affect the chance that Scalia voted no—that chance is still .9—since the causal link does not go from Thomas to Scalia. The only profile under which the proposal would be rejected, if Thomas voted yes, is a profile with Scalia and Ginsburg voting no. That chance is $.9 \times .5 = .45$. So $Q_T^0(V = 1|V_T = 1) = 1 - .45 = .55$. Hence, $D_T^0 = .55 - .05 = .50$. The argument for D_T^1 runs parallel and so $D_T = .50$.

Let us now calculate the D-value for Scalia. We first consider Q_S^0. If Scalia votes no, then the chance that Thomas votes no is .9. Given that Scalia votes no, we only get acceptance in the real world, if both Thomas and Ginsburg vote yes. And so the conditional chance that the proposal is accepted is $P(V = 1|V_S = 0) = .1 \times .5 = .05$. We now turn to $Q_S^0(V = 1|V_S = 1) = 1 - Q_S^0(V = 0|V_S = 1)$. If Scalia were to vote yes, this would affect the chance that Thomas votes no—that chance is now .1—since the causal link goes from Scalia to Thomas. The only profile under which the proposal would be rejected is a profile with Thomas and Ginsburg voting no. The chance is $.1 \times .5 = .05$. So $Q_S^0(V = 1|V_T = 1) = .95$ and $D_S^0 = .9$. The argument for D_S^1 runs parallel and so $D_S = .9$.

Finally, we assess the voting power of Ginsburg. If Ginsburg votes no, then this does not affect the chance that Thomas or Scalia vote no. Given that Ginsburg votes no, the chance that both Thomas and Scalia vote yes is $.9 \times .5 = .45$. So $P(V = 1|V_G = 0) = .45$. Suppose that Ginsburg asks herself what the chance would be that the motion had been accepted had she voted yes. The chance that both Thomas and Scalia would have voted no is $.9 \times .5 = .45$. So $Q_G^0(V = 1|V_G = 1) = .55$. Hence, $D_G^0 = .1$. The argument for D_G^1 runs parallel and so $D_G = .1$.

This is not unreasonable. Scalia does have more voting power than Thomas, because he takes Thomas along with him as he changes votes, but not vice versa. Ginsburg has

Table 1 Results under the opinion leader model

Voter	Scalia	Thomas	Ginsburg
D-measure	.9	.5	.1

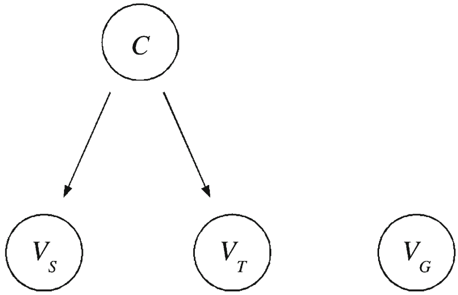


Fig. 2 The causal network for the common cause model

very little voting power because she faces a quasi-block vote of the two other voters. Our results for the *D*-measure are summarized in Table 1.

3.2 Common causes

Let us now change our assumptions in the following way:

1. We introduce a parameter ϵ which permits us to vary the extent to which the votes of Scalia and Thomas are correlated (correlations always imply stochastic dependency). The parameter ranges from –1 for full negative correlations over 0 for independence to +1 for full positive correlation:

$$P(V_T = 1|V_S = 1) = P(V_T = 0|V_S = 0) = .5 \times (1 + \epsilon). \tag{9}$$

   Hence,

$$P(V_T = 1|V_S = 0) = P(V_T = 1|V_S = 0) = .5 \times (1 - \epsilon). \tag{10}$$

2. Correlations do not arise due to a direct causal influence from Scalia to Thomas or vice versa as in Sect. 3.1. Rather, they are due to a common cause (see Fig. 2). Positive correlations are due to shared political views. Negative correlations are due to diverging political views. We can model this in the following way. We introduce a random variable which captures the nature of the proposal (cf. Balke and Pearl 1994). If $C = 0$, then the proposal is such that both Scalia and Thomas vote no; if $C = 1$, then the proposal is such that Scalia votes no and Thomas votes yes; if $C = 2$, then the proposal is such that Scalia votes yes and Thomas votes no; if $C = 3$, then the proposal is such that both vote yes. In the terms of Dretske (1988, pp. 42–44), C models a triggering cause for the voting behaviour of Thomas and Scalia. The features of the proposal trigger votes that match the political views of Thomas and Scalia. By specifying the probability values in Table 2 we can fix the degree to which Thomas and Scalia's votes are correlated or anti-correlated.

Table 2 The probability model for the common cause model

$C = k$	$P(V_S = 1 \mid C = k)$	$P(V_T = 1 \mid C = k)$	$P(C = k)$
$C = 0$	0	0	$.25(1 + \epsilon)$
$C = 1$	0	1	$.25(1 - \epsilon)$
$C = 2$	1	0	$.25(1 - \epsilon)$
$C = 3$	1	1	$.25(1 + \epsilon)$

The variable V_G is not included—it is independent from the other variables V_i and takes values of 0 and 1 with a probability of .5 each

Table 3 Results under the common cause model

Voter	Scalia	Thomas	Ginsburg
D-measure	.5	.5	$.5 \times (1 - \epsilon)$

In the following, we will omit the details of our calculations. A general algorithm for calculating the Q_i^0 and Q_i^1 is provided in Appendix. Applying this algorithm, we obtain the results shown in Table 3. Thus, under the common cause model, the D-values for Scalia and Thomas do not depend on the strength of the correlations. On our model, the voting power of a voter is the same regardless whether she shares political views with another voter or not. However, the voting power of Ginsburg depends on whether and how the votes of the other voters are correlated. If Scalia and Thomas always vote the same, then Ginsburg has no voting power. If, on the other hand, Scalia and Thomas always cast opposing votes, then Ginsburg has maximal voting power. Clearly, if the votes of Scalia and Thomas are independent, then every voter has a D-value of .5. Note that this D-value for independent votes coincides with the Banzhaf measure for this simple voting game. As we will show in Sect. 4, this is due to a more general connection between Banzhaf voting power and the D-value.

One might object that, intuitively, the voting power of Scalia (and Thomas) is greater in a court in which their votes are positively correlated than in a court in which they vote independently. But the objection confuses different notions of power. If Scalia's and Thomas's votes are positively correlated, these voters have more power in the sense that they are more likely to be successful—their votes are more likely to coincide with the outcome of the collective decision. But that does not mean that they have more power in the sense of being more decisive. It is the latter notion of power that we are interested in. The objection may also be based on a confusion of absolute power with relative power. Absolute power is the power that we have discussed in this paper thus far. Relative power is the power a voter has compared to the average power of all voters. It can be quantified by normalizing a measure of absolute power by the sum of those measures of all voters. In our example, Scalia and Thomas do have more relative power in the case of positive correlations than they have in case their votes are stochastically independent. In sum, the results that we obtain for the example are plausible for voting power as an absolute measure of decisiveness.

3.3 Causal influence on multiple votes

Let us now change our assumptions once more. Start with the opinion leader model from Sect. 3.1 and suppose that Scalia's causal influence extends to Ginsburg's vote as

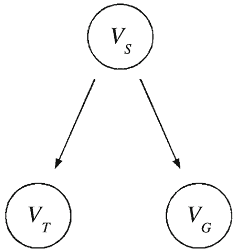


Fig. 3 The causal network with causal influence on multiple votes

Table 4 Results for a model in which Scalia's vote causally influences both other votes

Voter	Scalia	Thomas	Ginsburg
D-measure	$1 - .5 \times (1 - \epsilon)^2$	$.5 \times (1 - \epsilon)$	$.5 \times (1 - \epsilon)$

well. This leads to a causal model as depicted in Fig. 3. We define a family of probability models parameterized by ϵ. As in Sect. 3.2, ϵ can take values in the interval $[-1, 1]$. For $\epsilon = 1$, there are full positive correlations between Thomas's (or Ginsburg's) and Scalia's votes. In this case, Scalia is a very influential opinion leader and his vote will be copied by any other voter in the court. For $\epsilon = -1$, there are full negative correlations between Thomas's (or Ginsburg's) and Scalia's votes. For instance, both Thomas and Ginsburg might dislike Scalia and try to outvote him on every issue.

In mathematical terms, the model looks as follows: There is still equiprobability for yes and no votes in the marginal probabilities of the different voters. Given Scalia's votes, Thomas's and Ginsburg's votes are independent. Furthermore,

$$P(V_T = 1|V_S = 1) = P(V_G = 1|V_S = 1) = .5 \times (1 + \epsilon) \tag{11}$$

and

$$P(V_T = 0|V_S = 0) = P(V_G = 0|V_S = 0) = .5 \times (1 + \epsilon). \tag{12}$$

This fixes the probability model.

Results can be calculated following the method from the Appendix. They are shown in Table 4. For $\epsilon = 0$, the votes of the voters are independent and each voter's D-measure is .5. For full positive correlations ($\epsilon = 1$), Scalia has a D-measure of 1, whereas the other have a zero D-measure. This is a plausible thing to say, since, in this case, Thomas and Ginsburg lack the ability to affect the outcome of the collective decision by unilaterally casting a different vote.

For full negative correlations ($\epsilon = -1$), we obtain: $D_S = -1$, $D_T = 1$, $D_G = 1$. Let us first discuss the D-measure of Scalia for this case—this D-measure is negative. As we have said above, negative values of the D-measure mean that a voter is able to affect the outcome of a collective decision, but in an unexpected way. If she were to switch her vote from yes to no (from no to yes), this would make a difference, but it is not the probability of rejection (acceptance) that is increased, as one might

Table 5 Results for a voting rule under which Scalia is a dictator

Voter	Scalia	Thomas	Ginsburg
D-measure	1	0	0

expect, but rather the probability of acceptance (rejection).[6] In terms of the terminology introduced in Sect. 2, the voter tends to have negative influence rather than positive influence.

Of course, this unexpected way of influencing the outcome of a collective decision invites strategic voting. If a person were to know that she has negative voting power, then she could vote strategically and vote against her actual preferences. Suppose that her strategy is successful. One might then want to say that she has positive voting power—she is able to get the outcome she wants. One may want to turn this into an objection against our D-measure, since the D-measure is maximally negative. However, this objection does not work. Voting power simply is not about what people want. Rather, it concerns the relations between particular individual actions (casting particular votes) and the outcome of a collective decision. What somebody really wants is another matter.[7]

Let us now turn to the D-measure of Thomas and Ginsburg—which is 1. At first sight, this seems strange. The votes of Thomas and Ginsburg are fully caused in the sense that, given each particular vote of Scalia, both Thomas and Ginsburg vote the same way Scalia did. But on second reflection D being 1 for Thomas and Ginsburg turns out to be okay. The D-measure for Thomas focuses on the question what would happen, if Thomas switched his vote. For answering this question, the causes behind Thomas's actual vote simply do not matter. Rather, a world is imagined in which Thomas casts a different vote from that he cast in the real world. We conclude that our measure deals with the example appropriately.

3.4 Dictator

So far we have assumed simple majority voting with equal weights as the voting rule. Let us now change the weights as follows: Scalia has a block vote of weight three, whereas Thomas and Ginsburg have only votes of weight one, each. The Supreme Court issues a yes vote if and only if the weights of the yes votes add up to at least three. Thus, Scalia is a dictator, whereas Thomas and Ginsburg are dummies (cf. Def. 2.3.4 on p. 24 in Felsenthal and Machover 1998).

Whether we calculate these results under the model under which Scalia is an opinion leader (Sect. 3.1) or under the common cause model with any value of ϵ (Sect. 3.2) or under the model under which Scalia has influence on multiple votes with any value of ϵ (Sect. 3.3), we obtain the results listed in Table 5.

[6] Since the D-measure is the sum of two addends, it is sufficient for a voter's D-measure to be negative that both possible switches change the probability of acceptance in an unexpected way, and it is necessary that one of the switches changes this probability in an unexpected way.

[7] Cf. the notion of "epistemic power" in Morriss (1987, Chap. 8).

These results are very plausible. Only the dictator has voting power, whereas the dummies do not have voting power. This gives rise to the following observation. The D-value significantly depends on the voting rule. As we change the voting rule and keep the model for the voting profile fixed, the value of D will in general change. This is important, since, in voting theory, we are particularly interested in how different voting rules affect the influences that the voters can have on the outcome.

4 The relation to the Banzhaf measure

We will now show that the D-measure reduces to Banzhaf voting power under the condition of equiprobability and if there is causal and stochastic independence between the votes.[8] On grounds of equiprobability:

$$P(V_i = 0) = P(V_i = 1) = .5. \tag{13}$$

To understand the role of causal and stochastic independence, consider first causally and stochastically dependent voters, e.g. Thomas with Scalia as opinion leader. Suppose that Thomas votes no. This teaches us something about Scalia's vote in the real world. On our interpretation of counterfactuals, this knowledge has to be taken into account in calculating $Q_T^0(V = 1|V_T = 1)$. Thus, $Q_T^0(V = 1|V_T = 1)$ does not simply equal $P(V = 1|V_T = 1)$. Next consider Ginsburg under the same model. The fact that Ginsburg votes no in the real world teaches us nothing about other votes, because Ginsburg's vote is independent. So $Q_G^0(V = 1|V_G = 1)$ *does* equal $P(V = 1|V_G = 1)$. More generally, if there is causal and stochastic independence between all votes, we have

$$Q_i^0(V = 1|V_i = 1) = P(V = 1|V_i = 1) \tag{14}$$

and

$$Q_i^1(V = 0|V_i = 0) = P(V = 0|V_i = 0). \tag{15}$$

So by the definition of D_i and Eqs. 13, 14 and 15,

$$\begin{aligned} D_i = .5\,(&P\,(V = 1|V_i = 1) - P\,(V = 1|V_i = 0) \\ &+ P\,(V = 0|V_i = 0) - P\,(V = 0|V_i = 1))\,. \end{aligned} \tag{16}$$

We replace $P(V = 1|V_i = 0)$ by $(1 - P(V = 0|V_i = 0))$, and $P(V = 0|V_i = 1)$ by $(1 - P(V = 1|V_i = 1))$ and obtain:

$$D_i = 2\,(.5P\,(V = 1|V_i = 1) + .5P\,(V = 0|V_i = 0)) - 1. \tag{17}$$

[8] Here causal independence means that no vote exerts a causal influence on another vote.

On grounds of equiprobability (Eq. 13), we replace the factors of .5 by $P(V_i = 1)$ and $P(V_i = 0)$, respectively. Thus,

$$D_i = 2\left(P\left(V = 1 \wedge V_i = 1\right) + P\left(V = 0 \wedge V_i = 0\right)\right) - 1. \tag{18}$$

But $P(V = 1 \wedge V_i = 1) + P(V = 0 \wedge V_i = 0)$ is just the probability of i's vote coinciding with the outcome of the vote, call it ψ_i. Therefore,

$$D_i = 2\psi_i - 1. \tag{19}$$

We know that (see Felsenthal and Machover 1998, Theorem 3.2.16, p. 45)

$$\beta_i' = 2\psi_i - 1, \tag{20}$$

where β_i' denotes the Banzhaf measure of voting power for i. Hence, D_i coincides with β_i' in the special case of equiprobability and causal and stochastic independence.

5 Wilmers' example

Let us now turn to a more complex example. Suppose that we have a five person Supreme Court with voters A, B, C, D and E and with simple majority voting. The 12 profiles in which at least four voters cast the same vote each occur with probability $1/12 - \epsilon$ for $0 \leq \epsilon \leq 1/12$. The 20 remaining profiles each occur with probability $.6\epsilon$. At $\epsilon = 0$, we reproduce Wilmers' example, as specified by Machover (2007), p. 3: All 12 profiles with 4 or 5 voters casting the same vote are equiprobable and all other 20 profiles occur with probability zero. Probability models with a finite ϵ correspond to generalizations of Wilmers' example.[9]

If ϵ is set at zero, the probability of pivotality is zero for every voter. This shows that we cannot measure voting power by means of the probability of pivotality, because, as Machover (2007, p. 3) writes, "it would be absurd to claim that every voter here is powerless, in the sense of having no influence over the outcome of divisions." So let us examine whether our D-measure yields more fitting values.

For $0 < \epsilon \leq 1/12$, we specify a causal interpretation that is consistent with the probability model. We first calculate the conditional probabilities $P(V_B = 1|V_A = 1)$, $P(V_B = 1|V_A = 0), \ldots, P(V_E = 1|V_A = 0, V_B = 0, V_C = 0, V_D = 0)$. We then impose the following causal model. A is not influenced by any other voter. B's vote is influenced by and only by A's vote. C is influenced by both A and B's votes and so on. Thus, A's vote has a causal bearing on B through E's votes, B's vote has a causal bearing on C through E's votes, ... and E's vote has no causal bearing on any other vote. This is illustrated in Fig. 4.

Of course there are many other causal models that are consistent with the probability model. E.g. a permutation of A,..., E would also yield a causal model compatible with the probabilities. Furthermore, many common cause models could be provided

9 Our discussion in the introduction was based upon a simplified version of Wilmers' example.

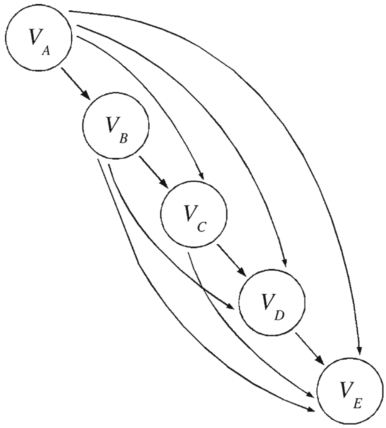


Fig. 4 The causal relations that we assume for the generalization of Wilmers' example

Table 6 Results for the extension of Wilmers' example

i	$\lim_{\epsilon\to 0} Q_i^0\ (V = 1\|V_i = 1)$	$\lim_{\epsilon\to 0} P\ (V = 0\|V_i = 1)$	$\lim_{\epsilon\to 0} D_i$
A	5/6	1/6	4/6
B	2/3	1/6	3/6
C	1/2	1/6	2/6
D	5/12	1/6	3/12
E	1/6	1/6	0

Because of the symmetry of the model, the second column also equals $\lim_{\epsilon\to 0} Q_i^1\ (V = 0|V_i = 0)$, and the third column also equals $\lim_{\epsilon\to 0} P\ (V = 1|V_i = 0)$

that are consistent with the probability model. But we will assume that our particular causal model appropriately represents the influences in the real world.

We calculate the D-values for this causal model following our methodology. As an example, we obtain $D_A = 2/3 - 9.2\epsilon$. Subsequently we calculate the limits as ϵ goes to 0 for all voters. In Table 6, we see that the D-values cascade downwards as we move from voters A to E. This squares very nicely with the fact that A is an opinion leader to more voters than B is, B is an opinion leader to more voters than C is etc.

What happens when we set ϵ at 0—as is the case of Wilmers' example (Machover 2007, p. 3)? For voters A, B, C and E, we obtain exactly the values that we have obtained in the limit $\epsilon \to 0$. But when $\epsilon = 0$ we face a problem in calculating the D-value of voter D. Suppose that $V_D = 0$ in the real world. We ask what the chance of acceptance would be, if D had voted yes rather than no, i.e. $Q_D^0(V = 1|V_D = 1)$. If we follow the algorithm from Appendix, we need the probability $P(V_E = 1|V_A = 1, V_B = 0, V_C = 0, V_D = 1)$. But this probability is undefined since $P(V_A = 1, V_B = 0, V_C = 0, V_D = 1) = 0$. Hence we cannot calculate the $Q_D^0(V = 1|V_D = 1)$ for a probability model with extreme values. For this reason, we stipulated a non-extreme ϵ-model and calculated the limiting value of the D-measure.

But one might object that there are other families of models that approach Wilmers' example in some limit. For example, we could set the probability of one profile with three yes-votes and two no-votes at $.4\epsilon$ and set the probability of another such profile at $.8\epsilon$. Again, we will recover Wilmers' example, as we set ϵ at zero. However, if we take the limit $\epsilon \to 0$, we might obtain a slightly different limit for the D-measure for D. But it can be shown that for all such families of models, the D-values for A, B, C and E are unaffected in the limit $\epsilon \to 0$ and the D-value for D ranges from 0 to 1/3, i.e. it takes the D-values of C and E as its bounds.

6 Discussion

Our assessment of a voter's voting power is based upon a causal model. But how do we know whether correlations in the votes are due to, say, opinion leaders or due to shared political views? And if they are due to opinion leaders, how do we know whether Scalia is an opinion leader for Thomas or vice versa? The standard line in Causal Learning is that a probability model yields conditional dependence structures that define a class of causal models. For this class, it may be possible to obtain bounds on the D-values of the voters. But if we want to identify a unique causal model we need additional information. One may look at the temporal structure: If Thomas always casts his vote after Scalia, then the causal direction is clear. Or one may appeal to experiment: For instance, one could toggle Scalia's vote and see whether Thomas follows suit. Once a unique causal model is specified, we can assess the D-value of each voter. But how the causal model is specified and whether this can be done in a unique way is not the subject of our inquiry.

In this paper we have only dealt with very simple causal models. It is sometimes appropriate to switch to models under which the causal pattern of votes depends on the issue under consideration. Suppose for instance that Scalia has positive causal influence on the other voters for economic issues—the other voters are prone to copy Scalia's vote (cf. the model from Sect. 3.3 with $\epsilon > 0$). But at the same time, Scalia has a negative influence on the other voters for issues of social morality—the other voters are prone to vote differently from Scalia (cf. the model from Sect. 3.3 with $\epsilon < 0$). In such a case, it is appropriate to construct two causal models and to calculate a D-measure for each type of issue that one wishes to consider.

Let us now deal with two objections against the proposed measurement of voting power.

First, one might object that voting theorists are interested in voting power, i.e. the power that a voter has *in virtue of her vote*. Here the vote figures as a particular resource that can be distinguished from other resources (see Morriss 1987, particularly pp. 18–19 and Chap. 19). Measures of voting power assess the influence that a voter can have on grounds of this resource. They thus specifically concern the voting rule. The objection then is that we are confusing this power with other kinds of power, e.g. the influence that a person can have on grounds of being an opinion leader. This objection motivates an alternative treatment of Wilmers' example: One can say that our intuitions regarding the example are confused, because no distinction is drawn between the extent to which a person can influence the outcome of the collective decision in virtue of her vote, and the extent to which she can do so in virtue of other things. In virtue of her vote, nobody has power, and this is exactly what is captured by the simple suggestion to generalize the Banzhaf measure.

In order to deal with this objection, we want first to point out an ambiguity. “In virtue of her vote” can mean (i) in virtue of her casting a vote—i.e. the real-life event of casting a vote—or (ii) in virtue of the vote that she has cast—i.e. the vote as it is registered in the voting profile. Let us then deal with the objection under both readings.

On reading (i), the objection does not get off the ground. For it is perfectly conceivable that the event of my voting yes causally bears on how other voters vote and clearly this causal influence should be incorporated into voting power, if we are to measure that extent to which a voter can influence the outcome of a collective decision in virtue of casting a vote in the first reading.

On reading (ii), we would indeed have to say that none of the voters in Wilmers' example can influence the outcome of the collective decision in virtue of her vote; and on this reading, our D-measure would take into account a kind of power that is not genuine voting power. Nonetheless, we take it that there are various reasons why the introduction of the D-measure is a valuable contribution regarding voting power. First, it is important to consider what we use measures of voting power for. Suppose that a candidate in an election uses voting power to make decisions about campaign resource allocation. (In a classic study on campaign resource allocation, Brams and Davis (1974) do in fact establish a connection between voting power and optimal resource allocation.) But then she does care about the influence that particular voters can have on the election outcome, not just through the voting rule but also through the influence that she exerts on other voters. If Scalia is an opinion leader, then it is worthwhile allocating resources to a campaign that targets Scalia. It is influence in the sense of the D-measure that matters and not just voting power on grounds of the voting rule. From the candidate's viewpoint it does not matter how a voter can make a difference to the election outcome, be it through the voting rule or through exerting influence on other voters. Second, our measure deepens the understanding of the Banzhaf measure, since we show that the Banzhaf measure is a special case of our measure. What we effectively establish is that the Banzhaf measure can be thought of as an average treatment effect under special circumstances, and this is a significant result. Third, even if one insists on restricting voting power to the power one has in virtue of one's vote only, the resistance to assigning zero voting power to the voters in Wilmers' examples shows that there is a common notion of power that is richer than voting power—call it whatever you wish. This notion takes into account both the voting rule and complex social interactions that bring about the votes, and these components are intertwined—they cannot just be isolated and added. Our D-value measures the influence that a voter can have due to a voting rule and due to her influence on others. This measure is worthwhile because it captures a natural interpretation of the influence that a voter can have on the outcome of a collective decision.

Second, one might object that it is too demanding in practice to require causal information for calculating the D-measure. Our reply is that one cannot quantify the influence that a voter can have in virtue of a voting rule and in virtue of her influence on others, without using causal information. The uncertainty in our causal information will be reflected in the uncertainty in the D-measure.

We conclude with a final comment on opinion leaders. In our example, opinion leaders can influence other people in virtue of casting their vote. But opinion leaders typically do not exercise influence through actually casting a vote, since voting is in many settings done secretly. Also, opinion leaders can only influence other voters through casting a vote, if there is sequential voting in which the opinion leaders vote before these other voters. Instead, opinion leaders tend to *express their political views in conversation* and that influences the way other people think and vote. We cannot directly use our D-measure to quantify the extent to which an opinion leader can influence the outcome of a collective decision *in virtue of expressing her views in conversation*, since the D-measure is the average treatment effect that a person's *voting* has on the outcome of a collective vote. But we can define a slight variation of the D-measure in order to quantify the extent of possible influence under consideration. For this, we introduce a random variable E_i that equals 1, if i expresses her views in conversation and 0 otherwise. For measuring the extent of the influence that an opinion leader can have in virtue of expressing her views in conversation, we consider the average treatment effect that the value of E_i has on the outcome of the vote. That is, in the definition of the D-measure, we replace V_i by E_i. In fact, in a similar way we can measure the influence that some person can have on some outcome O in virtue of doing A, whatever O and A are, quite generally.

Acknowledgments This paper was presented at LOFT8 and at an earlier time at the 5th workshop of the Voting Power and Procedures group at Warwick University (2007), supported by the Leverhulme Trust (Grant F/07 004/AJ). We are grateful to comments by Moshé Machover and two anonymous referees. CB thanks the Center for Philosophy of Science at the University of Pittsburgh and the German Academic Exchange Service (DAAD) for support.

Appendix: Algorithm for calculating the *D*-measure

We will follow Balke and Pearl (1994), though our notation diverges. Let us take $Q_i^0(V = 1|V_i = 1)$ as an example. This is the probability of the counterfactual that the proposal would have been accepted, had i voted yes, though i votes no in the actual world. i's vote is an event that is embedded in a causal structure. It is caused by certain events and it causes certain effects. Isolate all the non-effects of i's vote. If we learn that i votes no in the actual world, then this teaches us something about some of these non-effects of i's vote. We determine a joint probability model for the non-effects of i's vote, conditional on i voting no. Subsequently we set i's vote at yes, as if this came about, in Lewis's terms, by a miracle, that is, as if some exogenous force interfered in the course of nature and changed the event from voting no to voting yes (Lewis 1986). We evaluate how the effects of i's vote would be affected by the probability model over the non-effects of i's vote conjoint with i voting yes.

Formally, for evaluating $Q_i^0(V = 1|V_i = 1)$ we consider the actual world in which $V_i = 0$. Let the random variables $C_1, \ldots, C_n$ be the non-effects of V_i. The C_j may include other votes and variables representing common causes. We calculate the joint probabilities $P(C_1 = c_1, \ldots, C_n = c_n|V_i = 0)$ where the c_j range over the possible values for C_j for each j. For each combination of the $C_1 = c_1, \ldots, C_n = c_n$, we then multiply $P(C_1 = c_1, \ldots, C_n = c_n|V_i = 0)$ with $P(V = 1|C_1 = c_1, \ldots, C_n = c_n, V_i = 1)$. That is, we ask: What is the probability of acceptance, if $V_i = 1$, but if the non-effects of V_i are as they are in the actual world. By summing the products

$$P(C_1 = c_1, \ldots, C_n = c_n | V_i = 0) \times P(V = 1 | C_1 = c_1, \ldots, C_n = c_n, V_i = 1) \tag{21}$$

for every possible combination $C_1 = c_1, \ldots, C_n = c_n$, we obtain $Q_i^0(V = 1|V_i = 1)$. To calculate $P(V = 1|C_1 = c_1, \ldots, C_n = c_n, V_i = 1)$, we average over all the effect variables of V_i , call them $E_1,\ldots, E_m$:

$$\begin{aligned} &P\,(V = 1|C_1 = c_1, \ldots, C_n = c_n, V_i = 1) \\ &\quad = \sum_{e_1} \ldots \sum_{e_m} P\,(V = 1|C_1 = c_1, \ldots, C_n = c_n, V_i = 1, E_1 = e_1, \ldots, E_m = e_m) \\ &\quad\quad \times P\,(E_1 = e_1, \ldots, E_m = e_m | C_1 = c_1, \ldots, C_n = c_n, V_i = 1)\,. \end{aligned} \tag{22}$$

In terms of Bayesian Networks, we can characterize the algorithm as follows:

a. Construct a Bayesian Network with variables (1) for the votes of each voter, (2) for the common causes and (3) for the outcome of the collective decision. Insert arrows for opinion leaders as in Fig. 1, for common cause political views as in Fig. 2, and arrows from each voter into the outcome of the collective decision. The latter arrows model the voting rule.
b. Read off the prior probabilities $P(V_i = 1)$ and $P(V_i = 0)$ from this network.
c. Set the value of the variable for voter i at no and read off the probability of acceptance, i.e. $P(V = 1|V_i = 0)$.
d. Determine the joint probability distribution over the non-effect variables of V_i conditional on $V_i = 0$.
e. Construct a node for the combination of all non-effect variables $C_1, \ldots, C_n$ and insert this joint probability distribution as a new prior.
f. Erase the nodes for the individual non-effect variables along with their incoming and outgoing arrows.
g. Insert the requisite arrows from the combined non-effect variable to the effect variables of V_i, including the node for the outcome of the collective decision, and put in the concomitant conditional probability distributions. Note that V_i is a root node in this new network.
h. Set the value of the variable for voter i at yes in this new network. Read off the probability of acceptance. This is $Q_i^0(V = 1|V_i = 1)$.
i. An analogous procedure yields $P(V = 0|V_i = 1)$ and $Q_i^1(V = 0|V_i = 0)$.

References

Balke, A., & Pearl, J. (1994). Probabilistic evaluation of counterfactual queries. *Proceedings of the Twelfth National Conference on Artificial Intelligence, AAAII-94*, 220–237.

Beisbart, C. (2010). Groups can make a difference. Voting power measures extended. *Theory and Decision, 69*, 469–488; a former version is available in the Discussion Paper Series of the CPNSS at the LSE. http://www.lse.ac.uk/collections/CPNSS/CPNSS-DPS/LSEPhilosopyPapersList.htm.

Beisbart, C., & Bovens, L. (2008). A power measure analysis of Amendment 36 in Colorado. *Public Choice, 124*, 231–246.

Brams, S. J., & Davis, M. D. (1974). The 3/2's rule in presidential campaigning. *American Political Science Review, 68*, 113–134.

Dretske, F. (1988). *Explaining behaviour*. Cambridge, MA: MIT Press.
Dubey, P., & Shapley, L. S. (1979). Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research, 4*, 99–131.
Felsenthal, D. S., & Machover, M. (1998). *The measurement of voting power: Theory and practice, problems and paradoxes*. Cheltenham: Edward Elgar.
Gelman, A., Katz, J. N., & Bafumi, J. (2004). Standard voting power indexes don't work: An empirical analysis. *British Journal of Political Science, 34*, 657–674.
Gelman, A., Katz, J. N., & Tuerlinckx, F. (2002). The mathematics and statistics of voting power. *Statistical Science, 17*, 420–435.
Kaniovski, S., & Leech, D. (2009). A behavioural power index. *Public Choice, 141,* 17–29.
King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry. Scientific inference in qualitative research*. Princeton, NJ: Princeton University Press.
Laruelle, A., & Valenciano, F. (2005). Assessing success and decisiveness in voting situations. *Social Choice and Welfare, 24*, 171–197.
Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noũs*; also in Lewis, D. (1986). *Philosophical papers* (Vol. II, pp. 32–52). New York: Oxford University Press.
Machover, M. (2007). Discussion topic: Voting power when voters' independence is not assumed. Mimeo. http://eprints.lse.ac.uk/2966/.
Morgan, S. L., & Winship, C. (1999). *Counterfactuals and causal inference*. Cambridge: Cambridge University Press.
Morriss, P. (1987). *Power. A philosophical analysis* (2nd ed., 2002). Manchester: Manchester University Press.
Perl, J. (2000). *Causality. Models, reasoning, and inference*. New York: Cambridge University Press.
Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed.). Boston, MA: MIT Press.
Winship, C., & Morgan, S. L. (2007). The estimation of causal effects from observed data. *Annual Review Sociology, 25*, 659–707.