



## A Note on Measuring Voters' Responsibility

Dan S. Felsenthal

University of Haifa, Israel

(eMail: dfelsenthal@poli.haifa.ac.il)

Moshé Machover

King's College, University of London

(eMail: moshe.machover@kcl.ac.uk)

**Abstract** We consider a singular event of the following form: in a simple voting game, a particular division of the voters resulted in a positive outcome. We propose a plausible measure that quantifies the causal contribution of any given voter to the outcome. This measure is based on a conceptual analysis due to Braham (2008), but differs from his solution to the problem of measuring causality of singular events

*Keywords* Coleman's measure of power to prevent action, Degrees of causation, Extent of individual responsibility for collective decision, Necessary element of a sufficient set (NESS), Simple voting game (SVG).

### 1. Introduction

In their recent paper, Braham and van Hees (2008b) – building on Braham's essay (2008) – propose a measure of what they call 'degree of causation' of a singular event. Suppose a given outcome has occurred as result of the joint actions of several agents. The problem they address is: how to quantify the causal contribution of each of the agents to the occurrence of the outcome.

In the first four sections of their paper, they present a detailed informal conceptual discussion of the problem, which we regard as plausible, and which we wish to accept at least provisionally, for our present purpose. We shall not rehearse this discussion here, but simply refer the reader to Braham and van Hees (2008b).

But we disagree with the way Braham and van Hees proceed to define a

formal ‘degree of causation’, and we wish to propose a different formal solution.

However, rather than dealing with the whole range of cases addressed by Braham (2008) and Braham and van Hees (2008a,b), we shall confine our formal treatment to a very special type of case, to which we shall refer briefly as the *voting scenario*.<sup>1</sup> Suppose that we are given a simple voting game (SVG)  $\mathcal{W}$  whose assembly (set of voters) is  $N$ . The singular (or atomic) event we shall consider is represented by a positive outcome (say, approval of a proposed bill) resulting from a particular division of  $N$  i.e., partition of  $N$  into ‘yes’ voters and ‘no’ voters. Since the outcome is positive, the set  $S$  of ‘yes’ voters must be a winning coalition of  $\mathcal{W}$ . Here the casting of ‘yes’ votes by the members of  $S$  constitute their joint actions that resulted in the positive outcome. The problem is then to quantify, for each voter  $v \in S$ , the degree (or extent) to which  $v$  has contributed to causing the outcome.<sup>2</sup>

Our assumption that the outcome of the division was positive involves no loss of generality. The same treatment – with obvious modifications – applies to a negative outcome. Formally, dealing with a negative outcome amounts to considering the SVG dual to  $\mathcal{W}$  instead of  $\mathcal{W}$  itself.

We shall confine ourselves to addressing the implications of the more general discussion, definitions and prescriptions in Braham and van Hees (2008b) to this voting scenario. We do not claim to have sufficient expertise for discussing causation in a more general setting.

We should stress that the only information that we assume to be available (or at any rate admissible) is the SVG  $\mathcal{W}$  and the way each of the voters voted in this particular division. We exclude any other information on inter-relationships among the voters, or on their preferences. This assumption is similar to that made in the theory of a priori voting power, except that here we have one additional piece of admissible information: namely, that one particular division actually occurred on the single occasion under consideration. This implies that the actual probability of this division is positive – which is not saying very much, because this probability may well be extremely small. So in this sense our approach here is still largely aprioristic. In real-life application, additional information, relevant to quantifying causal contributions, may be available and admissible. Thus the measure we shall propose should be regarded as a benchmark or an ideal limiting case.

On these grounds of actual ignorance (or of going behind a veil of igno-

---

<sup>1</sup>For definitions and explanations of the technical terms used in this paragraph, and of other terms from the theory of voting power used below, we refer the reader to our book, Felsenthal and Machover (1998).

<sup>2</sup>It may perhaps be objected that the causal contribution ought to be attributed not to the agent (the voter  $v$ ) but to the action ( $v$ 's vote). However, in the context of the voting scenario, where each voter takes a single action, this distinction makes no difference.

rance) we feel justified in assuming that a priori all possible divisions (all but one of which are taken to be counterfactual) are equiprobable.<sup>3</sup> Like Braham and van Hees, we base our quantitative allocation of causal contribution on the qualitative NESS test. 'NESS' is an acronym for 'necessary element of a sufficient set'. Roughly speaking, a condition makes a causal contribution to an outcome just in case this condition is a necessary element of a set of conditions that are jointly sufficient for producing the outcome. In Section 2 we will consider how to formulate the NESS test in the voting scenario. We shall argue that it has two non-equivalent versions, each of which is reasonable, depending on circumstances.

In Section 3 we shall argue that the *extent* to which a voter makes a causal contribution to the outcome ought to be quantified in absolute terms. A voter's *relative share* in the total causation must be regarded as a derived quantity, rather than the primary one. We shall then propose two versions of such an absolute measure, corresponding to the two versions of the NESS test presented in Section 2.

In what follows we shall treat [*extent of*] *responsibility* as identical to [*extent of*] *causation* – at least so far as the voting scenario is concerned. Here we go against Braham and van Hees (2008b), who express reservations against this identification. Their reservations may well be justified if they have in mind a different sense of 'responsibility' from ours. After all, this term can have all sorts of interpretations. Let us just say that this is a matter of definition: we simply *define* the degree (or extent) to which a voter has *responsibility* for the outcome of a particular division as identical to the degree (or extent) to which the voter contributed to *causing* that outcome. So we shall use these terms interchangeably, for stylistic variation.

## 2. Which SVG?

In an earlier version of their paper, Braham and van Hees (2008a) stated the NESS test for condition *c* to be counted as *a causal condition* for an event *e* as follows:<sup>4</sup>

- (i) *c* obtains, (ii) *e* obtains, and (iii) *c* is a necessary element of a sufficient set of conditions for *e*.

---

<sup>3</sup>This is the usual justification in the theory of a priori voting power. Braham and van Hees (2008b) also make this assumption, albeit tacitly.

<sup>4</sup>They call it the 'weak Ness test', as distinct from another version, which they call the 'strong Ness test'. However, these two versions are in fact equivalent, at least as far as the voting scenario is concerned. So for our purposes the distinction can be ignored. In quoting this earlier version of the NESS test, we have changed upper-case *C* and *E* to lower case, for the sake of consistency with their later statement of the test.

Arguably, an additional proviso ought to be added here: surely, *all other* elements of such a sufficient set, whose existence is required in (iii), must *also* obtain.

To see this, let us for a moment put aside the voting scenario. Suppose that Ms Borgia wishes to kill Giovanni by poison. She serves him with poisoned mushrooms and poisoned wine, each of which is sufficient to kill him. He eats the mushrooms, drinks the wine – and dies. Here there are four conditions: (1) she poisons the mushrooms; (2) she poisons the wine; (3) he eats the mushrooms; (4) he drinks the wine. And each of the four conditions is clearly a causal condition for his death, because both the set  $A$  consisting of (1) and (3) and the set  $B$  consisting of (2) and (4) are sufficient for killing him. Moreover, each of (1) and (3) is necessary for the lethality of  $A$ ; and each of (2) and (4) is necessary for the lethality of  $B$ . But now suppose that she fails to poison the wine. He eats, drinks and dies. Here (4), drinking the wine, is no longer a causal factor of his death, because (2) – the other member of  $B$  – did not actually obtain.

Indeed, in the final version of their paper, Braham and van Hees (2008b) modified their statement of the NESS test by adding the proviso in question:

There is a set of events that is sufficient for  $e$  such that: (i)  $c$  is a member of the set; (ii) all elements of the set obtain; (iii)  $c$  is necessary for the sufficiency of the set.

However, in the special voting scenario matters are less clear cut: it is not quite so obvious that the extra proviso ('all elements of the set obtain') must indeed be stipulated. We shall argue, and illustrate using Example 2.1, that whether this extra proviso is warranted depends on which counterfactual divisions are considered reasonable; and this, in turn, depends on whether 'yes' and 'no' votes are treated as entirely symmetric alternatives, as well as on additional considerations.

In the voting scenario, we have an SVG  $\mathcal{W}$  with assembly  $N$ . In a given [single] division, the set of 'yes' voters was  $S$ . Here the earlier form of the NESS test says:

*Definition 2.1 (NESS test)* Voter  $v \in N$  bears some responsibility for the division having a positive outcome iff (i)  $v \in S$ , (ii)  $S$  is a winning coalition, and (iii)  $v$  is a critical member of some winning coalition  $A$ .<sup>5</sup>

And the additional proviso is that  $A \subseteq S$ ; so that instead of Definition 2.1 we have

---

<sup>5</sup>Recall that requirement (iii) means that  $A$  is a winning coalition but  $A - \{v\}$  is a losing coalition.

*Definition 2.2 (NESS\* test)* Voter  $v \in N$  bears some responsibility for the division having a positive outcome iff (i)  $v \in S$ , (ii)  $S$  is a winning coalition, and (iii\*)  $v$  is a critical member of some winning coalition  $A \subseteq S$ .

Note that in the special case where  $S = N$  – that is, where there is a unanimous ‘yes’ vote – there is no difference between (iii) and (iii\*), so Definitions 2.1 and 2.2 become equivalent.

But is (iii\*) indeed reasonable where there is no unanimity? To examine this question, consider the following example.

*Example 2.1* Let  $\mathcal{W}$  be the SVG with  $N = \{a, b, c, d, e, f\}$  whose minimal winning coalitions are  $\{a, b, c\}$ ,  $\{a, b, d\}$  and  $\{a, e, f\}$ ; and let  $S = \{a, b, c, d, e\}$ .

It is easy to see that according to either test, each of  $a, b, c$  and  $d$  undoubtedly bears some responsibility for the (overdetermined) positive outcome. Note that although neither  $c$  nor  $d$  is critical in  $S$  itself, each of them is critical in some winning subcoalitions of  $S$ .

But what about  $e$ ? This voter is not critical in any subcoalition of  $S$ , but is critical in some other winning coalitions, for example in  $A = \{a, e, f\}$ . Thus whether we regard  $e$  as bearing any responsibility for the actual positive outcome depends on whether we stipulate (iii) rather than (iii\*).

The question as to whether (iii\*) is justified turns out to be quite delicate. In order to hold  $c$  responsible for the outcome, we had to entertain some *counterfactual* division in which  $d$  would have voted ‘no’ (and vice versa, interchanging  $c$  and  $d$ ). In order to hold  $e$  responsible, we must entertain some *counterfactual* division in which  $f$  would have voted ‘yes’ – but which (iii\*) does not allow us to entertain. Is it reasonable to entertain the former counterfactuality but not the latter?

We wish to suggest a tentative answer: *it all depends on the status of ‘no’ votes.*

If casting a ‘no’ vote is considered to be an act symmetric to that of casting a ‘yes’ vote – that is, these two acts are regarded as similar in kind albeit opposite in direction – then we ought to stick to (iii) and discard (iii\*). This is because a counterfactual division in which a voter who in fact voted ‘no’ *would have* voted ‘yes’ is entirely on a par with a counterfactual division of the opposite kind: in which a voter who in fact voted ‘yes’ *would have* voted ‘no’. So in Example 2.1  $e$ ’s ‘yes’ vote must be regarded as having made a causal contribution to the positive outcome.

On the other hand, if a ‘no’ vote is considered to be the default condition, an absence of action,<sup>6</sup> then (iii\*) must be upheld. It would be unreasonable to hold a given voter  $v$  responsible for the positive outcome of an actual division on the grounds that  $v$ ’s ‘yes’ vote would have been critical in a counterfactual division in which some other voter who in fact voted ‘no’, and so failed to act, *would have* acted and voted ‘yes’. So in Example 2.1  $e$ ’s vote cannot be taken to bear any responsibility for the positive outcome. To do so would be analogous to claiming that Giovanni’s imbibing of unpoisoned wine made a causal contribution to his death, on the counterfactual grounds that *had* Ms Borgia poisoned the wine, it *would have* been sufficient to kill him.

Another conceivable justification for (iii\*) is that since the members of  $N-S$  voted against the actual positive outcome, they are irrelevant and should be taken out of consideration in assessing the responsibility of the ‘yes’ voters.

At the end of Section 3 we shall present a more powerful argument in support of (iii\*). But for the time being, rather than making a definite choice between the two definitions, 2.1 and 2.2, we shall retain both as plausible alternatives, depending perhaps on the real-life situation.<sup>7</sup>

Formally speaking, the choice between the two definitions amounts to the following. Consider the voting scenario: an actual [single] division under the SVG  $\mathcal{W}$ , in which the set of ‘yes’ voters was a winning coalition  $S$ . When calculating the extent to which some  $v \in S$  is responsible for the positive outcome, then according to Definition 2.1 we must take account of the entire SVG  $\mathcal{W}$ , and count all the winning coalitions of this SVG in which  $v$  is a critical member. On the other hand, according to Definition 2.2 we must ignore all the actual ‘no’ voters, and count only the winning coalitions that contain  $v$  as a critical member and belong to the *subgame*

$$\mathcal{W}^S := \{T \in \mathcal{W} : T \subseteq S\}.$$

Note that the assembly of  $\mathcal{W}^S$  is  $S$  rather than  $N$ . Of course, in the special case where  $S = N$  – that is, a unanimous ‘yes’ vote – we have  $\mathcal{W}^S = \mathcal{W}$ .

### 3. Absolute Measure of Causation

Our discussion in this section up to Postulate 3.1 applies equally, irrespective of the choice between Definitions 2.1 and 2.2, both of which we retain for the

<sup>6</sup>Arguably, this is the case under the so-called ‘qualified majority’ decision rule in the EU Council of Ministers: abstention has exactly the same effect as a ‘no’ vote.

<sup>7</sup>In the mathematical theory of voting power, ‘yes’ and ‘no’ votes are treated as entirely symmetric alternatives. But in real-life applications this is not always the case: voting for a proposed bill that will change the status quo may be regarded as more radical than voting against it and for maintaining the status quo. See Footnote 6.

time being as plausible alternatives. In this discussion, clauses that assume Definition 2.1 come first, followed by 'or\*' and the corresponding clause that assumes Definition 2.2.

We aim to propose a cardinal measure that *quantifies the extent* of the voters' respective causal contributions to the outcome in the voting scenario; and does so in a way that is an intuitively reasonable extension of the *qualitative* test of Definition 2.1 or\* 2.2.

Let us recall the definition of the *Banzhaf score* (or *count*)  $\eta_v[\mathcal{W}]$  of a voter  $v$  in an SVG  $\mathcal{W}$ : it is the number of winning coalitions of  $\mathcal{W}$  that contain  $v$  as a critical member.

The most natural way of using the qualitative test of Definition 2.1 or\* 2.2 as a basis for a quantitative measure of responsibility is to postulate that, for each voter  $v$ , the extent of  $v$ 's responsibility for the positive outcome of the actual division is proportional to  $\eta_v[\mathcal{W}]$  or\*  $\eta_v[\mathcal{W}^S]$ , which is the number of coalitions  $A$  that satisfy condition (iii) of Definition 2.1 or\* (iii\*) of Definition 2.2.

Now we must address the question whether the extent of responsibility for the outcome is best regarded as fundamentally a relative or an absolute magnitude. In other words: does it only make sense to speak of a given voter's *relative share* in the total responsibility, or should we rather seek to quantify each voter's responsibility in absolute terms?

We wish to argue for the latter view: the extent of responsibility is primarily absolute; whereas the voter's relative share of responsibility is a derivative quantity, obtained from the more basic absolute magnitude by normalization.

In our opinion the situation here is broadly similar to that in the theory of voting power. There it is erroneous to regard the relative Banzhaf index  $\beta$  as the correct measure of a voter's a priori degree of influence over the outcomes of an SVG. Rather, it is the *absolute* Penrose measure  $\psi$  that provides a reasonable formalization of a voter's a priori influence.<sup>8</sup> The relative Banzhaf index  $\beta$ , which is derived from  $\psi$  by normalization, quantifies the voters' *relative shares* in the total amount of absolute voting power (i.e., the sum of the absolute voting powers of all voters) – *a total which is by no means fixed*, but depends on the SVG (i.e., the formal decision rule).

Moreover, disregarding  $\psi$  and taking  $\beta$  as the basic quantity not only entails loss of vital information about a voter's influence, but – as the history of the theory of voting power demonstrates – can lead to serious confusion and

---

<sup>8</sup>In our book, Felsenthal and Machover (1998), we denoted the Penrose measure by ' $\beta'$ ' and referred to it as the '[absolute] Banzhaf measure'. We now prefer the present notation and terminology.

error.<sup>9</sup>

*Example 3.1* Consider two SVGs, with three voters each: the simple majority SVG  $\mathcal{M}_3$ , in which the winning coalitions are all those containing at least two voters; and the unanimity SVG  $\mathcal{B}_3$ , in which the only winning coalition is the entire assembly. In each case suppose that a division has taken place in which all three voters voted ‘yes’.

In both cases, by reason of symmetry each voter bears  $\frac{1}{3}$  of the total responsibility. But it is intuitively clear – at least it seems clear to us – that each of the voters of  $\mathcal{B}_3$  bears a greater amount of responsibility than a voter of  $\mathcal{M}_3$ : in the former case the voter’s ‘yes’ was *actually* essential, and s/he was *actually* able to prevent the outcome; whereas in the latter no voter was actually in this position, but only in the *counterfactual* divisions in which the other two voters *would have* voted in opposite ways.

Thus, the extent of causal contribution cannot be an essentially relative quantity.

Nor is it reasonable to apportion responsibility to the voters in non-overlapping portions of some fixed quantity. We have postulated that the extent of  $v$ ’s responsibility for the positive outcome of the actual division should be proportional to  $\eta_v[\mathcal{W}]$  or\*  $\eta_v[\mathcal{W}^S]$ . This number is the cardinality of the event  $E_v$  that  $v$  is a critical member of the ‘yes’-voting coalition in a division with positive outcome in  $\mathcal{W}$  or\* in  $\mathcal{W}^S$ .<sup>10</sup> But in general the events  $E_v$  and  $E_u$  for two distinct voters  $v$  and  $u$  may well overlap; so their cardinalities (or relative frequencies) *cannot* be treated as non-overlapping portions of some fixed quantity.

Ignoring these considerations invites the nemesis of pathological consequences. If we take the extent of responsibility as a fundamentally relative quantity, then we must take  $v$ ’s responsibility to be obtained from  $\eta_v[\mathcal{W}]$  or\*  $\eta_v[\mathcal{W}^S]$  by normalization. The result is the relative Banzhaf index of  $v$  in  $\mathcal{W}$  or  $\mathcal{W}^S$ , i.e.  $\beta_v[\mathcal{W}]$  or\*  $\beta_v[\mathcal{W}^S]$ . This is indeed what Braham and van Hees (2008b) end up doing.

But that would lead to unacceptable consequences. The following example is adapted from our book Felsenthal and Machover (1998: Ex. 7.8.14).

*Example 3.2* Consider the weighted voting game (WVG)

$$\mathcal{W} = [11; 6, 5, 1, 1, 1, 1].$$

---

<sup>9</sup>See discussion of this in our book and historical survey article: Felsenthal and Machover (1998, 2005).

<sup>10</sup>This event is a set of equiprobable atomic events, all of which, or all but one, are counterfactual.

Suppose that in an actual division all voters unanimously approved the proposed bill. So in this case  $\mathcal{W}^S = \mathcal{W}$ . Then according to the 'relativist' view of causal contribution the extent of causal contribution of voter 1 (the heaviest voter) to the positive outcome is  $\beta_1[\mathcal{W}] = \frac{11}{23}$ .

Now suppose the heaviest voter has bought up the single share of the third voter. We get a new WVG:  $\mathcal{W}' = [11; 7, 5, 0, 1, 1, 1]$ . Again, suppose that in an actual division all voters unanimously approved the proposed bill. Now, according to the same relativist view, the extent of the first voter's causal contribution to the positive outcome is  $\beta_1[\mathcal{W}'] = \frac{17}{36}$ , which is *smaller* than  $\frac{11}{23}$ .

This is highly paradoxical: in the latter case the first voter cast both his 6 old shares and the new share acquired from the third voter in support of the bill – and yet s/he seems to have had a smaller causal contribution to the outcome than in the former case, when s/he only had the 6 old shares!<sup>11</sup> Of course, this is not a real pathology of causation, but only a consequence of trying to measure it in purely relative terms rather than absolute ones.

What would be a reasonable way of measuring a voter's absolute responsibility? We do not presume to provide a definitive answer. But if, following Braham and van Hees (2008b), we agree to take Definition 2.1 or\* 2.2 as a starting point, then a plausible answer suggests itself.

Using Penrose's measure  $\psi$  of voting power is *not* the right answer, because it is a two-sided measure that treats positive and negative outcomes in a symmetric way; whereas here we are concerned with a positive outcome as the privileged one.<sup>12</sup>

However, the theory of a priori voting power does have a plausible measure, which has long been known: Coleman's measure, aptly termed by him 'the power to prevent action' (1971: 280–1). In the present context, 'action' means positive outcome of a division. Voter  $v$ 's power to prevent a positive outcome,  $\gamma_v[\mathcal{W}]$ , is by definition the conditional probability, given that the outcome would be positive, of the event that  $v$ 's 'yes' vote would be critical to the outcome. So, in the voting scenario in which the set of 'yes' voters in an actual division was a winning coalition  $S$ , we propose the following as a measure of  $v$ 's responsibility for (or absolute causal contribution to) the positive outcome.

<sup>11</sup>This example can be adapted – with equally paradoxical upshot – to another scenario described by Braham and van Hees (2008b: Examples 6.1, 6.2). Replace the voters by 'firms' and their voting weights by the respective amounts of some toxin they dump in a river; the quota 11 is now the minimal amount of the toxin which is sufficient to kill all the fish in the river.

<sup>12</sup>Of course, we could start from an actual negative outcome and try to measure the extent of voters' responsibility to it. But this is quite another matter: as we noted in the Introduction, it amounts to considering the SVG dual to  $\mathcal{W}$  instead of  $\mathcal{W}$  itself.

*Definition 3.1* If  $v \notin S$ , then  $v$ 's *absolute responsibility* for the positive outcome equals 0.

If  $v \in S$ , then  $v$ 's *absolute responsibility* for the positive outcome is given by

$$\gamma_v[\mathcal{W}] := \frac{\eta_v[\mathcal{W}]}{\omega} \text{ or* } \gamma_v[\mathcal{W}^S] := \frac{\eta_v[\mathcal{W}^S]}{\omega^S},$$

where  $\omega$  is the number of winning coalitions in  $\mathcal{W}$  and  $\omega^S$  is the number of winning coalitions in  $\mathcal{W}^S$ .

Roughly speaking, if  $v \in S$  then the extent of  $v$ 's responsibility for the actual positive outcome is measured by the probability that, in the event of a positive outcome occurring,<sup>13</sup>  $v$  would be in a position to prevent it, if s/he so wished.

*Example 3.3* Let us go back to Example 3.1 and apply our measure to the two SVGs,  $\mathcal{M}_3$  and  $\mathcal{B}_3$ , assuming again that in both cases the three voters unanimously voted 'yes'. In both cases,  $S = N = \{1, 2, 3\}$ , so our measure of responsibility of a voter is the value of  $\gamma$  using the entire SVG. It is easy to see that  $\gamma_i[\mathcal{M}_3] = \frac{1}{2}$  for  $i = 1, 2, 3$ ; and  $\gamma_i[\mathcal{B}_3] = 1$  for  $i = 1, 2, 3$ .

This is intuitively reasonable: where unanimity was necessary for the actual positive outcome, each voter bears full (maximal) responsibility for it. But where two 'yes' votes would have been sufficient, each voter bears less than full responsibility.

Now let us go back to Example 2.1. Here we get very different results depending on whether we use Definition 2.1 or\* 2.2.

First, let us apply Definition 2.1, so that we must use the whole of  $\mathcal{W}$  for computing  $\gamma$ . Of course,  $f$ , who actually voted 'no', bears 0 responsibility for the positive outcome. For the others, who voted 'yes', we get:

$$\gamma_a[\mathcal{W}] = 1, \gamma_b[\mathcal{W}] = \frac{9}{17}, \gamma_c[\mathcal{W}] = \gamma_d[\mathcal{W}] = \frac{3}{17}, \gamma_e[\mathcal{W}] = \frac{5}{17}.$$

Note that  $a$ , who is a vetoer (blocker) in  $\mathcal{W}$  bears full responsibility – which is intuitively as it ought to be.

But if we apply Definition 2.2, then we must use the subgame  $\mathcal{W}^S$  for computing  $\gamma$ . Again,  $f$  bears 0 responsibility. For the 'yes' voters we get:

$$\gamma_a[\mathcal{W}^S] = 1, \gamma_b[\mathcal{W}^S] = 1, \gamma_c[\mathcal{W}^S] = \gamma_d[\mathcal{W}^S] = \frac{1}{3}, \gamma_e[\mathcal{W}^S] = 0.$$

Note that both  $a$  and  $b$  are vetoers in  $\mathcal{W}^S$ , so both bear full responsibility. But  $e$  is a dummy in  $\mathcal{W}^S$ , and hence bears 0 responsibility.

---

<sup>13</sup>This event is a set of equiprobable atomic events, all but one of which are counterfactual. Which counterfactuals are admitted depends on whether we opt for Definition 2.1 or\* 2.2.

Finally, let us go back to Example 3.2. Here again we have assumed a unanimous 'yes', so our measure of responsibility of a voter is the value of  $\gamma$  using the entire SVG. For  $\mathcal{W} = [11; 6, 5, 1, 1, 1, 1]$ , labelling the voters in the order in which they are listed, we get

$$\gamma_1[\mathcal{W}] = 1, \gamma_2[\mathcal{W}] = \frac{31}{33}, \gamma_i[\mathcal{W}] = \frac{1}{33} \text{ for } i = 3, \dots, 7.$$

Whereas for  $\mathcal{W}' = [11; 7, 5, 0, 1, 1, 1]$  we get

$$\gamma_1[\mathcal{W}'] = 1, \gamma_2[\mathcal{W}'] = \frac{15}{17}, \gamma_3[\mathcal{W}'] = 0, \gamma_i[\mathcal{W}'] = \frac{1}{17} \text{ for } i = 4, \dots, 7.$$

Voter 1 is a vetoer in both  $\mathcal{W}$  and  $\mathcal{W}'$ , and so bears full responsibility for the outcome in both cases. But the formation of the bloc (annexation of voter 3 by voter 1) affects the other voters in different ways. The responsibility of voter 2 is reduced; voter 3 becomes a dummy, so bears 0 responsibility; but the responsibility of the remaining voters increases considerably – so much so, that the *relative share* of voter 1 in the total is somewhat smaller than before.

Having proposed two alternative measures of responsibility for the voting scenario (Def. 3.1), we are now in a position to put forward an argument for preferring one of them. Consider the following postulate:

*Postulate 3.1* In the voting scenario, in which the set of 'yes' voters in an actual division was a winning coalition  $S$ , if voter  $v$  is a critical member of  $S$  then the measure of  $v$ 's responsibility for the positive outcome must be at least as great as that of any voter who is not a critical member of  $S$ .

The intuitive justification for this postulate is that a voter whose contribution to the positive outcome was critical in the actual division should not bear less responsibility for that outcome than a voter whose contribution would only have been critical in a counterfactual division.

Now, it is clear that the measure  $\gamma[\mathcal{W}^S]$ , based on the subgame  $\mathcal{W}^S$ , clearly satisfies Postulate 3.1. Indeed, any voter  $v$  who is critical in  $S$  is a vetoer in  $\mathcal{W}^S$ , so  $\gamma_v[\mathcal{W}^S] = 1$ , which is the greatest possible value of  $\gamma$ .

On the other hand, the measure  $\gamma[\mathcal{W}]$ , based on the entire SVG  $\mathcal{W}$ , does not in general satisfy Postulate 3.1, as can be seen from the following example.

*Example 3.4* Let  $\mathcal{W}$  be as in Example 2.1; but now let  $S = \{a, b, e, f\}$ . Here  $a, e$  and  $f$  are critical in  $S$ , whereas  $b$  is not critical in  $S$  or indeed in any sub-coalition of  $S$ . We have

$$\gamma_a[\mathcal{W}^S] = 1, \gamma_b[\mathcal{W}^S] = 0, \gamma_e[\mathcal{W}^S] = \gamma_f[\mathcal{W}^S] = 1.$$

However,

$$\gamma_a[\mathcal{W}] = 1, \gamma_b[\mathcal{W}] = \frac{9}{17}, \gamma_e[\mathcal{W}] = \gamma_f[\mathcal{W}] = \frac{5}{17},$$

contrary to Postulate 3.1.

Thus if we adopt Postulate 3.1, which seems to us intuitively reasonable, then we must opt for the starred part of Def. 3.1. This, in turn, implies that we should prefer condition (iii\*) and opt for Def. 2.2 rather than Def. 2.1.

#### 4. Conclusion

The problem addressed in this note is how to measure the extent of causal contribution (which we identify with responsibility) in the voting scenario. Moreover, we confine ourselves to the quasi-aprioristic approach, in which the only admissible factual information is that a particular division occurred on a single occasion under a given SVG.

Our proposed solution is tentative. We do not claim it is the only plausible way of solving the problem. What we do claim is that *if* one adopts the conceptual framework of Braham and van Hees (2008b), based on the NESS test – which seems quite plausible to us – *then* one should go about it as we have suggested: by first defining an absolute measure of responsibility, using Coleman's (1971) power to prevent action.

We do not dismiss the relative degree of causation defined by Braham and van Hees (2008b); rather, we claim that it cannot be regarded as primary, but as a derived quantity, obtained from the primary absolute measure  $\gamma$  by normalization. The point is that by *ignoring* the latter and *starting* with the former one loses important information.

The relative degree of causation may in fact be needed for certain purposes: for example if some authority wishes to share among the voters a fixed reward (or a fixed penalty) in proportion to their respective causal contributions to the outcome. Of course, this may well lead to some paradoxical results, as can be seen from our Example 3.2.<sup>14</sup> But these results are not the 'fault' of causation. Rather, they are the consequence of trying, as it were, to fit a square peg into a round hole: to distribute non-overlapping shares of a fixed quantity in proportion to overlapping parts of a variable quantity.

Finally, let us note that the method advocated here for constructing a measure of responsibility may be extended to cover a positive outcome obtained

---

<sup>14</sup>This is particularly striking in the pollution scenario described in Footnote 11: if the firm that is the worst polluter takes over one of the lesser offenders, and dumps into the river the latter's toxins as well as its own, then it has to pay a *smaller* fine than in the case where there is no takeover!

under a ternary decision rule – also known as *ternary voting game*, briefly TVG – which admits abstention as a voter's input, distinct from both 'yes' and 'no'.<sup>15</sup> Clearly, a voter who actually voted 'no' must be assigned 0 responsibility for the outcome. If voter  $v$  actually voted 'yes' or abstained, then the extent of  $v$ 's responsibility for the actual positive outcome is measured by the conditional probability that, in the (actual or counterfactual) event of a positive outcome occurring,  $v$  would be in a position to prevent it, if s/he so wished. A tricky question, however, is which TVG ought to be used for computing this conditional probability: the original TVG or some sub-TVG; and in the latter case, how that sub-TVG is to be defined. This question is beyond the scope of the present paper.

### Acknowledgements

While working on this paper both authors were co-directors of the Voting Power and Procedures Programme at the Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science. This programme is supported by Voting Power in Practice Grant F/07 004/AJ from the Leverhulme Trust.

### References

- Braham, M. (2008), Social power and social causation: Towards a formal synthesis, in: M. Braham and F. Steffen (eds.), *Power, Freedom, and Voting: Essays in Honour of Manfred J. Holler*, Springer, 1–21.
- Braham, M. and M. van Hees (2008a), Degrees of Causation (earlier draft, mimeograph). Downloadable from: <http://tinyurl.com/232nja>
- Braham, M. and M. van Hees (2008b), Degrees of Causation, (mimeograph). Forthcoming in *Erkenntnis*. Downloadable from: <http://tinyurl.com/6d9n6q>
- Coleman J. S. (1971), Control of collectivities and the power of a collectivity to act, in: B. Lieberman (ed.), *Social Choice*, Gordon and Breach, 269–300.
- Felsenthal D. S. and M. Machover (1998), *The Measurement of Voting Power: Theory and Practice, Problems and Paradoxes*, Edward Elgar.
- Felsenthal D. S. and M. Machover (2005), Voting power measurement: A story of misreinvention, *Social Choice and Welfare* 25, 485–506.

---

<sup>15</sup> See Ch. 8 of Felsenthal and Machover (1998).