Collective Choice as Information Theory: Towards a Theory of Gravitas

Wilmers, George

2010

MIMS EPrint: **2010.12**

ISSN 1749-9097

Collective Choice as Information Theory: Towards a Theory of Gravitas

by

George Wilmers

June 26th, 2008

ABSTRACT

The present paper introduces a new approach to the the theory of voting in the context of binary collective choice, which seeks to define a dynamic optimal voting rule by using insights derived from the mathematical theory of information. In order to define such a voting rule, a method of defining a real-valued measure of the weight of independent opinion of an arbitrary set of voters is suggested, which is value free to the extent that it depends only on probabilistic information extracted from previous patterns of voting, but does not require for its definition any direct information concerning either the correctness or incorrectness of previous voting decisions, or the content of those decisions. The approach to the definition of such a measure, which the author calls gravitas, is axiomatic. The voting rule is then defined by comparing the gravitas of the set of those voters who vote for a given motion with the gravitas of the set of those who vote against that motion.

1 What Can We Learn from a Council of Elders?

As a motivating thought experiment[1] let us consider a precapitalist tribal society governed by a hereditary chief who takes all decisions *de jure*, but who is advised by a council of elders M which he chairs. Custom has determined that, after due deliberation, but prior to the chief making a final decision on any resolution before M, each elder must pronounce a declaration of opinion for or against the resolution: abstentions are not permitted. Let us imagine that M is considering a particular resolution. If the chief is wise then he will listen carefully to the advice he is given by the elders on the resolution; but how should he evaluate it? He may perhaps reason that, since his position is hereditary, he is unlikely to be wiser than the average elder, even though he happens to be possessed of a certain mathematical knowledge and ability; hence his best policy may be to efface entirely all his own subjective judgments about the matters under deliberation both now and previously, and also to efface all his personal opinions about the value of the previous judgments of the elders. However if the chief is to eliminate from consideration all such personal judgments, then he must find some objective way to compare the weight of opinion of the set of those elders who are in favour of the particular resolution against the weight of opinion of those who are against the resolution. How are these two weights of opinion to be measured?

Our chief could of course simply count up those in favour and those against the resolution, and compare the resulting cardinalities, as the leaders of the great western democracies would surely enjoin him to do[2], but his mathematical learning makes him extremely reluctant to throw away the extensive objective information which is contained in the pattern of advice given to him by the elders concerning previous resolutions. Also he has noticed that in the past the members of certain groups of elders have generally voted together in a rather predictable manner, so that he does not feel it appropriate to count their votes as if they represented quite separate opinions. He reflects that he would prefer to have at his disposal a measure of weight of *independent* opinion of each of the two sets of elders representing opposing viewpoints on the merits of approving the resolution. So, in order to ensure that his approach is truly objective, the chief decides to erase from his memory all details about the actual content of previous resolutions and of any advice he has been given previously, and to treat in a formal mathematical manner the information contained in the resulting abstract matrix of the elders' declarations for or against all previous proposals.

[1] The present section arose from my reflections on a conversation in 1968 between John Bell and my late father, recorded in John's autobiography *Perpetual Motion*.

[2] provided, of course, that the results of such a calculation were likely to be consistent with their own assessments of the correct decision.

The chief's mathematical problem is now how to extract from this remaining information two weights of independent opinion for comparison. This problem is in essence the subject matter of the present paper.

2 The Notion of Gravitas: A Preliminary Discussion

Our formal starting point is a fixed assembly M of n voters in a binary choice context, where M is endowed with a probability distribution σ on the set of possible of divisions $D(M)$ of M. Formally a *division* α of M is just a map from M to $\{0,1\}$ which represents the *event* that the members of M vote on the latest motion before the assembly in such a manner that for all $a \in M$ a votes yes if $\alpha(a) = 1$ and no if $\alpha(a) = 0$. Since we are identifying divisions with events, our notation will allow logical disjunctions of divisions also to be treated as events, so that e.g. for $\alpha, \beta \in D(M)$, $\alpha \vee \beta$ is the same event as $\beta \vee \alpha$.

We may think of σ as derived by some statistical rules from the evidence of previous voting records. We shall not concern ourselves here with exactly what statistical procedures are used to derive such an *a posteriori* probability distribution on $D(M)$, but will instead take it as given. Thus we start with a mathematical idealization of the problem in the previous section. In general σ will be dependent on time since it will change as further information of the voting records of the members of M is accumulated. In the discussion below however we shall mostly treat σ as if it were fixed at a particular moment in time, and will take it as given at that moment in time, even though the concepts defined below should properly be thought of as defined relative to $\sigma(t)$ and variable time t.

Our fundamental question can now be phrased as follows. Suppose a new motion is presented to M and a given subset A of voters of M vote one way on the motion while the complement of A in M, A^c, vote the other way. Does there exist some natural *measure* which we can define in order to compare the "weight of independent opinion" of A with that of A^c ?

In section 3 we shall approach this question from an axiomatic standpoint, and we shall call this idea of the weight of independent opinion the *gravitas* of A, denoted by $G^\sigma(A)$. We formulate strong natural axioms for gravitas for arbitrary σ, which generalise the special classical situation in which σ is taken *a priori* to be the uniform distribution on $D(M)$, i.e. where the voters are *a priori* considered to vote independently with each voter voting yes with probability $\frac{1}{2}$.

In particular the quantity $G^\sigma(A) - G^\sigma(A^c)$, or *gravitas margin* generalises the classical notion of margin. We show that there exists a measure which satisfies the axioms, which we call polarity-free entropy (PFE). Although it does not seem easy to find an alternative measure to PFE which satisfies the given axioms for gravitas (other than a trivial translation by a constant), it is as yet unclear if there exist *intuitively convincing* additional axioms which would make PFE the unique solution for G^σ.

Given a notion of gravitas G^σ, we may define a voting rule R_{G^σ} by setting, for any division α such that the set of those who vote yes in α is A,

$$R_{G^\sigma}(\alpha) = \begin{cases} 1 & \text{if } G^\sigma(A) > G^\sigma(A^c) \\ 0 & \text{otherwise} \end{cases}$$

R_{G^σ}, which we call the *gravitas majority rule*, may be regarded as a conceptual generalization of the simple majority rule for the case in which the information contained in σ is available. It may be described as a realization of the intuitive concept of *rule by weight of independent opinion.* Furthermore, by analogy with the classical margin $|A| - |A^c|$, we will call the quantity $G^\sigma(A) - G^\sigma(A^c)$ the *gravitas margin*; the intuitive idea of the gravitas margin is to provide an indicator of reliability of a judgment arrived at by applying the rule R_{G^σ}. In the remainder of this section we shall consider briefly the philosophical background to these ideas.

The axiomatic approach which we are adopting differs considerably from that of more traditional semantic constructions which are used to interpret the meaning of the act of voting. In particular we make no *a priori* assumption that in the act of voting the individual voters are expressing personal opinions or personal preferences which are in any sense independent of the opinions or preferences of other voters. Our intuitive philosophical focus is rather on analysing the properties of the *sets* of temporarily like-minded voters A and A^c which form when the assembly M is considering a particular motion, and on treating these subsets as being the important collective actors in an information theoretic analysis of voting.

There exists a large corpus of scholarly work on the mathematics of democratic choice, most of which can trace its philosophical origins either to the (quite separate) work of the 18th century luminaries Condorcet [85] and Rousseau [62], or to the 20th century game theoretic considerations of social choice theorists arising from the celebrated impossibility theorem of Arrow [63]. In the case of unicameral binary choice the former tradition, which we may loosely call the *epistemic* tradition[3], has been concerned primarily with the problem of examining the mathematical conditions under which a majority decision rule can be

[3] see Cohen[86] and Coleman and Ferejohn [86] for philosophical discussions of the concept of an epistemic justification of democracy, and also for a critical discussion of proceduralist approaches.

theoretically justified in the context where an objectively correct answer is assumed to exist[4], while the latter tradition is concerned with the reconciliation of individual subjective preference orderings and seeks typically to examine under what conditions decision rules can avoid certain types of paradox or inconsistency. However, to our knowledge, there has been no work done on the axiomatic or mathematical foundations of a theory which would attempt to generalise the classical ideas of either Condorcet or Rousseau to the situation in which *extra objective information* is available in the form of the probability distribution σ.

The notion of gravitas which we present here and its associated decision rule R_{G^σ} could naturally be seen as belonging to the epistemic tradition. However the author believes that the notion of gravitas is relevant not just to a "Condorcet jury" type of context where an objectively correct answer is assumed to exist, but to a much more general context in which we require only that a correct answer to a motion put before M is accepted as existing with a *normative but probabilistic sense given to the meaning of the word correct, as being defined relative to certain limited but precisely defined information.* In the present case the limited information is taken to consist of σ together with the actual division of the voters on the given motion. Thus correctness in such a sense is an information-theoretic and relative notion: on the basis of certain symmetry and information theoretic principles, if we strictly limit the information available as above, then a particular outcome is deemed probabilistically correct *in that context*. Such a notion of probabilistic correctness relative to a precise informational framework may be viewed as an attempt to transcend the traditional distinction between epistemic and proceduralist interpretations of voting, and to provide a common epistemic analysis of voting theory[5]. This is however an idea of for epistemic interpretation of voting which is very different from the usual notion of what might constitute an epistemic explication of voting[6]. We will not pursue the analysis of this idea further here, since it is peripheral to the development of the main ideas of this paper.

The general theory of voting is associated with probability theory in various

[4]See e.g. Grofman, Owen and Feld [83], Ladha [92], Borland [94], List and Goodin [01], and List [04] for details.

[5]It may be noted that the philosophical idea of a separate notion of probabilistic correctness relative to limited information makes sense even in the case when we suppose that there exists an "objectively true" answer. For example, in a jury trial, the criterion for conviction is typically that guilt is proved "beyond reasonable doubt". If therefore we make the reasonable assumption that all judgments in such trials are *de facto* made on a probabilistic basis, then, given that the information which can be made available to a jury is of necessity limited, a jury (or indeed an individual jury member) may in fact make a decision which is probabilistically *correct on the basis of the evidence available*, but which is nonetheless *incorrect in an absolute sense.* Our restriction of the admissible information available to the decision rule to σ together with the actual division of the voters, may in this case be interpreted as a uniform (or fair) method of reifying the information contained in the accumulated subjective judgments of jury members on the evidence available to them.(Of course this presupposes that an estimate for σ is actually available, which would not be the case for a one-time only jury).

[6]See Cohen [86] for an account of the latter.

ways, notably in the classical theory of voting power, and in Condorcet style justifications of majority decision rules. However we may reasonably ask the question why there has been so little theoretical work done at a foundational level on optimal collective decision rules in a context where additional objective information concerning prior individual voting records of members of an assembly is available, and in particular why that most powerful tool of mathematical reasoning under uncertainty, information theory[7], has been so strikingly absent from deliberations. There are two related reasons for this situation, both of which have their origins in the tradition of centuries. The first of these reasons is that the foundational principle of "one person one vote" (OPOV), however hierarchically modified, underlies in some form or other all modern institutional collective forms of decision making; thus since the academic field of study of collective decision making is dominated by a consideration of *existing* types of institution, rather than a study of what might be possible, the consideration of fundamentally more complex decision rules invoking the use of additional information is normally ruled out *a priori*[8]. The other, related, reason is that, despite its rather weak theoretical justification, OPOV and its natural corollary of majority rule are ideologically so closely associated with the contemporary political concept of democracy, that any suggestion that some other conflicting principle might be both more profound, more equitable, and might produce better collective judgments, is likely to meet with incredulity at best. In addition there are two further technical reasons why the point of view advocated in the present paper might not have appeared worthy of consideration until relatively recently. On the one hand the appropriate mathematical ideas from information theory have only been current in the last half century, while on the other hand the necessary technology of instant communication, and the computational power necessary to process the raw voting data in order to estimate numerical values for a notion of gravitas have only become available within the last twenty years[9].

In the general context of the idealised Condorcet jury, where there exists a clearly defined objectively correct answer associated with each motion put before the assembly, it will in many cases be possible to carry out experiments to determine empirically whether or not a rule R_{G^σ} associated with a particular definition of gravitas G compares favorably in the decisions it generates when comparison is made with the simple majority rule, or indeed with a rule $R_{G'^\sigma}$ where G' is some other notion of gravitas. For example, where a disease can be infallibly diagnosed by some laboratory test, a panel of medical experts could be

[7]in particular Shannon's notion of entropy [64]: see e.g. Paris [94] for a modern detailed axiomatic presentation of the use of entropy in probabilistic reasoning.

[8]We may note here that types of information other than that encoded in σ might in principle also be recorded and used in the calculations of a decision rule; for example normalised information about the strength of conviction which individual voters attach to individual judgments could be recorded and used in some way. A closely related point is made in Dummett's discussion of Arrow's theorem in Dummett [84].

[9]I do not intend by this statement to minimise the difficulty of the computational problems involved, which I have not addressed here, and which would certainly be substantial in the case of a large electorate.

asked to evaluate for a lengthy sequence of patients whether or not, on the basis of clinical evidence, each of them had the disease . The votes of the experts would in each case be aggregated separately using R_{G^σ} and the simple majority rule, and a comparison of the results could then be made with the objectively correct answers which would be supplied by applying the laboratory test. Extensive experiments of this kind over a variety of different scientific domains could be used to provide strong evidence for or against the efficacy of a rule R_{G^σ} for a particular definition of G. If it turns out that a particular definition of G can be uniquely characterised by a convincingly natural set of axioms, then empirical evidence of the above kind could provide powerful independent evidence in favour of adopting R_{G^σ} as a decision rule in a more general context.

We should remark that, although in the Condorcet jury context there has been considerable research carried out concerning the performance of various voting rules which employ extra information concerning the *competence* of individual voters, such voting rules have an entirely different nature to the approach adopted in the present paper, since they depend on information concerning the correctness of previous judgments, whereas our framework of analysis makes no assumption that such information is available.

3 Axioms for Gravitas

Before we state our axioms we need to establish some simple notation. The probability distribution σ on $D(M)$ extends naturally to a probability function on the set of disjunctions of elements of $D(M)$ and we shall identify σ with this extension, so that for example if $\alpha, \beta \in D(M)$ with $\alpha \neq \beta$, then $\sigma(\alpha \vee \beta) = \sigma(\alpha) + \sigma(\beta)$. Also for every $A \subseteq M$, σ induces a probability distribution σ_A on $D(A)$ the set of divisions of A. In fact for any $\alpha \in D(A)$ $\sigma_A(\alpha) = \sigma(\alpha)$.

We now introduce our axioms, and explain briefly the motivation behind them. It is understood that the axioms should hold for all possible M and σ. We also assume that the gravitas function G^σ takes real number values in the interval $[0, \infty)$.

Continuity Axiom

For any $A \subseteq M$, $G^{\sigma}(A)$ is continuous as a function of σ.

This axiom simply expresses mathematically the intuitive idea that the gravitas function should be a smooth function with respect to changes in σ: although gravitas might change quite rapidly as σ changes, there should be no sudden jumps in its values.

Locality Axiom

For every $A \subseteq M$ $G^{\sigma}(A)$ is a function of σ_A alone.

This axiom expresses the intuitive idea that the gravitas of the set of voters A should depend only on the behaviour of the voters in A, and should in particular be independent of how the remaining voters of M vote. While this property is very natural, there does exist however an alternative natural point of view, and we shall return to this in our considerations later.

Voter Renaming Axiom

Let π be a permutation of the voters of M which, given σ, induces the probability distribution σ^{π} on M defined by $\sigma^{\pi}(\alpha^{\pi}) = \sigma(\alpha)$ for each $\alpha \in D(M)$, (where α^{π} denotes the obvious permutation of the division α). For any $A \subseteq M$ let A^{π} denotes the image of A under π.

Then $G^{\sigma^{\pi}}(A^{\pi}) = G^{\sigma}(A)$.

This axiom is just a version of the familiar idea of anonymity; the gravitas of A should not depend on the names which the elements of A happen to possess but only on their properties as determined by σ.

Monotonicity Axiom

For any $A \subseteq M$ and $b \in M$, $G^{\sigma}(A) \leq G^{\sigma}(A \cup \{b\})$.

This axiom expresses the idea that adding a new member to a set of voters A cannot decrease the gravitas of A, given that the voting behaviour of the other members of A remains unchanged. Note that this natural assumption immediately implies that the voting rule R_{G^σ} is monotone.

Clone Axiom

For any $A \subseteq M$, if $a, b \in A$ are distinct voters such that the probability (calculated using σ) that a votes the same way as b is 1, then
$G^\sigma(A) = G^\sigma(A - \{b\})$.

This axiom just expresses the idea that if two voters in A behave identically, then one of them is redundant in calculating the the gravitas of A since the two voters vote systematically as if they were of one opinion. The axiom reflects the intuitive idea that in calculating gravitas we are seeking to count not voters, but independent points of view.

For any $A \subseteq M$ and $\alpha \in D(A)$, let $\overline{\alpha}$ denote the dual division to α in which each member of A votes the opposite way to the way they voted in α. We can now state our next axiom.

Polarity Free Axiom

For any $A \subseteq M$, $G^\sigma(A)$ depends only on the values $\sigma(\alpha \vee \overline{\alpha})$ where $\alpha \in D(A)$.

This axiom needs some explanation. The idea here is that the *actual* direction (for or against motions) in which voters vote is immaterial in calculating a measure of their independence: all that matters is their voting patterns *relative to each other*. So if σ were altered because a proportion of motions were arbitrarily replaced by their negations, this should not affect the value of $G^\sigma(A)$, assuming that the voters would reverse their votes in line with their beliefs. Obviously this axiom represents a strengthening of the Locality Axiom which could have been included in it. However because of its different and less obvious status, we have separated it from the Locality Axiom.

Let us denote by σ^*_A the probability distribution which is obtained from σ by considering just the set of events of the form $\alpha \vee \overline{\alpha}$ where $\alpha \in D(A)$. Thus the Polarity Free Axiom asserts that $G^\sigma(A)$ depends only on the information in σ^*_A. In the case when A is M we will write σ^* to denote σ^*_M.

We shall call an event of the form $\alpha \vee \overline{\alpha}$ a *polarity free division of* A. More generally any disjunction of polarity free divisions of M may be referred to as a *polarity free* event. Trivially polarity free events are closed under Boolean operations.

Clearly σ^* contains less information than σ. However the information which it contains has an interesting epistemological status as we now explain.

For the purpose of the present discussion let us now assume the existence of some objective notion of correctness for all motions presented to M. Then by complete analogy with σ there exists another probability distribution τ which encapsulates information about the *correctness* of the previous voting of voters in M. To make this precise we define for each division $\alpha \in D(M)$ an analogous event $\widehat{\alpha}$ by replacing "voting yes" by "voting correctly" and "voting no" by "voting incorrectly" in the definition of α. We call $\widehat{\alpha}$ a *truth-division of* M and we let $D^T(M)$ be the set of all truth-divisions of M. Also, given some $\widehat{\alpha} \in D^T(M)$, we may define its dual truth division $\overline{\widehat{\alpha}}$ to be $\widehat{\overline{\alpha}}$. Again by analogy we call an event of the form $\widehat{\alpha} \vee \overline{\widehat{\alpha}}$ a polarity free truth-division of M. Then analogously to σ there exists a probability distribution τ on $D^T(M)$ which could be defined in the same way from records as to whether voters voted correctly or incorrectly, *if such records existed*, as σ is defined from records of actual yes or no votes cast. Of course, unlike σ, the distribution τ will not in general be accessible to us, since except under rather special circumstances we will not have access to the data from which τ would be constructed.

$\tau(\widehat{\alpha})$ tells us the probability of the event $\widehat{\alpha}$ occurring, based solely on the record of correctness of the previous votes of members of M. However we may now notice that for any M and any $\alpha \in D(M)$ the events $\alpha \vee \overline{\alpha}$ and $\widehat{\alpha} \vee \overline{\widehat{\alpha}}$ are extensionally identical, and hence, provided that enough previous votes are taken into account, it will be the case that σ and τ nearly coincide for such events, i.e. that

$$\sigma(\alpha \vee \overline{\alpha}) \simeq \tau(\widehat{\alpha} \vee \overline{\widehat{\alpha}})$$

for any polarity free division $\alpha \vee \overline{\alpha}$ of M, where the approximation tends to equality as the number of previous votes taken into account increases. So *if we have no access to the records of correctness* from which τ would be constructed, we should regard the function σ restricted to these polarity free events as the best approximation available , say ρ^*, to the values which τ would give to the polarity free truth divisions; i.e. we define ρ^* by

$$\rho^*(\widehat{\alpha} \vee \overline{\widehat{\alpha}}) = \sigma^*(\alpha \vee \overline{\alpha})$$

for any polarity free division $\alpha \vee \overline{\alpha}$ of M.

Two linked foundational questions now arise. The first question may be stated as follows: if we are given *only* the information contained in σ^* how should ρ^* be extended to a probability distribution ρ defined on the whole of $D^T(M)$ in such a manner that from a purely information theoretic standpoint ρ is the best estimate of τ we can make in the absence of any other information? The second, related, question is: if instead we are given all the information contained in σ, how should we extend σ to a natural joint distribution on the Boolean algebra of events generated by $D(M) \cup D^T(M)$? These questions will not be considered further here but will be pursued in a later paper[10].

Our last two axioms generalise properties of the classical notion of margin. The absolute values of $G^\sigma(A)$ are intuitively less important than a comparison of the values of $G^\sigma(A)$ and $G^\sigma(A^c)$. For any measure of gravitas G we let $Mar_{G^\sigma}(A)$ denote the G^σ-*margin* of A in M, i.e. we define

$$Mar_{G^\sigma}(A) = G^\sigma(A) \text{ - } G^\sigma(A^c)$$

Now the classical margin of A (over A^c) is of course just $|A| - |A^c|$. So if $Mar_{G^\sigma}(A)$ is to generalise the classical margin we should expect that the two notions would coincide for the paradigm case of the uniform distribution on $D(M)$. Accordingly we may now state the following

Classical Margin Axiom

Let *unif* denote the uniform distribution on $D(M)$. Then for any $A \subseteq M$

$$Mar_{G^{unif}}(A) = |A| - |A^c|$$

In the case of the simple majority rule, the classical margin $|A| - |A^c|$ provides, in a Condorcetian analysis, an indicator of the probability that the majority decision is correct; although this analysis is dependent on absurdly idealised assumptions concerning voters' independence, nevertheless under these special conditions the classical margin possesses certain attractive invariance properties[11]. So it is natural that if we are seeking to generalise the concept of margin

[10] We confine ourselves here to noting that the first of these questions can reasonably be considered an analogue in collective choice theory of the problem in uncertain reasoning (by a single agent) of choosing a canonical probability distribution from a set of possible distributions constrained by certain data. The method of choice for solving the latter problem is the use of the maximum entropy principle (see e.g. Paris [94]), but without additional insight maximum entropy appears powerless to help in solving the former problem. The author believes however that the notion of gravitas can be used to provide the appropriate missing idea necessary to partially solve this problem.

[11] The importance of invariance properties of the notion of margin in a classical Condorcetian

to $Mar_{G^\sigma}(A)$ for arbitrary σ, then we should seek to ensure that this generalisation possesses a strong conceptual stability. Our final axiom below should be interpreted with this in mind.

For the purposes of defining our final axiom we will assume that $G^\sigma(A)$ satisfies the Locality, Polarity Free, and Voter Renaming axioms. Recall that we have insisted by the Locality and Polarity Free axioms that $G^\sigma(A)$ depend only on σ^*_A; in particular $G^\sigma(A)$ therefore depends only on the probability of polarity free events, i.e. of events which refer only to how the voters vote relative to each other, not to which way they actually vote. However, if we now take as given some such G and consider instead as a possible alternative notion of gravitas the expected value of G on A with σ^*_A conditioned upon the polarity free divisions corresponding to every possible way in which the members of A^c could divide, then we obtain a rather natural, but *not* locally defined quantity, which we will define below and will denote by $\mathcal{E}_{G^\sigma}(A)$. We call the function $\mathcal{E}_{G^\sigma}$ of subsets A of M the *polarity free expectation (over M)* of G^σ.

There is a slight notational difficulty in formally defining $\mathcal{E}_{G^\sigma}$. This difficulty arises because if we consider a polarity free division $\alpha \vee \overline{\alpha}$ where $\alpha \in D(A)$ and condition this event on the polarity free event $\beta \vee \overline{\beta}$ where $\beta \in D(A^c)$, then we need to consider the conditional probabilities of two possible alternative polarity free divisions of M, namely $\alpha\beta \vee \overline{\alpha}\overline{\beta}$ and $\overline{\alpha}\beta \vee \alpha\overline{\beta}$. The respective conditional probabilities of these events are given by

$$\frac{\sigma(\alpha\beta \vee \overline{\alpha}\overline{\beta})}{\sigma(\beta \vee \overline{\beta})} \text{ and } \frac{\sigma(\overline{\alpha}\beta \vee \alpha\overline{\beta})}{\sigma(\beta \vee \overline{\beta})} \quad .$$

Note that if $|A| = m$ say, then for each event $\beta \vee \overline{\beta}$ as above there are 2^m disjoint atomic polarity free divisions of M as above and 2^m corresponding conditionalised probability values summing up to 1. If we denote by $\sigma^*_{A|\beta\vee\overline{\beta}}$ the probability distribution just described, then this distribution may be thought of as a distribution on the 2^m polarity free divisions of the set $A \cup \{d\}$ where $d \notin A$ is treated as a placeholder element whose value in a division indicates the relative polarity of $\beta \vee \overline{\beta}$ to $\alpha \vee \overline{\alpha}$ corresponding to the two forms above. To be precise: given a fixed $\beta \in D(A^c)$, let $\gamma \in D(A \cup \{d\})$ be such that $\gamma(d) = 1$. Let $\alpha \in D(A)$ be defined by $\forall a \in A \ \alpha(a) = \gamma(a)$. Then the polarity free event $\gamma \vee \overline{\gamma}$ is identified with $\alpha\beta \vee \overline{\alpha}\overline{\beta}$.

With this interpretation we may define for non-empty A and A^c

$$\mathcal{E}_{G^\sigma}(A) = \frac{1}{2} \sum_{\beta \in D(A^c)} \sigma(\beta \vee \overline{\beta}) \ G^{\sigma^*_{A|\beta\vee\overline{\beta}}}(A \cup \{d\})$$

analysis of voting has been emphasized by List [04]

The factor of $\frac{1}{2}$ is present since otherwise because of the notation each event $\beta \vee \overline{\beta}$ would be counted twice. For the special cases when A or A^c is empty, it is natural to define $\mathcal{E}_{G^\sigma}(A) = G^\sigma(A)$.

Now, given some notion of gravitas G^σ, one can reasonably argue that, despite the nonlocality of its definition, $\mathcal{E}_{G^\sigma}(A)$ has an almost equally good claim to be considered as a measure of gravitas as $G^\sigma(A)$, since intuitively it just represents the expected value of $G(A)$ conditionalised on all possible appropriate events corresponding to the behaviour of the rest of the assembly, A^c. At first sight it would be be nice therefore if G^σ and its polarity free expectation $\mathcal{E}_{G^\sigma}$ could be made identically equal. This turns out to be too strong a requirement: it results in inconsistency. As we have stressed however the important function to be considered for possible invariance properties is the gravitas *margin* rather than gravitas itself. So it is pleasing to discover that the following strong axiom is in fact satisfiable:

Polarity Free Margin Invariance

For every $A \subseteq M$, $\qquad Mar_{\mathcal{E}_{G^\sigma}}(A) = Mar_{G^\sigma}(A)$

This concludes our list of axioms for the notion of gravitas.

4 Polarity Free Entropy (PFE)

In this section we define a measure, Polarity Free Entropy, or PFE, which satisfies all eight axioms for a notion of gravitas, G, described in the previous section, namely the Continuity, Locality, Voter Renaming, Monotonicity, Clone, Polarity Free, Classical Margin, and Polarity Free Margin Invariance axioms.

Definition: Given M, $A \subseteq M$ and σ,

$$PFE^\sigma(A) \quad = \quad -\sum_{\alpha \in D(A)} \frac{\sigma(\alpha \vee \overline{\alpha})}{2}\, log_2\, \frac{\sigma(\alpha \vee \overline{\alpha})}{2} \quad \text{if} \quad A \neq \emptyset$$

$$0 \qquad \text{otherwise}$$

Note that for $A \neq \emptyset$ the definition is just one plus the usual Shannon entropy (to the base 2) but taken over the set of polarity free events $\alpha \vee \overline{\alpha}$. Another perhaps more natural interpretation of $PFE^{\sigma}(A)$ is as the Shannon entropy of a hypothetical distribution $h(\sigma_A)$ obtained from σ_A by, for every $\alpha \in D(A)$, redistributing the probability of each polarity free event $\alpha \vee \overline{\alpha}$ equally over the events α and $\overline{\alpha}$. Here one can think of the entropy of $h(\sigma_A)$ as being a measure of the uncertainty in σ_A if one "forgets", or discards as irrelevant, the available information about the *particular* manner in which the value of each $\sigma_A(\alpha \vee \overline{\alpha})$ is subdivided by σ_A between α and $\overline{\alpha}$.

It follows from the above definition that the polarity free expectation of the PFE^{σ} function, $\mathcal{E}_{PFE^{\sigma}}$, as defined in the previous section, is given by

$$\mathcal{E}_{PFE^{\sigma}}(A) = 1 - \frac{1}{2} \sum_{\beta \in D(A^c)} \sigma(\beta \vee \overline{\beta}) \sum_{\alpha \in D(A)} \frac{\sigma(\alpha\beta \vee \overline{\alpha}\overline{\beta})}{\sigma(\beta \vee \overline{\beta})} \, log_2 \, \frac{\sigma(\alpha\beta \vee \overline{\alpha}\overline{\beta})}{\sigma(\beta \vee \overline{\beta})}$$

for $A \neq \emptyset, M$, with $\mathcal{E}_{PFE^{\sigma}}(A) = PFE^{\sigma}(A)$ otherwise.

It is now straightforward to verify that

Theorem 4.1

The measure of gravitas PFE defined above satisfies the Continuity, Locality, Voter Renaming, Monotonicity, Clone, Polarity Free, Classical Margin, and Polarity Free Margin Invariance axioms.

In addition PFE has the following two properties:

(1) For any $A \subseteq M$ and any σ

$$\mathcal{E}_{PFE^{\sigma}}(A) = PFE^{\sigma}(M) - PFE^{\sigma}(A^c) + \delta_A$$

where $\delta_A = 0$ if $A = \emptyset$ or M, and $\delta_A = 1$ otherwise.

It is this equation, a translation by δ_A of the equation satisfied by the classical Shannon entropy, which ensures that the axiom of polarity free margin invariance holds: the result then follows just by writing the equations corresponding to $\mathcal{E}_{PFE^{\sigma}}(A)$ and $\mathcal{E}_{PFE^{\sigma}}(A^c)$ and subtracting.

(2) In the special case of the uniform distribution, it is trivial to verify that PFE satisfies a strong form of the classical margin axiom, namely

$$PFE^{unif}(A) = |A|$$

We may remark here that although any translation of PFE^σ by the addition of a constant value K still satisfies the eight axioms of the previous section, if we also we require the property (2) above to be satisfied, then no nontrivial such translation is possible.

We should also note that if A is a singleton then $PFE^\sigma(A) = 1$ for any σ.

It is also worth remarking that if we define $G^\sigma(A)$ trivially to be $|A|$ for all σ, then this function satisfies all the axioms of the previous section except the clone axiom. It might therefore at first sight be concluded that the axioms are rather weak. This however is not at all the case: in the presence of the clone axiom the remaining axioms, and especially the polarity free margin invariance axiom, take on a far stronger meaning.

5 Conclusions and Open Problems

Much further research is necessary to elucidate the foundations of a theory of gravitas, together with the gravitas majority (or supermajority) decision rules which can be derived from the concept. The axioms suggested, in particular those involving the notion of polarity freeness, are by no means unchallengeable. In fact these axioms emerged because the author started by investigating a simpler notion of gravitas, consisting simply of the usual Shannon entropy of σ_A, namely $\sum_{\alpha \in D(A)} -\sigma(\alpha)\ log_2\ \sigma(\alpha)$. This definition has many pleasant properties and satisfies all the axioms given in section 3 above except the polarity free axiom and the polarity free margin invariance axiom . However this last axiom should not really be counted as a failure since Shannon entropy *does* satisfy the simpler margin invariance axiom which can be formulated in the absence of the "polarity free" requirement, by replacing $\mathcal{E}_{G^\sigma}$ by the conditional Shannon entropy E_{G^σ}, defined by

$$E_{G^\sigma}(A) = \sum_{\beta \in D(A^c)} \sigma(\beta)\ G^{\sigma|\beta}(A)$$

where $\sigma|\beta$ denotes the conditionalisation of σ_A to the event β.

With this change we get the axiom of

Margin Invariance

For every $A \subseteq M$, $\qquad Mar_{E_{G^\sigma}}(A) = Mar_{G^\sigma}(A)$

In the case when G is interpreted as Shannon entropy E_{G^σ} is simply the usual conditional entropy of A given A^c; then the analogous property to **(1)** of the previous section holds (i.e. with the constant term δ_A deleted), which immediately implies that the above axiom of margin invariance is satisfied. Furthermore Shannon entropy also possesses an at first sight attractive property which is not possessed by PFE: namely it is additive for the union of two disjoint sets of voters A and B in the case when the probability distributions over A and over B are independent of each other. Nevertheless Shannon entropy possesses some difficult counterintuitive properties as a measure of gravitas. We can see this by looking at the example of a singleton A. Here the Shannon entropy varies between 0 and 1 depending on how close to $\frac{1}{2}$ is the probability that the unique member of A votes yes. This does not seem to make much sense as a measure of gravitas: in a two person committee we would surely not prefer, in the absence of other information, the judgement of a voter whose previous record indicated she was equally likely to vote yes or no, against that of a voter who previously almost always voted no, but on this particular occasion voted yes! This example is not really a problem for PFE however since, as noted above, PFE gives each individual voter an equal gravitas of 1.

Another reason to distrust Shannon entropy as a measure of gravitas is that the Shannon entropy of A satisfies a very strong symmetry property which we may call

Division Renaming

$G^\sigma(M)$ is invariant under any permutation of $D(M)$ if the probability distribution σ is adjusted to reflect the permutation[12].

For ease of notation we have stated the property for M, but it is clear that in the case when G is Shannon entropy (or in the presence of the Locality axiom) we could replace M by an arbitrary subset A. This property is far stronger than voter renaming: in a sense it makes the voters almost redundant to the calculation of G^σ since the divisions now all acquire identical status as abstract objects, and their original relationship to the voters of A appears to be irrelevant. This does not seem to have any intuitive justification as far as the notion of gravitas is concerned; furthermore it is inconsistent with the polarity free

[12] i.e. if π is a permutation of $D(M)$ and σ^π is defined by $\sigma^\pi(\pi(\alpha)) = \sigma(\alpha)$ for each $\alpha \in D(M)$, then $G^{\sigma^\pi}(M) = G^\sigma(M)$. Note that whereas a permutation of M always induces a permutation of $D(M)$ the converse is not true.

axiom, modulo only the trivial requirement that $G^\sigma(M)$ be not independent of variations of σ.

Turning to the rule R_{G^σ} discussed briefly in section 2 as a motivation for the study of gravitas, it is worth noting that in a dynamic context where $\sigma(t)$ is changing over time, the clone and continuity axioms appear to ensure that such a rule would have the effect of strongly discouraging the formation of factions, by penalizing the voting power (or success) of any such faction: over time this would occur quite irrespective of whether the factions existed as formal entities.

To analyse this claim further let us define the *success rate* of a voter as the probability that that voter will belong to the winning camp when a new motion is presented and a vote is taken[13]. One of the most intractable problems arising from the use of conventional static voting rules is the fact that such voting rules are prone to instabilities caused by the increase in success rates which a set of voters may achieve individually by forming a faction which votes as a block, using an internal voting rule to decide which way all the members of that faction vote. The formation of one such faction in turn encourages the formation of other factions in a process which is inherently unstable unless one faction is large enough to constitute by itself a winning coalition, i.e a majority dictatorship[14]. More serious than the instabilities however is the fact that voters are no longer voting honestly on the individual motions presented to them. In a political context the factions formed in the above manner are often given a formal status and called parties; we are now culturally so accustomed to this phenomenon that, even though the negative effects of party discipline on the discourse of politicians are well recognised, the existence of parties is regarded as an intrinsic part, or even a *sine qua non*, of the process of democratic decision making. Yet the possibility of a dynamic voting rule such as $R_{G^{\sigma(t)}}$ indicates that the phenomenon of dishonest voting can at least be actively discouraged. For if $R_{G^{\sigma(t)}}$ is taken to be the voting rule then as soon as a faction has formed for long enough for the effects of block voting to be partially reflected in the probability distribution $\sigma(t)$, the gravitas of any set of voters including that faction as a subset would tend to decrease; hence it seems likely that the formation of a faction would result in at least some members of that faction suffering a decreased success rate shortly after its formation, thus undermining the *raison d'être* of the faction. This can be seen as encouraging honest voting, and as a strong disincentive to the formation of factions. While this positive effect seems intuitively clear for a gravitas majority voting rule, rigorous mathematical results along these lines are likely to be hard both to formulate and to prove.

In the light of the above, it is interesting to consider Rousseau's observations concerning the problems arising from the formation of factions in a political

[13]see Laruelle, Martinez, and Valenciano [06] for an analysis of the concept of success in the context of voting systems.

[14]Effects of this kind have been studied in a number of recent papers on voting power; see Felsenthal and Machover [02] and [06] and Gelman [03].

context. According to Rousseau's notoriously ill-defined, but sometimes unfairly maligned, intuitive concept of "general will", the general will is always correct, but may well be at variance with the vote of the majority[15]. In Rousseau's conception the general will cannot be directly accessed, but while the opinion of the majority provides an indication of the general will, presumably in some probabilistic sense, it can be "mistaken". Reasons given by Rousseau as to why such "errors" can occur include insufficient or incorrect information available for the formation of judgments, and especially distortions caused by the formation of factions:

> "If, when the people, being furnished with adequate information, held its deliberations, the citizens had no communication one with another, the grand total of the small differences would always give the general will, and the decision would always be good. But when factions arise, and partial associations are formed at the expense of the great association, the will of each of these associations becomes general in relation to its members, while it remains particular in relation to the State: it may then be said that there are no longer as many votes as there are men, but only as many as there are associations. The differences become less numerous and give a less general result. Lastly, when one of these associations is so great as to prevail over all the rest, the result is no longer a sum of small differences, but a single difference; in this case there is no longer a general will, and the opinion which prevails is purely particular. It is therefore essential, if the general will is to be able to express itself, that there should be no partial society within the State, and that each citizen should think only his own thoughts..."[16]

There is in this quotation from Rousseau a serious conceptual problem which arises from the first sentence, and which has been much commented upon. It is difficult to understand how people could be both well-informed and have thoroughly considered a question, if they have no communication with one another[17]. It seems here as if Rousseau is grappling with an intractable difficulty, because while he desires an informed people who have fully deliberated the questions to be decided, he is painfully aware of the negative effects on the voting outcome which may be caused by factional substructures. However we can now see that the use of a gravitas majority rule could well cut through the difficulty which Rousseau was facing; under a gravitas majority rule, the influence of factions is likely to evaporate soon after they start to be formed, and the outcome of the voting rule can once again be considered, figuratively, the result of weighing "sums of small differences" of independent opinions, even though there is nothing like a bijective correspondence between independent opinions and individual voters.

[15] see Rousseau [62].

[16] Rousseau, ibid. Book 2 ch. 3

[17] The original French text of the first sentence, which is difficult to translate exactly, is as follows: "Si, quand le peuple suffisamment informé délibère, les citoyens n'avaient aucune communication entre eux, du grand nombre de petites différences résulterait toujours la volonté générale, et la délibération serait toujours bonne."

If indeed some sense can be made of Rousseau's idea of the general will, a question concerning which this author takes no position, then the notion of gravitas margin seems likely to provide a far more plausible indication of the general will than that provided by the classical margin. Indeed if the general will is interpreted as the limit of a probabilistic notion, then it may well be possible, using the notion of gravitas, to give the general will a more precise sense, which would be reasonably faithful to Rousseau's underlying idea.

For many reasons the ideas put forward in the present paper are likely to be met with a certain scepticism; hence it seems appropriate to end this paper by posing two precise philosophical challenges to those who would question whether a notion of gravitas majority can serve any useful purpose in the context of human systems of collective choice:

1. In defining a decision rule for collective choice is there some convincing philosophical principle which would exclude as unreasonable the use of additional information about the abstract relationships between voters' previous choices such as that encoded by σ?
2. Suppose that it could be demonstrated by empirical methods as suggested in section 2 that a particular gravitas majority rule R_{G^σ} performed systematically better than the simple majority rule when applied in contexts in which the correctness of M's decisions could be compared with an independently accessible objective truth. To what extent would such empirical evidence validate the use of the rule R_{G^σ} in contexts (a) where there exists an independent standard of objective truth or correctness which is not in general accessible, or (b) where there exists no independent standard of objective truth or correctness?

In conclusion I wish to express my thanks to Moshe Machover, Alena Vencovska, Hykel Hosni, Luc Bovens and Greg Holland for helpful comments on earlier versions of some of the ideas presented in this paper.

School of Mathematics
University of Manchester
Manchester M13 9PL

6 References

Arrow, K.J.[63] Social Choice and Individual Values, Yale University Press, 1963

Boland,P.J. [89] Majority Systems and the Condorcet Jury Theorem, The Statistician 38 (1989)

Cohen, J. [86] An epistemic conception of democracy, Ethics 97(1), pp. 26-38

Coleman, J., and Ferejohn, J. [86] Democracy and social choice, Ethics 97(1)

Condorcet, Marquis de, [85] Essai sur l'application de l'analyse la probabilité des décisions rendues la pluralité des voix, Paris, de l'imprimerie royale, 1785

Dummett,M [84] Voting Procedures, Oxford University Press, 1984

Estlund,D.M.., Waldron, J., Grofman, B., Feld, S.L.[89] Democratic Theory and Public Interest: Condorcet and Rousseau Revisited, American Political Science Review, 83(4), Dec. 1989

Felsenthal, D.S. and Machover, M. [02], Annexations and alliances: when are blocs advantageous a priori? Social Choice and Welfare, 19: 295 312, 2002

Felsenthal, D.S. and Machover, M. [06], Further Reflections on the Expediency and Stability of Alliances, preprint, London, Revised September 2006, http://eprints.lse.ac.uk/2566/

Gelman, A. [03], Forming voting blocs and coalitions as a prisoners dilemma: a possible theoretical explanation for political instability, Contributions to Economic Analysis and Policy: 2(1), Article 13, 114, 2003

Grofman, Bernard N., Guillermo Owen and Scott L. Feld. [83]. Thirteen theorems in search of the truth. Theory and Decision, 15

Grofman, B., and Feld S., [88], " Rousseau's General Will : A Condorcetian Perspective ", American Political Science Review, 82, 567-576.

Ladha, Krishna K.,[93] Condorcet's Jury Theorem in light of de Finetti's Theorem, Soc. Choice Welfare 1993(10)

Laruelle, A., Martinez, R., and Valenciano, F. [06] “Success Versus Decisiveness Conceptual Discussion and Case Study”, J. of Theoretical Politics, vol. 18 no.2, 185-205, 2006

List,C. [04] On the Significance of the Absolute Margin, Brit. J. Phil. Sci. 55, 2004

List, C., and Goodin, R. E. [01] Epistemic Democracy: Generalizing the Condorcet Jury Theorem, Journal of Political Philosophy 9(3)

Paris J.B. [94] The Uncertain Reasoner’s Companion, Cambridge University Press, 1994

Rousseau J.J. [62] Du Contrat Social, 1762, text available online at http://abu.cnam.fr/BIB/index.html, English translation by G.D.H.Cole available at http://www.constitution.org/jjr/socon.htm

Shannon C.E,.and Weaver,W. [64] The Mathematical Theory of Communication, University of Illinois Press, 1964