Collective Choice as Information Theory: Towards a Theory of Gravitas. What Can We Learn From A Council Of Elders?*

by

George Wilmers

August 2007

*The author wishes to express his thanks to Moshe Machover, Alena Vencovska, and Greg Holland for helpful comments on earlier versions of some ideas presented here.

1 Motivation: A Thought Experiment

As a motivating thought experiment 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. Let us imagine that M is considering a particular proposal. If the chief is wise then he will listen carefully to the advice he is given by the elders on the proposal; 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 judgements about the matters under deliberation both now and previously, and also to efface all his personal judgements about the value of the previous judgements of other elders. However if the chief is to eliminate all such personal judgements, 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 proposal against the weight of opinion of those who are against the proposal. How are these weights of opinion to be measured? Our chief could of course simply count up those in favour and those against the proposal, and compare the resulting cardinalities, as the leaders of the great western democracies would surely enjoin him to do[1], 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 proposals. So, in order to ensure that his approach is truly objective, the chief decides to erase from his memory all details about actual content 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' opinions for or against all previous proposals. The chief's mathematical problem as to how to extract weights of opinion from this information is essentially now our problem.

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

2 The Notion of Gravitas: A Preliminary Discussion

Our formal starting point is a fixed assembly M of n voters in binary choice context, endowed with a probability distribution σ on the set of possible of divisions $D(M)$ of M. We may think of σ as derived by some statistical rules from the evidence of previous voting records. We shall not concerned ourselves here with exactly what statistical procedures are used to derive such an *a posteriori* probability distribution but will take it as given. Thus we start with a mathematical idealization of the problem in the previous section 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.

The question we shall address is from our point of view a question which should be asked *prior* to a discussion of the correct interpretation of voting power in the context above. The question is 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. Is there some natural *measure* which we can define in order to compare the "weight of independent opinion" of A with that of A^c ?

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 which generalise the special classical situation in which σ is taken *a priori* to be the uniform distribution, where the voters are *a priori* considered to vote independently. 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 would certainly not be easy to find an alternative measure to PFE which satisfies the given axioms for gravitas, 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 1 in α is A,

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

In the opinion of the author, R_{G^σ} is an appropriate generalization of the simple majority rule for the case in which the information contained in σ is available. It may be considered to be a realization of the concept of rule by weight of independent opinion.

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 [ ]. In the case of unicameral binary choice the former tradition, which we may loosely call the *epistemic* tradition[2] has been concerned primarily with the problem of examining the mathematical conditions under which a majority decision rule can be theoretically justified in the context where an objectively correct answer is assumed to exist[3], while the latter tradition is concerned 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 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 could 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, 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 information is taken to consist of σ together with the actual division of the voters on the given motion[4].

The general theory of voting is associated with probability theory in various 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 judgements of members of an

[2]see Cohen[86] for philosophical discussions of the concept of an epistemic justification of democracy

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

[4]It may be noted that the idea 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, if we make the reasonable assumption that all judgements 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, 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 judgements of jury members.

assembly is available, and in particular why that most powerful tool of mathematical reasoning under uncertainty, information theory[5], has been so strikingly absent from deliberations. There are probably two related reasons for this situation, both of 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 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*[6]. The other, related, reason is that, despite its rather weak theoretical justification from any standpoint, 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 judgements, is likely to meet with incredulity at best.

Nonetheless both Aristotle and Rousseau[62] recognized the fact that majority decisions could be "incorrect". According to Rousseau's notoriously ill-defined, but also sometimes unfairly maligned, intuitive concept of "general will", the general will is always correct, but might well be at variance with the vote of the majority. The notion of gravitas margin seems to the author to provide a far more accurate indicator of Rousseau's intuitive concept of general will than the classical margin. Indeed if the general will is understood as a probabilistic notion as discussed above then it may well be possible to give this notion a more precise sense, using the notion of gravitas, which corresponds closely to Rousseau's intuitive concept.

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

[6]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 judgements 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].

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.

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 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.

Invariance under Isomorphism

If π is a permutation of the voters of M which, given σ, induces the probability distribution σ^π on M, then for any $A \subseteq M$, if 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

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 decision rule R_{G^σ} is monotone.

Clone Axiom

For any $A \subseteq M$ and $a, b \in A$ 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 that voter adds nothing in terms of independent opinions.

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 $A \subseteq M$, $G^\sigma(A)$ should depend 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 calculation of $G^\sigma(A)$. Obviously this axiom represents a strengthening of the Locality Axiom which could have been included in it. However because of its 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.

Clearly σ^*_A contains less information than σ_A. However the information which it contains has an interesting epistemological status. The probability distribution σ_A encapsulates information about the directions (for or against motions) in which voters in A voted. On the other hand, in the presence of some notion of correctness or truth, there exists another completely analogous probability distribution,τ_A, on the subsets of A, such that for each subset B of A, $\tau_A(B)$ records the probability that the members of B voted correctly and the members of $A - B$ incorrectly. Each of the probability distributions σ_A and τ_A would on its own induce the same probability distribution σ^*_A, so that the latter may in some sense be thought of as encapsulating the information which is common to both σ_A and τ_A. There is an interesting informational symmetry here; where τ_A is not given to us, we might expect that σ^*_A should encapsulate the most appropriate information to estimate it.

Our final 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 let $Mar_{G^\sigma}(A)$ denote the *G-margin* of A in M, i.e. the quantity

$G^\sigma(A)$ - $G^\sigma(A^c)$

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

Classical Margin Axiom

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

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

Our final axiom also concerns an invariance property of the margin which also relates to our initial observations on the Locality Axiom. Although we have insisted by the Locality and Polarity Free axioms that $G^\sigma(A)$ be a property of σ^*_A, if we then take G and consider its expected value on A with σ^*_A conditioned upon the "polarity free" information from every possible way in which the members of M not in A could divide, then we obtain a non-locally defined quantity, which we 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. For reasons of notational awkwardness we avoid giving the mathematical general definition of $\mathcal{E}_{G^\sigma}(A)$ here, but will instead illustrate it with the particular definition in the case when G is taken to be our tentative solution, the PFE function defined in the next section. This will clarify the content of the general definition.

Despite the nonlocality of its definition $\mathcal{E}_{G^\sigma}$ has an excellent claim to be considered as another candidate for a measure of the intuitive notion of gravitas. 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 powerful axiom is in fact satisfiable:

Margin Invariance

For every $A \subseteq M$,

$Mar_{\mathcal{E}_{G^\sigma}}(A) = Mar_{G^\sigma}(A)$

4 Polarity Free Entropy (PFE)

In this section we define a measure, Polarity Free Entropy, or PFE, which satisfies all seven axioms for a notion of gravitas, G, described in the previous section.

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

$$PFE^{\sigma}(A) = -\frac{1}{2} \sum_{\alpha \in D(A)} \sigma(\alpha \vee \overline{\alpha}) \; log_2 \; \sigma(\alpha \vee \overline{\alpha})$$

Note that the definition is just the usual Shannon entropy (to the base 2) but taken over the set of polarity free events $\alpha \vee \overline{\alpha}$. The factor of $\frac{1}{2}$ is present because otherwise because of the notation each event $\alpha \vee \overline{\alpha}$ would be counted twice.

The definition of the polarity free expectation of PFE^{σ} function, $\mathcal{E}_{PFE^{\sigma}}$ is given by

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

It is now straightforward to verify that the PFE^{σ} function satisfies all seven axioms of the previous section. In addition it has the following properties:

(1) For any $A \subseteq M$ and any σ

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

It is this fact, typical of entropy type functions, with ensures that margin invariance holds.

(2) In the special case of the uniform distribution, for A non-empty,

$$PFE^{unif}(A) = card(A) - 1$$

In particular if A is a singleton then $PFE^{\sigma}(A)$ is always zero. This makes sense if we think of gravitas as a property emerging from the *collectivity* A. A singleton has no *inherent* collective independence from itself.

5 Final Remarks

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. This definition has many pleasant properties and satisfies all the axioms given in section 3 above except the polarity free axiom and the margin invariance axiom in the form given above. However margin invariance is not in fact a failure here since Shannon entropy does satisfy the (simpler) form of margin invariance which can be formulated in the absence of the "polarity free" requirement. Furthermore Shannon entropy does possess 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 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 a priori to side with a voter whose previous record indicated he was equally likely to vote yes or no, against a voter who previously almost always voted no, but on this particular occasion voted yes! This particular example is not really a problem for PFE however since PFE gives both voters an equal gravitas of zero.

As regards the rule R_{G^σ} discussed briefly in section 2 as a motivation for the study of gravitas, we would note that in a dynamic context where σ is changing over time, such a rule would have the effect of strongly discouraging the formation of factions by penalising the voting power (or success) of any such faction: over time this would occur quite irrespective of whether the factions existed as formal entities. This can be seen as a strong incentive to encourage honest voting . While this effect seems intuitively clear, rigorous mathematical results along these lines are likely to be hard both to formulate and to prove.

6 References

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 probabilité 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

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)

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 JJ [62] Le Contrat Social, 1762

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