MASTERCLASS

Rational Consensus in Science and Society

(Lehrer and Wagner, 1981)

Department of Philosophy, Logic, and
Scientific Method

The London School of Economics
and Political Science

20 October 2009

First Session (11.30 – 13.00)

THE FRENCH – DE GROOT – LEHRER

MODEL OF CONSENSUS

Carl Wagner
Department of Mathematics
The University of Tennessee

1. Averaging functions

The simplest average of a sequence $x_1,\ldots,x_n$ of real numbers is their *arithmetic mean*,

$$A(x_1,\ldots, x_n) := (x_1 + \cdots + x_n)/n.$$

The arithmetic mean is just one of an infinite number of *quasi-arithmetic means*, defined for each strictly monotonic function φ, by

$$A_\varphi(x_1,\ldots, x_n) := \varphi^{-1}[(\varphi(x_1) + \cdots + \varphi(x_n))/n].$$

When $\varphi(x) = \log(x)$, we get the *geometric mean*,

$$G(x_1,\ldots, x_n) := (x_1 \cdots x_n)^{1/n}.$$

When $\varphi(x) = 1/x$, we get the *harmonic mean*,

$$H(x_1,\ldots, x_n) := [(x_1^{-1} + \cdots + x_n^{-1})/n]^{-1}.$$

When $\varphi(x) = x^2$, we get the *root-mean-square*

$$RMS(x_1,\ldots, x_n) := [(x_1^2 + \cdots + x_n^2)/n]^{1/2}.$$

An even broader class of averaging functions is furnished by the *weighted quasi-arithmetic means,*

$$\varphi^{-1}[w_1\varphi(x_1) + \cdots + w_n\varphi(x_n)],$$

where $w_1,\ldots,w_n$ is a sequence of nonnegative real numbers and $w_1 + \cdots + w_n = 1$ and φ is any strictly monotonic function.

Initially, we will restrict attention to averaging functions from the class of *weighted arithmetic means,*

$$w_1x_1 + \cdots + w_nx_n .$$

Other possibilities will be discussed later in this session, or in session two.

2. A Formal Theory of Social Power

J.R.P. French (1956), *Psychological Review,* Vol. 63, No.3, pp. 181- 194.

- A group of n individuals, each with an opinion regarding the most appropriate value of some numerical decision variable. Their individual assessments at time t = 0 are recorded in an n x 1 column matrix $A^{(0)} = (a_1^{(0)},\ldots, a_n^{(0)})^{Tr}$, where $a_i^{(0)}$ denotes the initial assessment of individual i.

- Over time, individuals revise their assessments, with the column matrix $A^{(t)} = (a_1^{(t)},\ldots, a_n^{(t)})^{Tr}$ recording their assessments at time t, for t = 0,1,2,….

- French models the transition from $A^{(t)}$ to $A^{(t+1)}$ by the matrix equation

(2.1) $$A^{(t+1)} = WA^{(t)},$$

where $W = (w_{ij})$ is a fixed n x n *weight matrix*, that is,

(i) All entries of W are nonnegative real numbers ; and

(ii) All row sums of W are equal to 1, i.e.,

(2.2) $w_{i1} + w_{i2} + \cdots + w_{in} = 1$ for $i = 1,\ldots,n$.

Remark. Weight matrices occur in the theory of Markov chains, where they are termed *stochastic matrices*, and the quantities w_{ij} represent certain conditional probabilities.

● When $i \neq j$, the weight w_{ij} represents the power that individual j can bring to bear on individual i (resulting in i's giving weight w_{ij} to j's opinion).

● Similarly, w_{ii} represents i's power to resist the influence of other individuals in the group group(resulting in i's giving weight w_{ii} to his/her own opinion).

● The power parameters w_{ij} are conceived of as enduring over time, and manifest themselves by modifying individual i's assessment $a_i^{(t)}$ at time t to

(2.3) $a_i^{(t+1)} = w_{i1}a_1^{(t)} + w_{i2}a_2^{(t)} + \cdots + w_{in}a_n^{(t)}$

at time t+1, i.e., to a certain weighted arithmetic mean (with weights coming from the i^{th} row of W) of the n assessments recorded in $A^{(t)}$.

● If $m_t := \min \{ a_i^{(t)} : i = 1,\ldots,n \}$ and

$M_t := \max \{ a_i^{(t)} : i = 1,\ldots,n \}$, then

$$m_t \leq a_i^{(t+1)} \leq M_t , \quad i = 1,\ldots,n.$$

● But the standard deviation of the assessments $\{ a_i^{(t+1)} : i = 1,\ldots,n \}$ can exceed the standard deviation of the prior assessments $\{ a_i^{(t)} : i=1,\ldots,n\}$.

● If the rows of W are identical, then the entries of the column matrix $A^{(1)} = WA^{(0)}$, are identical, i.e., *unanimous assessments are attained at* t = 1. This can occur even if the rows of W are not identical:

$$W \quad X \quad A^{(0)} = A^{(1)}$$

$$\begin{matrix} ½ & 0 & ½ \\ ¼ & ½ & ¼ \\ ½ & 0 & ½ \end{matrix} \quad X \quad \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} = \begin{matrix} 2 \\ 2 \\ 2 \end{matrix} .$$

● Even if no $A^{(t)}$ has identical entries, it may be the case that

$$\text{(2.4)} \qquad \lim_{t \to \infty} A^{(t)} = A,$$

where A is an n x 1 column matrix with identical entries (*convergence to unanimity*).

- Since $A^{(t)} = W^t A^{(0)}$, convergence to unanimity is ensured if

(2.5) $$\lim_{t\to\infty} W^t = L,$$

where L is an n x n weight matrix with identical rows, in which case,

(2.6) $$\lim_{t\to\infty} A^{(t)} = LA^{(0)}.$$

French offers only elementary observations about convergence conditions. The subject is treated in somewhat more detail in F. Harary (1959, *A criterion for unanimity in French's theory of social power*, in Studies in Social Power, D. Cartwright, ed., Institute for Social Research, Ann Arbor, Mich.,pp. 168-182), but Harary appears to have misinterpreted one of the results in classical Markov chain theory in applying it to the consensus problem.

Summary: French's model is *descriptive* (though highly idealized) and *diachronic*, with repeated episodes of resistance and acquiescence driven by power relationships that are assumed to endure over time. And no justification is furnished for the use of weighted *arithmetic* averaging.

Apparently unaware of French's work, DeGroot proposed in 1974 a normative, synchronic model of consensus based on the very same mathematics:

3 . DeGroot's Normative Model of Consensus

(1974, *Reaching a Consensus*, Journal of the American Statistical Association 69, pp.118-121)

- Suppose that n individuals separately assess probability measures $p_1^{(0)},\ldots,p_n^{(0)}$ on a sigma algebra **A**, and they wish to aggregate these measures to produce a single, consensual measure. DeGroot proposed the following

method for attempting to arrive at such a consensus:

Let each individual i assign weight w_{ij} to individual j, *based on* i*'s assessment of* j*'s expertise relative to other members of the group*. The weight matrix $W = (w_{ij})$ is identical in form (though not interpretation) with the weight matrices employed in French's model. It is assumed that *weights are assigned before individuals are apprised of the probability measures of their colleagues.*

With $P^{(0)}$ = the column vector $(p_1^{(0)},\ldots, p_n^{(0)})^{Tr}$ and $P^{(t+1)} = WP^{(t)}$, t = 0,1,…, it again follows that convergence of powers of W to a weight matrix L with identical rows $(\lambda_1,\ldots, \lambda_n)$ is sufficient for the convergence of the column vectors $P^{(t)} = W^t P^{(0)} = (p_1^{(t)},\ldots, p_n^{(t)})^{Tr}$ of probability measures to

(3.1) $LP^{(0)} = (p, p,\ldots, p)^{Tr}$, where

(3.2) $p = \lambda_1 p_1^{(0)} + \cdots + \lambda_n p_n^{(0)}$,

and DeGroot advocates that the group adopt p as its consensual probability measure.

● A necessary and sufficient condition for convergence of powers of W to a weight matrix with identical rows:

Theorem 3.1. (Doob 1953) Powers of W converge to a weight matrix L with identical rows $(\lambda_1,\ldots, \lambda_n)$ if and only if some power of W contains a column with *exclusively positive* entries. Moreover, the consensual weights $\lambda_1,\ldots, \lambda_n$ are the unique solution to the simultaneous linear equations

(3.3) $(x_1,\ldots, x_n)W = (x_1,\ldots, x_n)$

$$x_1 + \cdots + x_n = 1.$$

● Doob's convergence criterion is actually decidable:

Theorem 3.2. (Isaacson and Madsen 1974) Some power of the n x n weight matrix W has exclusively positive entries if and only if some column of $W^{(n-1)(n-2)+1}$ has exclusively positive entries.

Remarks on DeGroot's model

1. Apart from noting that weighted arithmetic means of probability measures are probability measures, with no need for subsequent normalization (as is necessary, for example, in the case of weighted geometric or harmonic means), DeGroot offers no justification for arithmetic averaging.

2. Doob's convergence criterion is left in purely mathematical form, with no attempt to furnish a salient interpretation.

3. Despite being indexed on a variable t, DeGroot's model is essentially synchronic, with the infinite process

$$P^{(0)} \rightarrow WP^{(0)} \rightarrow W^2P^{(0)} \rightarrow \cdots$$

conceived as a single operation.

4. The only justification offered for repeated multiplication by W is that the weights in W capture *general expertise*, and are thus appropriately employed in averaging *any probability assessments of members of the group*, not simply the initial assessments recorded in $P^{(0)}$. This seems like an insufficient response to the concern that averaging beyond the first stage involves an unjustified "double counting."

In *Rational Consensus in Science and Society*, Keith Lehrer and I attempted to deal with the apparent deficiencies in DeGroot's model, and also to generalize that model.

4. The L-W Model of Consensus

● Lehrer was originally unaware of the work of French and DeGroot, but devised a normative model of rational consensus identical in mathematical form to DeGroot's.

● Lehrer's interpretation of the weights differed from DeGroot's in a significant way. He always regarded the sequence $W, W^2, W^3, \ldots$ of matrix powers as a simplification of the more general sequence of matrix products

(4.1) $W_1,\ W_2W_1,\ W_3W_2W_1, \ldots,$

with weights in W_1 expressing individuals' assessments of the expertise of their colleagues as, say, physicists; in W_2 as judges of physicists; in W_3 as judges of judges of physicists, etc.

- If $\lim_{t\to\infty} W_t W_{t-1} \cdots W_1 \to L$,

a weight matrix with identical rows $\lambda_1,\ldots, \lambda_n$, then these consensual weights are to be used in in averaging individuals' initial assessments of whatever decision variables or probability measures are in question.

- In the LW approach, deliberation is regimented as follows:

1. Discussion of the most appropriate values of the decision variables is carried out by exchanging anonymous position papers. Individuals' assessments at "dialectical equilibrium" are registered in a matrix A.

2. If consensus fails in A , authors of the papers are revealed, and a discussion of the most appropriate weights to assign individuals as evaluators of the original decision variables (e.g., as physicists) is carried out by exchanging a second round of anonymous position papers.

3. Individuals' assessments of these first order weights at dialectical equilibrium are recorded in a matrix W_1. If consensus obtains in W_1A, the group adopts the entries of any of the identical rows of W_1A as consensual values of the initial decision variables. If consensus fails in W_1A (which implies that consensus fails in W_1), the authors of the second round of papers are revealed, and a discussion of the most appropriate weights to assign individuals as judges of physicists is carried out by exchanging a third round of anonymous position papers.

4. Individuals' assessments of these second order weights at dialectical equilibrium are recorded in a matrix W_2. If consensus obtains in W_2W_1, or in $(W_2W_1)A$, the group adopts the entries in any of the identical rows of $(W_2W_1)A$ as consensual values of the initial decision variables. If not, consensus must fail in W_2. Then the authors of the third round of papers are revealed, and a discussion of the most appropriate weights to assign individuals as

judges-of-judges-of physicists is carried out by exchanging a fourth round of anonymous position papers.

5. Individuals' assessments of these third order weights at dialectical equilibrium are recorded in a matrix W_3. If consensus obtains in W_3W_2, or in $(W_3W_2)W_1$, or in $((W_3W_2)W_1)A$, the group adopts any of the identical rows of $((W_3W_2)W_1)A$ as consensual values of the initial decision variables……

This deliberative protocol insures that higher order evaluations do not covertly influence lower order evaluations. So repeated averaging is insulated from the possibility of double counting.

- Remark. The above protocol is elaborated in C.Wagner, Consensus through respect: a model of rational group decision-making, *Philosophical Studies* 34 (1978), 335-349.

5. Convergence to Consensus

Case 1. $W_i \equiv W$, $i = 1,2,\ldots$

Let $W = (w_{ij})$.

● We say that i *respects* j if $w_{ij} > 0$.

● If $i_0, i_1,\ldots,i_r$ is any sequence of individuals in which i_k respects i_{k+1} for $k = 0,\ldots, r-1$, we say that there is a *chain of respect* (*of length* r) from i_0 to i_r .

Theorem 5.1. Let i and j be individuals. There is a chain of respect of length r from i to j if and only if the entry $w_{ij}^{(r)}$ in the i^{th} row and j^{th} column of the matrix W^r is positive.

Doob's convergence criterion (Theorem 3.1) can thus be reformulated as follows:

Theorem 5.2. Powers of W converge to a weight matrix L with identical rows $(\lambda_1,\ldots, \lambda_n)$ if and only if, for some $r \geq 1$, and for some individual j, there is a chain of respect of length r from every other individual to j, as well as from j to j, in which case the consensual weights are the unique solution $(\lambda_1,\ldots, \lambda_n)$ to the simultaneous linear equations

$$(x_1,\ldots, x_n)W = (x_1,\ldots, x_n)$$

$$x_1 + \cdots + x_n = 1.$$

● Theorem 5.2, with its insistence on the existence of chains of respect *of uniform length* is still not completely satisfactory. But it can easily be seen to imply the following more natural convergence condition:

Theorem 5.3. Let W be a weight matrix and let E be the set of all individuals j such that there is a chain of respect from every other individual to j. If (i) E is nonempty, and (ii) at least one individual j in E respects him/herself ($w_{jj} > 0$), then powers of W converge to consensus. The consensual weights are the unique solution $(\lambda_1,\ldots, \lambda_n)$ to the simultaneous linear equations

$$(x_1,\ldots, x_n)W = (x_1,\ldots, x_n)$$

$$x_1 + \cdots + x_n = 1,$$

and $\lambda_j > 0$ if and only if j belongs to the set E.

- What if (i) holds, but (ii) fails?

It seems reasonable that the pattern of respect captured in (i) should materialize in some sort of consensus, but (i) alone does not guarantee that powers of W converge to consensus. Example:

$$W = \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}.$$

The following theorem of Berman and Plemmons (*Nonnegative Matrices in the Mathematical Sciences*, Academic Press, 1979, p.244) suggests a way around this problem:

Theorem 5.4. Let W be a weight matrix and let E be the set of all individuals j such that there is a chain of respect from every *other* individual to j. The set E is nonempty if and only if there is a unique solution $(\lambda_1,\ldots, \lambda_n)$ to the simultaneous linear equations

(5.1) $(x_1,\ldots, x_n)W = (x_1,\ldots, x_n)$

(5.2) $x_1 + \cdots + x_n = 1.$

Claim: If there there are unique defensible consensual weights $(\lambda_1,\ldots, \lambda_n)$ implicit in W, they should satisfy (5.1).

Justification: If such weights fail to satisfy (5.1), then the weights $(\lambda_1^*,\ldots, \lambda_n^*) := (\lambda_1,\ldots, \lambda_n)W$ would compete with $(\lambda_1,\ldots, \lambda_n)$ as the proper weights for averaging assessments in $A^{(0)}$.

● The above argument avoids iterated averaging altogether and endorses the unique *fixed point weight vector* $(\lambda_1,\ldots, \lambda_n)$, if it exists, as the proper sequence of consensual weights implicit in W.

But here is an iterative justification, based on the convergence of powers of a weight matrix W_ε that is as "close" to W as we wish:

Theorem 5.5. Let W be a weight matrix and let E denote the set of all individuals k such that there is a chain of respect from every other individual to k. If E is nonempty, then either

(1) powers of W converge to consensus ; or

(2) for all ε in (0,1), powers of
$W_\varepsilon := \varepsilon I + (1 - \varepsilon)W$ converge to consensus independently of ε.

In each of the above cases the consensual weights constitute the unique fixed point weight vector of W, and individual i receives positive consensual weight if and only if i belongs to E.

Case 2. Possibly different weight matrices $W_1, W_2, \ldots$ at every level of evaluation.

● A sufficient condition for $W_n W_{n-1} \cdots W_1$ to converge to a weight matrix with identical rows as $n \to \infty$:

Theorem 5.6. (Chatterjee and Seneta 1977) Let μ_i denote the smallest element of W_i. If the infinite series $\sum_{i \geq 1} \mu_i$ diverges to infinity, then $W_n W_{n-1} \cdots W_1$ converges to a weight matrix with identical rows as $n \to \infty$.