Evaluating the Competence of Experts, Pooling Individual Judgements into a Collective Choice, and Delegating Decision Responsibility to Subgroups[1]

B. Grofman*, S. L. Feld** and G. Owen***

**School of Social Sciences, University of California, Irvine, Irvine, California 92717, USA*

***Department of Sociology, State University of New York at Stony Brook, Stony Brook, New York, USA*

****Department of Mathematics, University of the Andes, Bogota, Colombia*

ABSTRACT

We deal with situations where there can be, at least in principle, a "correct" answer (e.g. a parameter estimate), and where individuals (some of greater, some of lesser expertise) will be asked to make judgments. We show how majority group decisions vary with the competence of the individuals comprising the group, indicate ways to improve upon simple majority decision-making if the relative level of expertise of participants can be assessed in the group judgmental task. We also show how, under certain circumstances, delegation of parameter estimation to a subgroup operating under a unanimity or near unanimity rule, but with a tie-breaker provision for referral of conflicts to the larger group if unanimity is not obtained, can be superior in judgmental accuracy to direct simple majority decision-making in the larger group.

KEYWORDS

Group decision-making; judgmental accuracy; committees; majority rule; experts; information pooling.

"Ye shall know the truth, and the truth shall make you free".
New Testament, John, VIII, 32.

---

[1]An earlier version of this paper was delivered at the Fifth International Congress of Cybernetics and Systems, Mexico City, August 1981. This research was supported by Grant SES 80-07915, National Science Foundation, Political Science Program. The first portion of Section II of this paper is drawn from material in Grofman, Owen, and Feld (1982a; 1982b).

"The many, when taken individually, may be quite ordinary fellows, but when they meet together, they may well be found collectively better than the few". Aristotle, Politics, Book III.

I. INTRODUCTION

In many decision contexts, collective decision-making bodies will be faced with conflicts between the judgments of the members of the group and judgments of experts. In the first section of the paper, extensions of the Condorcet jury theorem (Condorcet, 1785) are developed to show under what circumstances decisions reached by majority vote among the members of a group are more likely to be correct than those of a given expert (or a panel of experts) of specified competence. When recommendations of experts differ, a method is proposed for combining opinions using a weighted voting rule which assigns weights proportional to the log odds of each individual's judgmental competence. This method, which is shown to be a direct extension of Bayes's Theorem, has the property of maximizing the likelihood that the group decision will be the "best" that can be reached given the available information input. When the competence of particular experts is unknown, a method for estimating competence is discussed which involves comparing the judgments of each one of a group of experts over a series of dichotomous choices with the decisions that would have been reached by a majority vote of the panel of experts as a whole.

In the second section of the paper, the problem of optimal delegation of decision-making is analyzed. Consider a group of size N reaching a decision on a series of complex issues by a majority vote of its members. For some set of problems under consideration, let p_i be the judgmental competence of the i^{th} member of the group (i.e. the probability that, in a dichotomous choice situation, the individual will correctly choose the "better" of two proposed solutions to the problem). We do not know individual p_i but may assume that $\bar{p}_i > 1/2$. Let the group decide between one of two options: Option (1): Decide all issues themselves (by majority vote). Option (2): Divide the group's membership into distinct subgroups (committees) of size k, each of which will be delegated decision responsibility for exactly r issues, but require those subgroups to use a supramajoritarian decision vote (e.g. subgroup unanimity); and furthermore, require that if a subgroup deadlocks, then some procedure be used to break the deadlock (e.g. the convening of a new committee to consider the problem de novo, as for example, a new jury trial being ordered if the first trial results in a hung jury).

In general, we expect that requiring supramajoritarian decisions will increase the probability of deadlock and thus lead to irresolute decision-making. However, using ideas of Bayesian sequential sampling, we show that for a range of values of $\bar{p}$ and N there exist small committees and optimal tie-breaking mechanisms which give rise to decisions at least as good or better than those which would be reached by a majority vote of the larger parent body, and we conjecture that for a wide range of $\bar{p}$ and N the expected number of individuals required before the committee decision process becomes resolute (known in the sampling literature as ASN, the "average sampling number") may still be less than the size of the parent group. In such cases, delegation is clearly the preferred choice, since it leads to decisions which are both resolute and accurate, while requiring fewer personnel. In other cases of interest (in particular for "very easy" decisions and "very hard" decisions), we believe that it can be shown that committee decisions made by committees with only a relative handful of members can be shown to be, with very high probability, nearly as good as decisions of a much larger parent body. Hence, if there are many decisions to be made and the transaction costs of empaneling the full group to consider them are large, while the cost in reduced "legitimacy" of decisions is relatively slight, delegation of decision-making responsibilities to a small task force may be a strongly preferred strategy.

In the concluding sections of the paper we offer advice to decision-makers on how the decision-making process can be better structured.

Notation

- p_i = judgment competence of the i^{th} voter ($0 < p_i < 1$) in a dichotomous choice situation, i.e. probability that the voter will make the correct choice (or the "better" choice) of the two available to him where each is assumed to be a priori equally likely to be the correct choice, and where "better" choice can be defined post hoc even if not known a priori
- N = number of voters in the group (for simplicity, N will generally be taken to be odd)
- m = a majority = $(N+1)/2$ for N odd
- $\bar{p}$ = average judgmental competence of voters in the group
- p = judgmental competence of a voter in a homogeneous group
- p_N = probability that at least a majority of voters will make the correct choice in a dichotomous choice situation, where N is the number of voters in the group
- w_i = weighted vote of the i^{th} voter in a group using a weighted voting rule (for simplicity, we set $\Sigma w_i = 1$)
- p_H = in a group vote, the probability that a group will be unable to reach a decision

II. POOLING OF JUDGMENTS

The basic theorem on the effect of pooling of judgments is due to the French mathematician and philosopher, the Marquis de Condorcet (1785). This theorem, which can be thought of as a variant of the well-known "law of large numbers", was, however, "lost" for a number of years until rediscovered by Black (1958). (For a history of the theorem, see Grofman, 1975). The Condorcet Theorem in its original form assumes that jurors are homogeneous, i.e. $p_i = p_j$ for all i, j. Hence, for homogeneous jurors, we shall drop the subscript.

Theorem I (Condorcet Jury Theorem): If $1 > p > 1/2$, then p_N is monotonically increasing in N and $\lim_{N\to\infty} p_N \to 1$; if $0 < p < 1/2$, p_N is monotonically decreasing in N and $\lim_{N\to\infty} p_N \to 0$; while if $p = 1/2$, $p_N = 1/2$ for all N.

For a proof see Black (1958) or Grofman (1978). If $p > 1/2$, this theorem can be interpreted as "vox populi, vox dei". It is rather remarkable how fast p_N goes up (down) with N if $p > 1/2$ ($p < 1/2$).
We show results for N=1 through N=19 in Table 1.

TABLE 1 The Probability that a Majority of Jurors Will Reach a Correct Verdict for Various Values of N and p

N	p = .2	.4	.5	.6	.8
1	.2000	.4000	.5000	.6000	.8000
3	.1040	.3520	.5000	.6480	.8960
5	.0580	.3174	.5000	.6826	.9420
7	.0335	.2858	.5000	.7102	.9666
9	.0196	.2666	.5000	.7334	.9804
11	.0116	.2466	.5000	.7534	.9884
13	.0070	.2288	.5000	.7712	.9930
15	.0042	.2132	.5000	.7868	.9958
17	.0026	.1990	.5000	.8010	.9974
19	.0016	.1860	.5000	.8140	.9984

N = group size, p = the probability that an individual member of the group will reach a correct judgment. Source: Grofman (1975).

Grofman (1978) generalizes this theorem to the case where the p_i are normally distributed with a variance equal to the binomial variance, with $\bar{p}$ replacing p in the expressions above. The present authors have further generalized the Condorcet theorem to deal with any distribution of p_i values, where $\bar{p}$ is fixed. Algorithms for optimizing and counteroptimizing the distribution of competence subject to a fixed $\bar{p}$ constraint are given in Owen, Grofman and Feld (1981) and Grofman, Owen, and Feld (1982a; 1982b).

Grofman (1975; 1978) has looked at the question of whether the group accuracy of a large group of not-so-smart people will be higher or lower than the p_N value for a smaller "blue-ribbon" group. Grofman (1975) provides a closed form expression, which gives us a way of expressing trade-offs between p (or $\bar{p}$) and N in the form of "isocompetence curves". (See Fig. 1). Grofman (1975) also demonstrates the paradoxical result that it is sometimes possible to raise p_N by adding members to a group who actually lower the average $\bar{p}$; the increase in N compensates for the decrease in $\bar{p}$. This idea has been further examined in Margolis (1976), who asserts that this paradox can only occur if the average competence of the added members is above .5. Grofman (1978) looks at the question of whether the pooled group judgment can be expected to be more accurate than that of its best member. It is sometimes asserted that this can never occur (see e.g. Einhorn, Hogarth and Klempner, 1977: 168), but that is incorrect. The Grofman (1978) results on this question are, however, reported in the form of lengthy tables and do not lend themselves to ready summary: we refer the interested reader to the original source[2].

[2]For discussion of the expected distribution of the competence of the "best" member of the group, see Steiner and Rajaratnam (1961). For closely related models, see Lorge and Solomon (1959) and Steiner (1966).

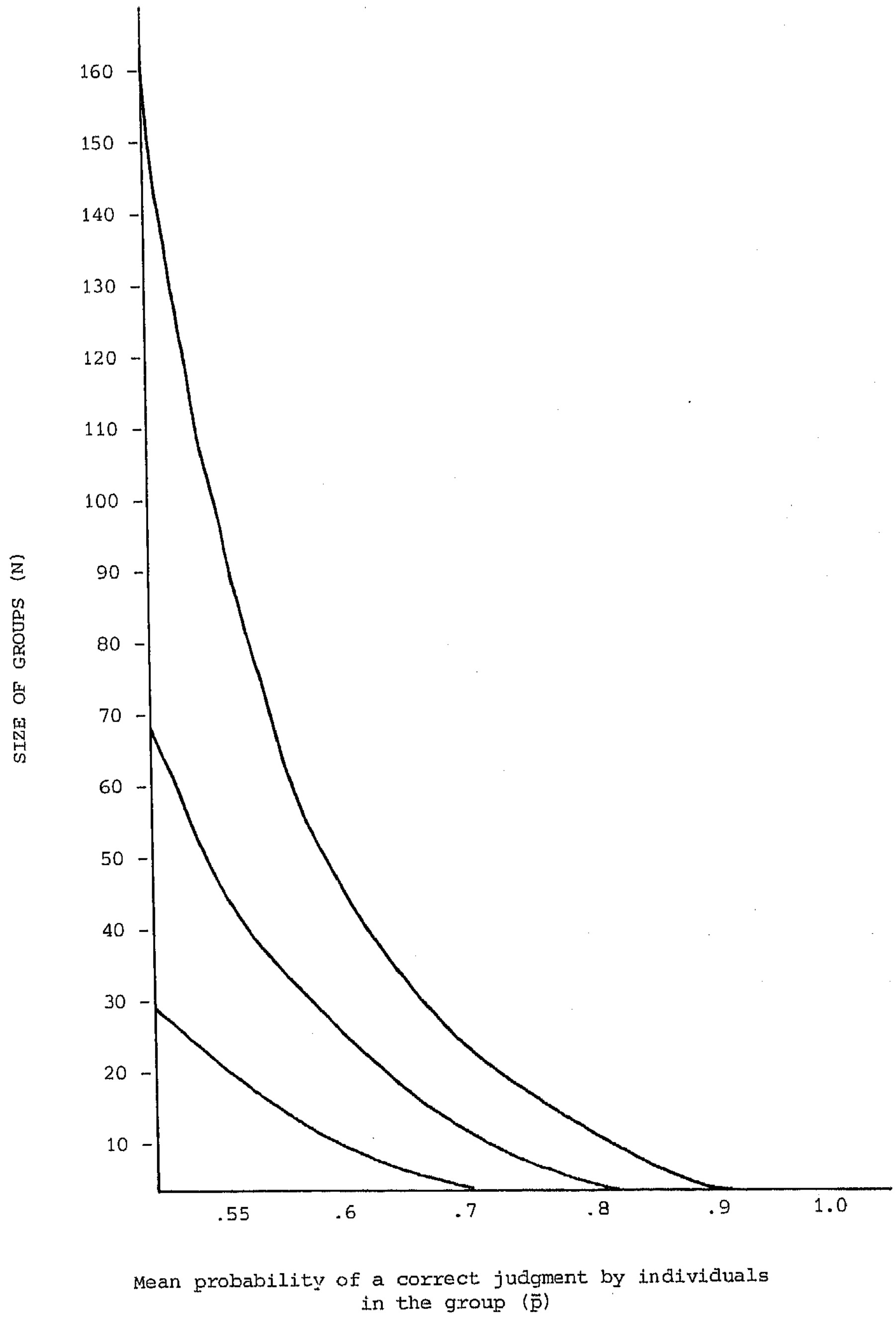

Fig. 1. Isocompetence curves for various values of mean group competence ($\bar{p}$) (Grofman, 1975)

We now turn to another result that tells us how to optimally assign weights to group members' inputs into the decision process when members are of unequal competence.

Theorem II (Corollary to Bayes' Theorem: The Bayesian Optimal Group Decision Rule)[3]: In a heterogeneous group, the decision rule which maximizes p_N is given by using a weighted voting rule where a majority of the weighted votes are required for passage and by assigning the weights, w_i, such that the w_i are proportional to

$$\log\left(\frac{p_i}{1-p_i}\right) \qquad (1)$$

Note that, once we pick a logarithmic base, the weight assignment we give to an individual is a function purely of his competence and is independent of the competence of the other members of the group. This result is a quite counterintuitive one. In the light of a proof of this theorem which shows it to be, in effect, a restatement of Bayes's Theorem, this result turns out to be equivalent to the well-known fact that the posterior Bayesian probability is independent of the order in which evidence is inputted.

Some examples of this theorem will be useful in showing its counterintuitive power. The first example is due to Grofman and provided the incentive for Shapley's derivation of the theorem. Consider a group with competences (.9, .9, .6, .6, .6). If we let the most competent members of the group decide, $p_N = .9$; if we let the group decide by majority rule, $p_N = .87$; but if we let the group decide using the weighted voting rule (1/3, 1/3, 1/9, 1/9, 1/9) then $p_N = .93$. This weighted voting rule is equivalent to giving the two most competent members of the group one vote each and letting the three least competent share a vote among themselves which is to be cast by a majority vote among the three of them.

If we look at a three-member group with competences (.55, .60, .70), then we may show that the optimal rule is to assign weights (0, 0, 1); on the other hand, if the competence of the first member is adjusted upward to obtain (.65, .60, .70), then the optimal voting rule is simple majority; i.e. improving the competence of one voter dramatically affects the power of all the voters in the group.

It might appear that Theorem II is of little value because it requires us to know the true competences of all group members or consulting experts. However, if we "examine" group members/experts by asking them a series of questions which pose the sort of problems that the group must solve (posed, of course, in the form of dichotomous choices), then even if we don't know the true answers to the questions, we can estimate the skill of the experts ! Of course, this requires assuming that the preponderance of expert opinion has a greater than 50 percent probability of being correct. Note that we do not require that the correct answer to the question be known. Rather we are

[3]This theorem has been proved by Shapley and by Grofman (see Shapley and Grofman, 1981), and independently by Nitzan and Paroush (1980a); see also Pierce (1961), Minsky and Papert (1971), and Duda and Hart (1973), which contain the theorem but in quite different substantive contexts, in which the theorem is presented as a result in pattern recognition theory or in automata theory. It appears in Grofman, Owen and Feld (1982a).

estimating the "correct" answer in much the same way one might estimate the "true" value of an unknown population parameter through use of sampling methods. Implicitly, we are requiring that there be some criterion by which the "correctness" of an answer can in the long run be judged, even if a priori we cannot tell which choice is the "better" or "best" one.

Theorem III (Correcting a True-False Exam Without an Answer Key): Let r_i be the proportion of time that an individual with competence p_i agrees with the majority verdict of a group of which he is a part. If $\bar{p}_i > .5$, first,

$$(r_i - .5) \text{ is proportional to } (p_i - .5) \quad (2)$$

and, if $.55 < p_i \leq .76$, then

$$p_i - .5 \text{ is proportional to } \log\left(\frac{p_i}{1-p_i}\right) \quad (3)$$

Hence, (a) we can approximate an individual's true competence on a task set by scoring the percentage of his agreement with the majority choices; and (b) we can approximate the optimal weights prescribed by Theorem II by scoring the percentage of agreement with the majority choices and assigning w_i proportional to $(r_i - .5)$.[4]

It should always be possible to set up such a mechanism for rating experts by designing the appropriate series of tests and comparing expert judgments. However, in any case, it remains important to determine how robust the optimal weights are, i.e. whether the accuracy of the group of experts is much affected by small modifications in the weights. If the accuracy is little affected, then it is likely that even a weighting system where the weights are informally estimated on the basis of subjective judgments of experts' competence can be highly effective; but if the accuracy is strongly affected by weightings, then subjective informal estimated weightings must be looked upon with suspicion.

To answer this type of question, we have developed an approximation to the accuracy of a group's majority decisions with given competences and weights. The basis for the approximation is the assumption that the distribution of outcomes tends towards normality for large enough numbers of individuals (or even for small numbers of individuals if their competences are symmetrically distributed). The mean proportion of correct votes is just the weighted mean of the competences, while the variance has two components: the variance in individual weights and the variance among the competences. To determine how often the group makes the correct decision, we determine how many standard deviations .50 is away from the mean proportion of correct votes. This we express in standard deviation units as a z score, z_N, and transform that standard score to a probability using a normal distribution table. (See Fig. 2).

[4]This theorem has been proved by Scott Feld (1979 unpublished) using a Taylor expansion around $p_i = .5$ to obtain the approximations given above. It appears in Grofman, Owen and Feld (1982a).

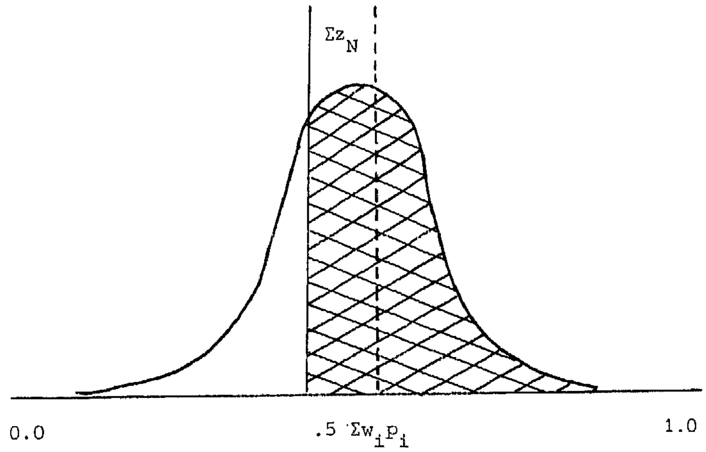

Fig. 2. Approximate distribution of proportion of correct votes with specified competences and weights for a group of size N.
(Shaded area indicates outcomes where the majority is correct.)

The larger the value of z_N the greater the probability that the group majority will reach the correct decisions.

$$z_N = \frac{\sum_{i=1}^{n} w_i p_i - .5}{\sqrt{\sum_{i=1}^{n} w_i^2 p_i (1-p_i) + \sum_{i=1}^{n} w_i^2 (p_i - \sum_{i=1}^{n} w_i p_i)^2}} \qquad (4)$$

Where the distribution of weights and competences is disparate, and N is small, one may wish to do an exact calculation of the probability that the group is accurate. The basic strategy is to list all possible combinations of votes, determine the proportion of weighted votes that are correct for each combination, and determine the probability of that combination given the individual competences. For example, if there are three individuals with competence distribution (.6, .7, .8) and respective weights of .2, .4, and .4, then Table 2 gives the relevant information.

This weighted voting scheme will produce correct majorities 78.8% of the time compared with 80% of the time for the optimal weighting scheme. The optimal weights would be .15, .32, and .53 (or equivalently 0, 0, 1) and would allow the most competent individual to decide all of the majority decisions. Clearly, the actual distribution of outcomes departs strongly from the normal distribution of Fig. 2, so an exact calculation was necessary given the small numbers involved. No systematic sensitivity analysis has been done using either approximate or exact calculations of outcome distributions, but examples indicate that the results are robust over small modifications in the weights.

In those cases where p_i values are not known and where the nature of the decision process rules out using concordance with majority opinion as a surrogate for judgmental accuracy, an alternative is to use subjective estimates of relative competence based on the judgments of each member of the group as to the judgmental accuracy of his fellows. Berge (1975) provides an iterative algorithm to aggregate individual relative competence judgments. This algorithm will in general converge to a unique set of competence loadings (see also Mirkin, 1979, chapter 4; and Kreweras, 1965).

TABLE 2 The Probability that a Majority Vote of a Three-Member Group with Distribution of Competence (.6, .7, .8) Will Reach a Correct Verdict under a (.2, .4, .4) Weight Assignment.

A	B	C	total correct weighted votes	probability
+	+	+	.2+.4+.4 = 1.0	.6x.7x.8 = .336
+	+	−	.2+.4 = .6	.6x.7x.2 = .084
+	−	+	.2+.4 = .6	.6x.3x.8 = .144
+	−	−	.2 = .2	.6x.3x.2 = .036
−	+	+	.4+.4 = .8	.4x.7x.8 = .224
−	+	−	.4 = .4	.4x.7x.2 = .056
−	−	+	.4 = .4	.4x.3x.8 = .096
−	−	−	= 0	.4x.3x.2 = .024
				1.000

An interesting and important implication of Theorem II is that if individuals are appropriately weighted, then the likelihood that the group decision is correct is directly related to the size of the majority. Just as the optimal weight for individual i is $w_i = \log(p_i/(1-p_i))$, so the size of the weighted majority, M = total weights for minus total weights against, is the weight that the group should be given in aggregating it with the other decision-makers. It is apparent that $M = \log(p_N/(1-p_N))$ and hence $p_N = \exp M/(\exp M)+1$. Thus, the larger the size of the majority for a given group, the greater the probability that the group majority is correct. Any system that is concerned with accuracy should take this into account, because it implies that even if a particular group is correct 99 times in 100 on average, if it reaches some given decisions by a bare majority, one can have little confidence in that particular decision. It would in general be best never to have to accept a decision based upon a slim majority. This provides another form of justification for supramajoritarian decision requirements (Cf. Crain and Tollison, 1977).

When the optimal weights are not used, then a bigger majority need not guarantee a higher probable judgmental accuracy. Consider the distribution of competence (.8, .8, 0). If equal weights are used, when exactly 2 voters are in agreement, they are correct with probability 16/25. If all 3 voters are in agreement they are correct with probability zero. However, using equal weights, we may prove that if all of the members of a group have at least a .5 probability of being correct, then a larger majority necessarily indicates a greater probability that the group is correct (Grofman, Owen and Feld, 1982a). For most purposes, we can assume that every individual has at least a coin-toss chance of being correct and so can feel much more confidence with large majorities than with small majorities, even using equal weights. Also, if N is large enough, then using our normal approximation from Fig. 2, it is easy to see that no matter what the distribution of underlying weights and competences, if the outcomes are distributed normally with mean greater than .5, then the larger the majority the larger the expected probability that the group is correct in its decision. Again, of course, the meaningfulness of such a statement requires the existence of a criterion by which, in the long run at least, correctness can be judged[5]. In many real world situations, the context is one where preferences conflict and there exist no objective standards for establishing trade-offs among conflicting criteria. Here, the assumptions of our models would be inapplicable in principle. Rather, our models deal with contexts in which (a) there is a "right" answer or (b) there is consensus as to what ought to constitute the public interest, but also a divergence in estimates of particular facts required to evaluate outcomes in terms of their social utility.

III. OPTIMALLY DELEGATING DECISION RESPONSIBILITY TO TASK FORCES

In many cases it may be very expensive in person-power to decide all issues by a vote of a large group such as a legislature. Instead, decisions will be delegated to smaller task forces. Sometimes those task forces are delegated full responsibility to make policy. More commonly, they must refer decisions back to the larger body (or sometimes an intermediate body larger than the task force but smaller than the group's entire membership), but with the practical consequence that unless the small group is divided on an issue, its recommendations will almost certainly be accepted by the parent body. We wish to consider here this problem of optimal delegation. We shall, however, consider only some very simple special cases, but we believe that these results can be generalized.

Consider a situation where all voters have a competence of .8. Let us then choose between two delegation procedures. In one, all decisions are reached by a majority vote of a three-member committee. Such three-person groups have probability of .896 of reaching a correct majority decision. On the other hand, each decision could be delegated to a two-person group with a requirement that they agree but with the promise that if there is disagreement, then the case be submitted to a three-person group for a de novo consideration.

It is straightforward to show that this second system actually can be expected to require fewer committee members than the simple three-person committee procedure using majority rule. The two-person committees reach correct decisions in 64% of the cases and incorrect decisions in 4% of the cases, so tie-breaking committees are only used in the remaining 32% of the cases. So on

---

[5]We have been assuming that group members or experts each reach conditionally independent judgments. It may be that there is interdependence because some members of the group are taking their cues as to how to vote from the same source. Let us consider a simple case which reveals the essential features of such a situation. Let α be the probability that a voter agrees with the choice of an opinion leader. Let $1-\alpha$ be the probability that a voter chooses independently of the preference of this opinion leader. It is assumed that there exists only one opinion leader and that α is the same for all voters. For simplicity let us assume that the opinion leader has competence $\bar{p}$.

It is easy to see that $\alpha > 0$ lowers the judgmental competence of the group majority. Moreover, the effect of α can be huge. For example, if $\alpha = .2$ and $\bar{p}_i = p = .6$, then $E(p_N) = .6$ for all N; i.e. the group majority is only exactly as competent as the opinion leader and the Condorcet Jury Theorem effect of raising the group competence toward 1 if $\bar{p} > .5$ is lost entirely. Shapley and Grofman (1981) consider the independence problem in greater detail.

the average, we require five committee members only 32% of the time and two committee members 64% of the time, which is less than three persons average. In addition, the second system is more accurate than the first. The two-person committees are accurate in 64% of the cases (94% of the cases that they resolve), and their tie-breaking mechanism is correct in 89.6 percent of the 32% of the cases on which it is convened. So the second system is correct 92.7% of the time compared with the 89.6% correct judgments of the three-person committee operating under simple majority. The two step system can be expected to provide more correct verdicts using fewer people. It should also be noted that fewer decisions are made by a single vote, so fewer problems are resolved with high risk of error.

An investigation of the robustness of this example shows that if the competence of the committee members is less than .7, then the resoluteness of the two-stage delegation process declines, and it can be expected to "use up" more committee members than the three-member majority rule committee. However, the two-tiered system will continue to provide more correct verdicts than the three-person committees. If the competence of committee members is very high, then the proposed system still provides better decisions at substantially lower expected costs in "person-power".

Consider a second example. If we compare five-person majority committees with two-person unanimous committees with a five-person committee called in to resolve deadlock, we find that even if the two-person groups have .5 average competence or below, this system reduces person-power requirements. If the committee members have relatively low average competence (i.e. under .8), then this system provides more accurate judgments than the five-person committees alone (see Table 3); but when the competence of the committee members increases, then the simple five-person majority committee is better. It is interesting to note that if competence is high enough (i.e. > .85), then it again becomes efficient to use three-person unanimous committees to initially decide the cases, followed by a five-person committee to break any deadlocks.

TABLE 3 Comparing 5-Person Committees Using Majority Rule to 2- and 3-Person Committees under Unanimity with a 5-Person Committee Used to Resolve Deadlock

Competence of jurors (p)	5-person majorities correct	2-person Dual System correct	3-person Dual System correct
.6	.683	.688	.708
.7	.837	.842	.870
.8	.942	.941	.964
.9	.991	.988	.997

The data in Table 3 suggest that a two-stage process with unanimous decisions required at the first stage is likely to provide high accuracy if committee members are of low competence but may decrease accuracy if committee members are of very high competence. However, the differences in Table 3 are of remarkably small magnitude.

More generally, these examples led us to conjecture that, rather paradoxically, under certain not implausible circumstances, small task forces using a unanimity rule with a tie-breaking mechanism could actually dominate large group majority decisions both in yielding a higher expected accuracy and in yielding a lower expected manpower requirement.

IV. DECOMPOSING PROBLEMS INTO BINARY CHOICES

In work in progress, the present authors are looking at the accuracy of group judgments under alternative election procedures. Our conjecture is that the use of binary choice procedures increases both individual and group judgmental accuracy. Let us turn, however, from the question of accuracy of group judgments to the question of "reasonable" procedures which groups might use.

We know from literature in cognitive psychology (see also Armstrong, Denniston and Gordon, 1975) that problem-solving is usually speeded up when tasks are decomposed into subcomponents which can be tackled in isolation and sequentially. Thus, one piece of advice we would give to decision-makers (and not just those in developing nations) is (to paraphrase a line from a recent Broadway comedy) "Take human bites !" i.e. tackle problems in manageable chunks and avoid input and information overload.

There are a number of procedures which can be used to structure decisions among multiple alternatives. One of these is a pairwise elimination tournament structure (see Fig. 3).

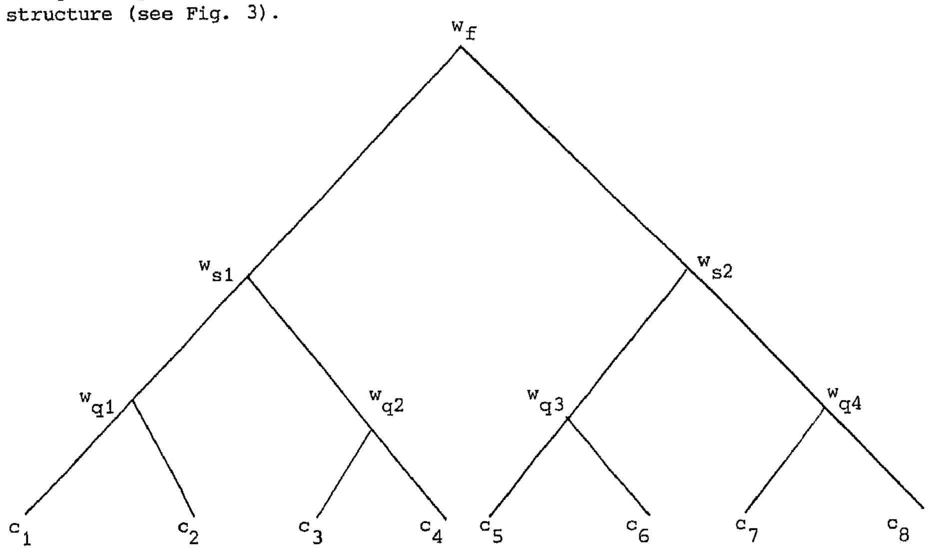

Fig. 3. Pairwise elimination tournament (w_q = winner of the quarter finals; w_s = winner of the semifinals; w_f = overall winner).

Here we begin with 2^k alternatives (k=3 in Table 3) and reduce these to a single choice after 2^k-1 pairwise contests.

Let us briefly consider some other decision procedures. In general, we would wish to select the choice, if any, which would be preferred to each of the others in a paired competition, but if m were large we would not want to actually conduct all (m(m-1))/2 paired comparisons among m alternatives. There are several procedures which will require at most m-1 pairwise ballots. One of these (which has a good deal to recommend it) is "standard amendment procedure" (see Farquharson, 1969) which is simply a process of sequential pairwise elimination. Under this procedure, the alternatives would be ordered in some fashion and the bottom alternative paired against the next alternative, with the winner of that contest paired against the next alternative and so on.

Let us look at a classic problem, drawn from Athenian jurisprudence, described in Farquharson (1969). A Greek jury must decide between three alternatives, a = acquittal, b = banishment, and c = condemnation (death). Perhaps, a more "natural" way than standard amendment procedure to structure the jury deliberation in the Greek tribunal case would be to first have a vote between acquittal and conviction and then a vote between banishment and condemnation <u>if</u> conviction carried. Such a procedure would involve what Farquharson (1969) referred to as the "successive procedure". We show this procedure in Fig. 4.

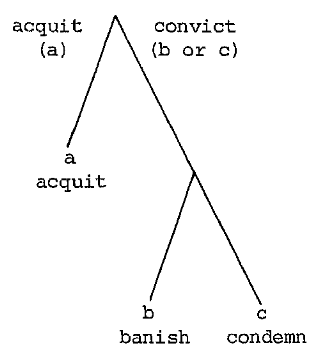

Fig. 4. The successive procedure for the three-alternative case of a jury tribunal choosing between conviction and acquittal and between two levels of sentence severity

While the successive procedure has certain important drawbacks (see Farquharson, 1969; Miller, 1977), it does have the advantage, in the problem we are considering, of partitioning the decision-making in a reasonable way into two separate (if not actually separable) issues - conviction vs. acquittal and nature of sentence (if conviction).

In Anglo-Saxon parliamentary procedure there is a motion called "division of the question". This motion calls for the matter under consideration to be divided into separate issues, each of which can be considered on its own merits for decisions each of which do not <u>logically</u> entail decisions on any other issues. A similar attempt to keep decision-making "confined" to single-issue dimensions is found in legislative rules requiring germaneness of amendments (Froman, 1967; Shepsle, 1979).

V. USING INFORMATION FEEDBACK TO IMPROVE THE ACCURACY OF GROUP DECISION-MAKING INVOLVING COMPLEX ESTIMATION OR CHOICE TASKS

Many complex decisions involve problems of simultaneous estimation. What is an endogenous parameter for one purpose is an exogenous parameter for another. An important parameter, say, projected energy costs over a 20-year period, may appear in many parts of a government planning process, e.g. in transport policy, import policy, urban/suburban redevelopment, etc. All too often, interrelated issues are parceled out to different decision-makers and/or totally different estimates of the same parameter are being used by different branches of government or even by different sections of the same bureau.

The DELPHI technique has been proposed as a means of providing information feedback to improve the accuracy of the group judgmental process. One of the earliest of the DELPHI experiments (taking place in the early 1950s, but not reported until Dalkey and Helmer, 1963) is also the most useful illustration for our purposes because it involved a single-complex decision task rather than a series of relatively simple discrete and unrelated decisions (Cf. Helmer, 1963; Brown and Helmer, 1964; Dalkey, 1969a, 1969b; Dalkey and others, 1970). In this experiment the authors required U.S. experts to seek to select (from the viewpoint of a Soviet strategic planner) an optimal Soviet missile targeting of U.S. industrial capacity and to generate an estimate of the number of Soviet A-bombs required to reduce U.S. munitions output by a prescribed amount. The technique used was repeated written interrogation of each expert (as to the factors entering his decision, his own parameter estimates for these factors, and information as to the kind of data that the expert would like to have in order to arrive at a better appraisal of these factors and therefore a more confident and reliable answer to the question(s) posed), interspersed with controlled opinion feedback. The information fed back to the experts between rounds of questioning consisted of the kinds of "available data" previously requested by one of the experts (e.g. output statistics for steel mills) or information about "factors and considerations suggested as potentially relevant by other respondents (e.g. the extent to which power transmission facilities permit reallocation of electric power)" (Dalkey and Helmer, 1963). All interaction between subjects was indirect, mediated by the experimenters. Communications relevant to the parameter estimation process were in written form.

In a complex decision task, the notion that each expert can be characterized by a single p_i value is erroneous. Rather, we may think of the problem as factorizable into subcomponents, with each expert having a differing level of competence with respect to the various "dimensions" of the problem. Ideally, feedback from fellow experts modifies the views of experts on those dimensions of the problem where they are least competent to make an independent judgment and about which they should be expected to be least confident and most willing to change their mind (at least, if confidence in one's competence and actual competence are correlated - which does seem to be true; see Dalkey and others, 1970). In such a case, the result of information feedback will be to improve the accuracy of group decision-making, since most shifts toward a group consensus will be in the direction of improved accuracy.

There are three features of the DELPHI technique that we wish to single out for special attention. First is the stress on <u>identification</u> of the factors (task subcomponents) which are relevant to the decision-making process. There is an important literature in mathematical psychology (see e.g. Dawes and Corrigan, 1974; Einhorn and others, 1977) which, as we read it, argues that the crucial stage of the decision task is the identification of all the factors relevant to prediction/choice; in particular, this literature suggests that this stage of the decision process is more important to the eventual accuracy of the final decision than the stage involving the attempt to accurately estimate the relative weight to be given to those factors which have been identified as relevant. Moreover, once all relevant factors have been identified, appropriate expertise can be brought to bear on what is now a more nearly decomposed problem.

A second key feature of this DELPHI experiment is the stress on fact-oriented experimenter-mediated non-face-to-face interactions.

> This mode of controlled interaction among the respondents represents a deliberate attempt to avoid the disadvantages associated with more conventional use of experts, such as round-table discussions or other milder forms of confrontation with opposing views. The method employed in the experiment appears to be more conducive to independent thought on the part of the experts and to aid them in the gradual formation of considered opinion. Direct confrontation, on the other hand, all too often induces .... an inclination to close one's mind to novel ideas, a tendency to defend a position once taken, or alternatively and sometimes alternately, a predisposition to be swayed by persuasively stated opinions of others (Dalkey and Helmer, 1963: 459).

A third facet of the DELPHI procedure used by Dalkey and Helmer (1963) worth calling attention to is the fact that the experiment was conducted in stages, with repeated rounds of feedback. While it may be good advice to students to always stick with their first answer on a multiple-choice exam rather than agonize about it and eventually make the wrong choice, for most decision-making the proverb "Act in haste, repent in leisure" is the more appropriate. Moreover, sequencing the decision process makes it possible to delegate responsibility for further investigation of the information particularly relevant to some given piece of the jigsaw puzzle, and to conduct parallel rather than serial processing.

VI. CONCLUSIONS

If we wish to optimize the probable accuracy of judgmental estimates of unknown parameters, then our principal conclusions are these:

1) Experts should be evaluated in terms of their performance record on a series of dichotomous "predictions". This evaluation is done best by looking at their skill at "retrodiction" (i.e. prediction of events which have already happened), thus allowing a direct measure of judgmental accuracy; but if $\bar{p} > 1/2$, then the relative competence of a set of experts on some choice dimension can be evaluated with respect to their conformance to the majority judgments of their fellow experts, and this concordance can be used to estimate Bayesian-optimal weights. Also, expert weights can be estimated by asking group members to evaluate each other's competence and then using a convergent estimation technique for combining of those estimates of individual competence.

2) Group decisions which integrate expert and nonexpert opinion, involving matters where there can be (even if only judged post hoc) determinations as to the accuracy of group parameter estimates, should use Bayesian-optimal weight assignments assigned as described above.

3) Full responsibility for many decisions should be delegated to relatively

small task forces operating under a unanimity (or near unanimity) requirement, but with a proviso for the convening of a larger body to consider the problem de novo if the task force deadlocks.

4) Wherever possible, decisions should be partitioned into a series of sequential dichotomous choices involving distinct issue (problem) dimensions.

5) Where final decisions must be based on multiple dimensions of interrelated parameter estimations (e.g. in predicting population growth making use of complex simultaneous estimation techniques and pooling of judgments), then information feedback among the experts involved in estimating the various pieces of the problem is highly desirable. One technique which has shown promise is DELPHI.

REFERENCES

Armstrong, J.S., W.B. Denniston, and M.M. Gordon (1975). The use of the decomposition principle in making judgments. Organizational Behavior and Human Performance, 14, 257-63.

Berge, C. (1975). Introduction à la Théorie des Hypergraphes. Montreal University Press, Montreal.

Black, D. ((1958). The Theory of Committees and Elections. Cambridge University Press, London.

Brown, B., and O. Helmer (1964). Improving the Reliability of Estimates Obtained from a Consensus of Experts. Rand Corporation, P-2986.

Condorcet, N.C. de (1785). Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix. Paris.

Crain, W.M., and R.D. Tollison (1977). Legislative size and voting rules. Journal of Legal Studies, 6, 235-240.

Dalkey, N.C. (1969a). The Delphi-Method: An Experimental Study of Group Opinion (RM-5888-PR). The Rand Corporation, Santa Monica, California.

Dalkey, N. (1969b). An experimental study of group opinion: The Delphi method. Futures, 1, 408-426.

Dalkey, N., and O. Helmer (1963). An experimental application of the Delphi method to the use of experts. Management Science, 9, 458-467.

Dalkey, N., B. Brown, and S. Cochran (1970). Use of self-rating to improve group estimates: Experimental evaluations of Delphi procedure. Technological Forecasting, 1, 283-291.

Davis, J.H. (1973). Group decision and social interaction: A theory of social decisions schemes. Psychological Review, 80, 97-125.

Dawes, R.M., and B. Corrigan (1974). Linear models in decision making. Psychological Bulletin, 81, 95-106.

Duda, R.O., and P.E. Hart (1973). Pattern Classification and Scene Analysis. Wiley, New York.

Einhorn, H.J., R.M. Hogarth, and E. Klempner (1977). Quality of group judgment. Psychological Bulletin, 84, 158-172.

Farquharson, R. (1969). Theory of Voting. Yale University Press, New Haven, Connecticut.

Froman, L.A. (1967). The Congressional Process: Strategies, Rules, and Procedures. Little Brown, Boston.

Gelfand, A., and H. Solomon (1973). A study of Poisson's models for jury verdicts in criminal and civil trials. Journal of the American Statistical Association, 68, 271-278.

Gelfand, A., and H. Solomon (1974). Modelling jury verdicts in the American legal system. Journal of the American Statistical Association, 69, 32-37.

Gelfand, A., and H. Solomon (1975). Analyzing the decision-making process of the American Jury. Journal of the American Statistical Association, 70, 305-310.

Gelfand, A., and H. Solomon (1977). Comparing juries: An argument in favor of twelve-member juries. Jurimetrics, 17, 292-313.

Gelfand, A., and H. Solomon (1977). An argument in favor of twelve-member juries. In S. Nagel (Ed.), Modeling the Criminal Justice System, Sage Criminal Justice System Annual, 7. Sage Publications, Beverly Hills, California, pp. 205-220.

Grofman, B. (1974). Mathematics and politics: Mathematical reasoning and optimal jury decision processes. And "Reply". In M. Black (Ed.), Problems of Choice and Decision: Proceedings of a Colloquium Held in Aspen, Colorado, June 24-July 6, 1974. Photo-offset, Cornell University Program on Science, Technology and Society and Aspen Institute for Humanistic Studies, pp. 331-337; 544-547.

Grofman, B. (1978). Judgmental competence of individuals and groups in a dichotomous choice situation. Journal of Mathematical Sociology, 6, 47-60.

Grofman, B. (1979). The case for majority verdicts. Trial Magazine, 29, 23-25, 47-48.

Grofman, B. (1980). Preliminary models of jury decision-making. In G. Tullock (Ed.), Frontiers of Economics. pp. 3.

Grofman, B. (1981). Mathematical modeling of jury/juror decision making. In B.D. Sales (Ed.), Perspectives in Law and Psychology, Volume II: The Jury, Judicial and Trial Processes. Plenum, New York.

Grofman, B., G. Owen, and S.L. Feld (1982a). Thirteen theorems in search of the truth. Theory and Decisions. Forthcoming.

Grofman, B., G. Owen, and S.L. Feld (1982b). Average competence, variability in individual competence, and the accuracy of statistically pooled group decisions. Psychological Reports. Forthcoming.

Grofman, B., G. Owen, L. Shapley, and S.L. Feld (1980). Eleven Theorems in Search of the Truth. Paper presented at the Annual Meeting of the American Political Science Association, Washington, D.C..

Helmer, O. (1963). The Systematic Use of Expert Judgment in Operations Research. P-2795, RAND Corp.

Klevorick, A.K., and M. Rothschild (1979). A model of the jury decision process. Journal of Legal Studies, 141-164.

Kreweras, G. (1965). A model to weight individual 'authority' in a group. In S. Steinberg and others (Eds.), Mathematics and Social Sciences I. Mouton, Paris, pp. 111-118.

Lorge, I., and H. Solomon (1959). Two models of group behavior in the solution of eureka-type problems. Psychometrika, 20, 139-148.

Lorge, I., D. Fox, J. Davitz, and M. Brenner (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. Psychological Bulletin, 55, 337-372.

Luce, D. (1959). Individual Choice Behavior. Wiley, New York.

Margolis, H. (1976). A note on incompetence. Public Choice, 26, 119-129.

Miller, N. (1977). Graph theoretical approaches to theory of voting. American Journal of Political Science, 21, 769-803.

Minsky, M., and S. Papert (1971). Perceptrons: An Introduction to Computational Geometry. MIT-Press, Cambridge.

Mirkin, B. (1979). Group Choice. Wiley, New York.

Moore, E.E., and C.E. Shannon (1956a). Reliable circuits using less reliable relays, I. Franklin Institute, Philadelphia Journal, 262, 191-208.

Moore, E.E., and C.E. Shannon (1956b). Reliable circuits using less reliable relays, II. Franklin Institute, Philadelphia Journal, 262, 281-297.

Nitzan, S., and J. Paroush (1980a). Investment in human capital and social self-projection under uncertainty. International Economic Review.

Nitzan, S., and J. Paroush (1980b). Are Qualified Majority Rules Special ? Unpublished manuscript.

Owen, G., B. Grofman, and S.L. Feld (1981). The Generalized Condorcet Jury Theorem. Unpublished manuscript.

Pierce, W. (1961). Improving Reliability of Digital Systems by Redundancy and Adaptation. Unpublished Ph.D. thesis in Electrical Engineering, Stanford University.

Poisson, S.D. (1837). Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile: Précedées des Règles Générales du Calcul des Probabilités. Bachelier, Paris.

Sattler, H.E. (1966). Effect of group variability on pooled group decisions. Psychological Reports, 18, 676-678.

Shapley, L., and B. Grofman (1981). Optimizing Group Judgmental Accuracy. Unpublished manuscript.

Shepsle, K. (1979). Institutional arrangements and equilibrium in multidimensional voting models. American Journal of Political Science, 23, 27-59.

Steiner, I.D. (1966). Models for inferring relationships between group size and potential group productivity. Behavioral Science, 11, 273-283.

Steiner, I.D., and N. Rajaratnam (1961). A model for the comparison of individual and group performance. Behavioral Science, 6, 142-147.