Models of Consensus for Validation of Expert Systems

Daniel E. O'Leary and Karen V. Pincus
University of Southern California

ABSTRACT Models of consensus are used in the validation process to develop a basis for system performance. This paper develops two analytic models of consensus that can be useful in the validation process. The first model employs the binomial model to study the probability that the consensus judgment is correct or incorrect. This problem is critical for the validation of expert systems, since the model leads us to conclude that in some cases consensus judgment is not appropriate. That basic model is extended to account for both different levels of expertise and unequal prior odds. The second model is a Bayesian model of the use of consensus judgment and the validation process. There are two applications of the Bayesian model. The first application of the Bayesian model finds that, in some cases, consensus judgment is not appropriate and to generate a hypothesis as to the conditions for the use of the consensus judgment. The second application compares an expert systems performance to the consensus judgment. It can be used to assist in deriving stopping criteria for the validation process.

1. INTRODUCTION

Validation of expert systems and other computer models is dependent on the ability to test the correctness of the computer program or model. Unfortunately, there are few situations where the responses of an expert system can be determined as objectively correct. As a result, validators must depend on surrogate measures of correctness of system responses. An important and frequently used approach is that of expert consensus.

The purpose of this paper is to develop and explore analytic models of consensus in order to (a) study conditions under which consensus is a reasonable basis of comparison, (b) to structure the use of consensus as a validation tool and (c) model the validation process that compares the expert system to the consensus solution. In so doing, this paper provides theoretic foundations of consensus, and a basis for the use of consensus in the practice of validation.

1.1 Consensus as a Basis of Comparison

Lack of consensus among a group of experts implies that some of the experts are not correct (or are employing a different model). However, even complete agreement among the experts does not guarantee a correct solution.

International Journal of Expert Systems Pp. 237–249, Volume 6, Number 2, ISSN 0894-9077

Unfortunately, there is little evidence on the relationship between consensus and correctness. Empirically, researchers have found that experts in some domains have been correct only 50 to 60% of the time (Sorenson et al., 1982). Thus, it is reasonable to assume that in some situations, consensus judgments will not be correct. Thus, there is interest in determining some of the characteristics of consensus, e.g., under what conditions should we expect it to be used.

1.2 System Comparison to Consensus

Oftentimes in the validation of expert systems and other more complex models, it is not clear which answers are correct or incorrect. Thus, the system is compared to experts. For example, in Yu et al. (1979), the system MYCIN was compared to a set of experts, in order to determine which had the best solution to a set of problems.

In many situations, experts do not agree with each other. As a result, consensus of a panel of experts is used as a surrogate for the correct response. For example, given a set of three experts, if the system's response agrees with two or more of the experts, we assume that the system is correct. Otherwise, the system's response is judged as incorrect.

1.3 This Paper

This paper proceeds as follows. Section 2 develops a basic model of the correctness of consensus judgments. That section summarizes some classic research as applied to the consensus problem. Section 3 investigates some extensions of that model, by relaxing assumptions inherent in that model. Some of the results of section 3 are new, such as the conditions for use of consensus in the situation of unequal prior odds. Section 4 studies when the consensus judgment is correct in the context of a Bayesian model. Section 5 uses that same Bayesian model for the comparison of an expert system to consensus judgment. Section 6 reviews some of the implications of these models and briefly discusses issues in their implementation. Section 7 provides a brief summary and some extensions.

2. A CONSTANT PROBABILITY MODEL

Throughout this paper, it is assumed we are concerned with dichotomous decisions or recommendations. Thus, we are concerned with a set of experts who will choose between those two alternatives over a series of situations or test problems. For example, bankers must decide whether loan applicants will default or not default.

The validation process is assumed to employ n experts, each with an equal probability of being correct. The probability of success is constant for each problem. The experts are assumed to arrive at their decisions independently. The expert's decisions are then summarized by some unbiased source to determine the consensus judgment.

2.1 Background and Assumptions

Condorcet (1785) first recognized that Bernoulli's (1713) work on the binomial could be used to model the probability of reaching correct decisions under different voting systems. Condorcet's (1785) work become the basis of modern research in voting (e.g., Black (1958)) and jury decision making (Gorfman & Owen, 1986). One of the common themes of that research is to determine the probability that the consensus position is correct.

The binomial consists of n independent trials, where each dichotomous choice decision has a probability p of success and a probability (1-p) of failure. In a validation setting, the use of the binomial would assume that each of the validating panel would have equal competence. In addition, it is also assumed that each of the two alternatives is equally likely to be correct. This assumption of equal prior odds creates a special case of the binomial, analogous to the case of using a fair coin. Each of these assumptions will be relaxed later in the paper.

2.2 A Model of the Consensus-Correctness Relationship

Let n be the number of experts in the validation experiment. Let M be the minimum number of experts necessary to establish a majority. When n is odd, M=(n+1)/2, when n is even M=(n/2)+1. Let m be the number of experts for a given consensus, where m = M, ..., n. Let P_C be the probability that consensus judgment is correct.

Given the two assumptions from the previous section,

$$P_C = \sum_{m=0}^{n} \binom{n}{m} p^m (1-p)^{n-m}.$$

A set of binomial table values for P_C for some values of p and n is given in Table 1.

2.3 Some Results from the Model

Condorcet (1785) found a number of important results from the use of the binomial as a model of consensus. Assume that n is odd and $n \geq 3$.

Table 1. Probability of Consensus Being Correct Assumes Equal Prior Odds

n	*p=.10*	*p=.30*	*p=.50*	*p=.70*	*p=.90*
3	.028	.216	.500	.784	.972
5	.009	.163	.500	.837	.991
7	.003	.126	.500	.874	.997
9	.001	.099	.500	.901	.999

Result 1
If $p > .5$ then $P_C > p$.
Result 2
If $p > .5$ then P_C is monotonically increasing in n with a limit of 1.
Result 3
If $p = .5$ then $P_C = .5$.
Result 4
If $p < .5$ then P_C is monotonically decreasing in n with a limit of 0.
Result 5
If $p < .5$ then $P_C < p$.

Result 1 indicates that if $p > .5$, then the probability that the consensus decision is correct, is greater than the probability that a single decision is correct. In this situation, consensus is a useful surrogate for correctness.

Result 2 suggests that, if $p > .5$, then the more the number in the panel of experts evaluating the system, the higher that the probability of consensus is correct.

Result 3 finds that in this specific case nothing is gained by going from individual judgments to consensus judgments. If the probability of all decision makers being correct is .5, then the probability that consensus of those decision makers is correct is also .5.

Result 4 indicates that, if $p < .5$, then the more the number in the panel of experts evaluating the system, the lower that the probability of consensus is correct.

Finally, Result 5 finds that if $p < .5$, then the probability that the consensus decision is correct, is less than the probability that a single decision is correct. In this situation, consensus actually results in a lower probability of correctness.

3. EXTENSIONS OF THE BASIC MODEL

There were two primary assumptions in the model of the previous section: equal competence of experts and equal prior odds. This section extends the model of the previous section by relaxing these assumptions.

3.1. Relaxation of the Equal Competence Assumption

It is reasonable to assume that different experts will have a different probability of providing the correct decision. For example, experts are often delineated as having different titles indicating gradation in expertise. Thus, it is reasonable to assume that the experts come from a number of different classes, where within each class, each expert is equally competent, yet there is an ordering of the competence of the different classes.

Assume there are two different groups of experts, A and B (this assumption could be extended to more than two groups). It is assumed that within either of those two groups the probability of correctness is equal. Let p_i be the probability that an individual expert in group i is correct, i = A or B. Assume that $.5 < p_A \leq 1$ and that $p_B < p_A$. Let $P_{C(i)}$ be the probability that a consensus decision of group i is correct, i = A, B or, A and B (written as A,B).

Margolis (1976) examined the model with this revised assumption and developed the following three results.

Result 6
If $p_B \leq .5$ then $P_{C(A,B)} < P_{C(A)}$.
Result 7
If $p_B > .5$, then there exists some cardinality of group B, referred to as a critical value B*, such that $P_{C(A,B)} > P_{C(A)}$.
Result 8
There exists some value $p_{B^*} < p_A$, such that if $p_B > p_{B^*}$ then $P_{C(A,B)} > P_{C(A)}$.

Result 6 indicates that if the value of p_B is low enough, then it does not make sense to aggregate the experts of the two classes in order to develop a consensus value. Result 7 indicates that for p_B of an appropriate level, if group B is large enough then it makes sense to integrate the members into one large group of A and B, that will make the consensus decision. Result 8 indicates that if p_B is large enough, then group B should be integrated with group A, regardless of the size of group B. These results are surprising to a certain extent, since they indicate that, in some situations, lower quality decision makers should be integrated in with higher quality decision makers for consensus judgments.

Result 7 may lead to the requirement that group B be quite large, so as to be impractical in the case of validating expert systems. If there are 30 members of A, $p_A = .7$ and $p_B = .51$, then B* would be several hundred, and thus beyond the scope of an expert system validation.

Using results from Margolis (1976), the critical point nature of Result 8 can be exemplified as follows. If $p_A = .9$ then $p_{B^*} = .82$. If $p_A = .8$ then $p_{B^*} = .70$. If $p_A = .7$ then $p_{B^*} = .62$. If $p_A = .6$ then $p_{B^*} = .55$.

These results can be extended. For example, the following result indicates that if a subset of some set of experts is being used to develop a consensus judgment, then it is always better to add more of those same equal experts to the set of experts from which consensus is being developed.

Result 9
Let A* be a subset of A. $P_{C(A)} \geq P_{C(A^*)}$ for all A*, not equal to A.

3.1.2. Limitation of Equal Class Behavior

The limitation of this approach is that it is assumed that within a class, all experts are equally likely to be correct. This limitation can be addressed by using some approximations to the binomial distribution normal and poisson.

3.1.3. Normal Approximation

The normal distribution can be used as an approximation to the binomial Feller (1950). Using that approximation, an alternative approach has been developed by Grofman (1978) and Grofman et al. (1983) that employs this result. Rather than multiple distinct sets of experts, they treat the set of experts as a single class, with competency normally distributed with a mean of p# and a variance of p#(1-p#)/n. Thus, expert correctness has a distribution. In

that case the conclusions of the equal competence model will hold, with p# substituting for p.

3.1.2. Poisson Approximation

The poisson distribution also can be used to approximate the binomial Feller (1950), where the poisson is defined as $p(k;L) = e^{-L}*L^k/k!$. In the same sense that the normal approximation to the binomial can be used to develop an alternative approach to the multiple classes, so can the poisson distribution. In the approximation of the poisson distribution, the parameter L is equal to n*(1-p). With L specified as n*(1-p) the same results as in section 2 hold. Unlike the binomial, the only constraint on L, is that L reflects the density of correct judgments in the group of experts.

3.2. Relaxation of the Equal Prior Odds Assumption

The model in section 2 also assumes that there are equal prior odds as to which of the alternatives is correct. This assumption is equivalent to the "fair coin" assumption that both a head and a tail are equally likely on each toss. However, in most decision making situations it is unlikely that the relevant states of nature are equally likely. For example, in the case of the prediction of bankruptcy, roughly 3% of the firms in the United States go bankrupt each year. If the choice is between predicting bankrupt or not bankrupt, then the prior odds are, respectively, .03 and .97.

Let p_S be the probability of one state of the dichotomous decision occurring. Let $p_{S'} = (1 - p_S)$, be the probability of the other. In the case of equal prior odds, $p_{S'} = p_S = .5$. If the prior odds are not equal, then there is no longer interest in p, instead the concern is with a revised probability that captures the difference in the prior odds. Let p_R be the probability of the expert making the correct decision, given the prior odds for the state of nature S, assuming all experts are of equal competence. Let $p_{R'}$ be the probability of the expert choosing alternative R′, making the correct decision, given the prior odds for the state of nature S′

Table 2. Probability of an Individual Decision Being Correct Given Unequal Prior Odds and Various Competencies When Prior Odds are Equal

	Competency (p) for Equal Prior Odds Decisions				
Prior Odds	*p=.10*	*p=.30*	*p=.50*	*p=.70*	*p=.90*
.10	.012	.045	.100	.206	.500
.20	.027	.097	.200	.368	.692
.30	.045	.155	.300	.500	.794
.40	.069	.222	.400	.609	.857
.50	.100	.300	.500	.700	.900
.60	.143	.391	.600	.778	.931
.70	.206	.500	.700	.845	.955
.80	.308	.632	.800	.903	.973
.90	.500	.794	.900	.955	.988

and assuming equal competence. Using Bayes' theorem, we have $p_R = (p*p_S)/[(p*p_S) + (1-p)*p_{S'}]$ and $p_{R'} = (p*p_{S'})/[(p*p_{S'}) + (1-p)*p_S]$. Some example values are given in Table 2.

3.2.1. Relationship Between p and p_S, and p_R

There are a number of relationships between p, p_S, and p_R that map the revised model into results obtained for the basic model, as discussed in results 1–5.

Result 10
If $p + p_S > 1$ then $p_R > .5$.

Proof of Result 10
$p_R = (p*p_S) / [p*p_S + (1-p)*(1-p_S)]$
$p_R = (p*p_S) / [2p*p_S + (1 - p - p_S)]$
Since $(1 - p - p_S)$ is less than 0, $p_R > .5$

Result 11
If $p + p_S < 1$ then $p_R < .5$.

Result 12
If $p + p_S = 1$ then $p_R = .5$.

Proof of Result 12
$p_R = (p*p_S) / [p*p_S + (1-p)*(1-p_S)]$
$p_R = (p*p_S) / [2p*p_S + (1 - p - p_S)]$
Since $p + p_S = 1$, $p_R = .5$

How is the probability that a consensus judgment is correct impacted by the unequal prior odds? Results 1–5 when combined with results 10–12 provide us with the answer. We should use consensus only if $p + p_S > 1$. Thus, the quality of consensus judgments is a function of both those probabilities.

3.2.3. Monotonicity Result for Revised Model

In addition, we can establish a monotonicity result for p_R. In particular, the following result indicates that p_R is monotonically increasing as the prior odds increase.

Result 13
p_R is monotonically increasing in p_S.

Proof
Let $p_S > p_S''$, then

$(p*p_S) / [p*p_S + (1-p)*(1-p_S)] > (p*p_S'') / [p*p_S'' + (1-p)*(1-p_S'')]$

$(p*p_S) [2p*p_S'' + 1 - p - p_S'')] > (p*p_S'') [2p*p_S + 1 - p - p_S)]$

$p*p_S - p*p*p_S > p*p_S'' - p*p*p_S''$

$p*p_S - p*p_S'' > p*p*p_S - p*p*p_S''$

Since $p_S > p_S''$, the inequality holds and p_R is monotonically increasing in p_S.

This can be a useful result. For example, we can make the following two statements. First, if we know p and have a conservative estimate of p_S, such that $p + p_S > 1$, then we know that we should use consensus. We do not need to know p_S exactly. We may be able to use simply a lower bound. Second, if the prior odds are greater than .5, we know that the simplified equal odds model underestimates p_R. Thus, in some cases the equal prior odds model helps bound the case where the prior odds are not equal.

4. A BAYESIAN MODEL OF CORRECT CONSENSUS JUDGMENTS

If the individual decisions are independent and the probability of a correct decision remains constant, then the binomial can be used as a model, as noted earlier. In the case of the basic binomial model, with equal prior odds and competence, if $p > .5$, $P_C > .5$, and the consensus solution should be used for each validation test problem.

However, it is reasonable to assume that one correct decision leads us to a higher probability that the next judgment will be correct. This is consistent with a Bayesian approach where prior probabilities are updated with experience, e.g., using a test data approach.

4.1. Test Data Analysis

Test data is one of the primary approaches to validating an expert system. For example, O'Leary (1991) found that over 50% of the validation effort and effectiveness was judged to come from test data. In some cases that analysis found that validation efforts only used test data.

It will be assumed that validation is done using a test data approach. If the decision is correct for one set of test data, then it will be assumed that this result increases our subjective probability that the judgment will be correct for the next set of test data. In this situation, we can model the process of determining the probability of a correct judgment using a Bayesian model.

4.2. Prior Probability

Bayesian analysis (for the mathematical relationship see Raiffa and Schlaifer (1961, pp. 50–51) of a binomial often assumes that the decision maker's prior probability distribution is distributed according to a beta distribution with parameters k and n. In this paper it is assumed that p represents the underlying probability of a correct decision. In addition, it assumed that p is distributed according to a beta distribution, with k and n chosen to represent prior feelings and information about the expert system to be validated. The choice of the parameters k and n is important, since the expected value of p is k/n.

As a result, assume that the prior distribution for p is $Pr(p \mid k,n) = [(n-1)!/(k-1)!(n-k-1)!]\, p^{k-1}(1-p)^{n-k-1}$, where $1 \geq p \geq 0$, and $n > k > 0$.

4.3 Probability Revision

There are some useful properties associated with treating the beta as the prior for p. If there are n# test problems and there are k# correct decisions, then the parameters of the revised distribution, the posterior, are written as the sum of the parameters in the prior, plus these changes, resulting in the revised parameters n+n# and k+k#. The expected value of the posterior is (k+k#)/(n+n#). Thus, p=k/n, becomes p# = (k+k#)/(n+n#). Thus, this distribution allows us to capture the success of using the test data in updating the parameters of the distribution.

4.4 Example

Suppose that, in general, three out of five validation judgments are correct on the test data, i.e., k = 3 and n = 5. In that situation, for the first validation test problem the probability of a correct judgment is .6 and the probability of an incorrect judgment is .4. In the case of a second test problem there are four branches in the decision tree. If the first judgment is correct then the probability that the second judgment is correct is 4/6 and the probability that the second judgment is incorrect is 2/6. If the first judgment is incorrect, then the probability that the second judgment is correct or incorrect is 3/6. Thus, for the second judgment, the probability that the judgment is greater than .5 is (3/5) * (4/6) or .4. The probability that the second decision is correct is either 4/6 or 3/6 depending on what happened with the first problem.

5. A BAYESIAN MODEL OF VALIDATION

In this section, it is assumed that validation of a knowledge-based system is done using a series of system tests. Associated with each test the expert system is either correct or not correct (e.g., consistent with the consensus or the no consensus positions). As long as the model is correct, the validation process continues. If at any point in time it is not correct, the validation process stops. At that point, the system is in need of additional knowledge acquisition, knowledge maintenance or else the project should be eliminated.

Unlike sections 2, 3 and 4, the concern is not with the probability of correctness of the consensus position. Instead the focus of this section is on the correctness of the expert system, i.e., consistency with the consensus position which is assumed correct.

5.1. Bayesian Structure

The test process has a dichotomous outcome: either there are results correct (e.g., consistent with the consensus position, C) or the results are not correct (consistent with the no consensus position, NC). As a result, the outcomes over a set of test problems form a binomial process, with a parameter p.

A Bayesian structure is assumed. If the system is consistent with the consensus position, then it is assumed that it will change the probability, that for the next test case the system

Table 3. Relationship Between Probability that System Agrees with Consensus and Not Consensus Groups Given i Test Cases

k	*n=2*	*n=3*	*n=4*	*n=5*	*n=6*	*General n*
1	(i)	(i)/2	(i)/3	(i)/4	(i)/5	(i)/(n–1)
2		(i+1)	(i+1)/2	(i+1)/3	(i+1)/4	(i+1)/(n–2)
3			(i+2)	(i+1)/2	(i+2)/3	(i+2)/(n–3)
4				(i+3)	(i+3)/2	(i+3)/(n–4)
5					(i+4)	(i+4)/(n–5)

will be consistent with the consensus position. The same Bayesian model from section 4 will be assumed to model the process. Thus, the prior is a beta distribution.

5.2. Case of k=3 and n=5

Consider the case of k=3 and n=5, with a two test validation problem. For the first set of test data p=.60. If the system is tested and consistent with the consensus judgment, then the parameters become (3+1) and (5+1), respectively. In that case p becomes 4/6, for the second case of test data. Alternatively, if the system is not consistent with the consensus judgment, then the parameters become (3) and (5+1), respectively and then p becomes 3/6. As a result, the probability of being consistent with consensus in the second period is (3/5)*(4/6) and *the probability of being consistent with the no consensus judgment in the second period is* (3/5)*(2/6). Similarly, the probability of consistency with the consensus judgment in the third period is (3/5)*(4/6)*(5/7), and the no consensus judgment in the third period is (3/5)*(4/6)*(2/7).

Thus, in general we have the infinite sequence e.g., Knopp (1956) (3/5), (3/5)(4/6), (3/5)(4/6)(5/7), ... But that sequence is equivalent to (3/5), (3*4)/(5*6), (3*4)/(6*7), (3*4)/(7*8),..., (3*4)/(i)*(i+1), ... (i = 4, ...,)

In a similar manner we have the infinite sequence associated with the lack of consensus, (2/5), (3/5)(2/6), (3/5)(4/6)(2/7). But that sequence is equivalent to (2/5), 24*(1/4)(1/5)(1/6), 24*(1/5)(1/6)(1/7),..., 24/(i-1)*(i)*(i+1), ... (i = 4, ...,).

Each of these two sequences is monotonically decreasing and is bounded from above. Thus, both sequences converge. In addition, we can see that, for this case of k and n, the probability of consensus, through the ith test case (i = 1, 2, ...), is (i+2)/2 times greater than the probability of no consensus at that ith case, after i-1 test cases of consensus.

5.3. The General Case

Using an approach similar to the case of k=3 and n=5, general results on the relationship between the probability of the system agreeing with the consensus group and the non-consensus group, can be developed. These results are summarized in Table 3.

6. IMPLICATIONS AND IMPLEMENTATION

This section discusses some of the implications of the models in this paper and their implementation.

6.1. Implications

The basic model and its extensions, discussed in sections 2 and 3, has a number of implications. First, the model indicates that the decision on whether or not consensus should be used is a function of the sum of two parameters: p and p_S. Consensus should not be used indiscriminately. Second, in the consensus decision in the basic model, where $p > .5$, it is always beneficial for the use of a complete set of the best experts. If all the top experts cannot participate, then it is likely that the next highest class of experts should also be used in the development of the consensus judgment. Third, the models imply some stopping criteria or design criteria for consensus analysis. For example, we can see from table 1, that if there are equal prior odds and if $p = .70$ and we wish a $p_C \geq .8$, then we must use at least 5 experts to develop the consensus judgment.

The findings of the Bayesian model have a similar set of implications. It appears that if the probability the consensus judgment is correct is greater than or equal to .5, then it is better to use the consensus judgment.

The Bayesian model discussed in section 5 has an implication for the extent of validation effort. As the number of test problems increases, the multiplicative relationship between the probability of agreement with the consensus or not consensus group increases rapidly. Validators could choose the number of test problems used in the validation effort based on that relationship.

6.2. Some Implementation Considerations

In order to implement the models in this paper, basic knowledge of the underlying parameters is required. In the first model of consensus, the probabilities p and p_S, were necessary. The competency levels p could be difficult to obtain. One approach would be to use a set of experiments where the experiments had a known answer. Prior odds of events, p_S could be obtained from experience. However, there is little in the literature about the quality of expertise in even broad categories of events. In a similar manner, the parameters k and n for the Bayesian model could be obtained using an analysis of empirical data or similar experimental approaches.

7. SUMMARY AND EXTENSIONS

This paper has developed two basic models that can be useful in the validation of expert systems and other complex models. The first model was based on the binomial, but was

extended to include multiple levels of expertise and unequal prior odds. The results presented here summarized some classic results and presented new results.

The second model was a Bayesian model, designed to study the correctness of the consensus judgment and the validation process. The first application was important, since it investigated a situation where the probability that consensus judgment was correct changed with different test problems. The second application corresponded to many real world validation efforts, where validation continues through some set of test problems as long as the validation of the system is successful. If at any point in time the system fails, a decision may be made that more knowledge acquisition or scrapping the system is appropriate.

7.1. General Extensions

The results presented here are not limited to the validation of expert systems and other complex models. Instead they could be extended to general consensus problems.

Each of the two models assumed a dichotomous decision. Thus, rather than the basic binomial models presented in this paper, more general multinomial models could be developed.

The models presented in this paper ignore the impact of multiple systems. In effect, each system is treated separately. Thus, another extension would account for such portfolio effects.

7.2. Extensions of the Basic Binomial Model

The basic model was limited to simple majorities as the means of the definition of consensus. Alternative approaches used by other organizations may include a two-thirds majority. These alternative definitions of consensus could be accounted for in the model developed above.

7.3. Extensions of the Bayesian Model

Although the Bayesian model was couched in the context of a comparison of the correctness of expert judgment and a comparison of expert system and consensus judgment, other formats could be used. The validation process could be structured as a sequential voting process. For example, rather than multiple time periods, we could view the resulting decision tree from the perspective that each branch is a different expert and the probabilities relate to a single test problem. In that case, the agreement of the first expert with the system would lead to the posterior with a parameter of $(k+1)/(n+1)$.

In the development of the Bayesian model, a beta prior was assumed. Other alternative prior distributions could have been developed. For example, Raiffa and Schlaifer (1961) discuss a normal distribution with similar updating properties to the beta distribution. Other extensions might also be generated for the Bayesian model.

Acknowledgment: The authors would like to acknowledge the helpful comments of the three anonymous referees and the editors.

REFERENCES

Bernoulli, J. (1899) [1713]. *Ars Conjectandi (The Art of Conjecturing)*, first published posthumously in Latin (Later published as Wahrscheinlichkeitsrechnung, Leipzig, Engelmann).

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

Condorcet, M. (1785). [Marie Jean Antoine Nicolas Caritat, Marquis de Condorcet] *Essai sur l'application de l'analyse a la probabilitie des voix*, [Essay on the Application of Analysis to the Probability of Majority Decisions], Paris, Imprimerie Rayale.

Einhorn, H. (1974). Expert judgment: some necessary conditions and an example. *Journal of Applied Psychology, 59*(5), 562–571.

Feller, W. (1950). *An Introduction to Probability Theory and its Applications.* John Wiley, London.

Goldberg, L., & Werts, C. (1966). The reliability of clinician's judgments: a multi-method approach. *Journal of Consulting Psychology*, June, 199–206.

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

Grofman, B., & Owen, G. (1986). Condorcet Models, Avenues for Future Research. In B. Grofman & G. Owen (Eds.), *Information Pooling and Group Decision Making: Proceedings of the Second University of California, Irvine, Conference on Political Economy*. Greenwich, CT: JAI Press.

Grofman, B., Owen, G., & Feld, S. (1984). Thirteen theorems in search of the truth. *Organizational Behavior and Human Performance, 33*, 350–359.

Knopp, K. (1956). *Infinite Sequences and Series.* New York: Dover.

Margolis, H. (1976). A note on incompetence. *Public Choice, 26*, 119–127.

O'Leary, D. (1991). Development, design and validation of expert systems. In *Verification, Validation and Testing of Expert Systems*, John Wiley.

Raiffa, H., & Schlaifer, H. (1961). *Applied Statistical Decision Theory.* Harvard, Cambridge, MA.

Sorenson, J., Grove, H., & Selto, F. (1983). Detecting Management Fraud: An Empirical Approach. *Symposium on Auditing Research 1982*, University of Illinois, (pp. 72–116) Champaign.

Yu, V., Fagan, L., Wraith, S., Clancey, W., Scott, A., Hannigan, J., Blum, R., Buchanan, B., & Cohen, S. (1979). Antimicrobial selection by computer: a blinded evaluation by infectious disease experts. *Journal of the American Medical Association, 242*, 1279–1282.

Address all correspondence to Daniel E. O'Leary, School of Business, University of Southern California, Los Angeles, CA 90089-1421. e-mail: oleary@mizar.usc.edu.