INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

- 1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.**
- 2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.**
- 3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.**
- 4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.**
- 5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.**

Xerox University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106

74-21,271

FUNARO, Joseph Francis, 1941-
AN EMPIRICAL ANALYSIS OF FIVE DESCRIPTIVE
MODELS FOR CASCADED INFERENCE.

Rice University, Ph.D., 1974
Psychology, experimental

University Microfilms, A XEROX Company, Ann Arbor, Michigan

RICE UNIVERSITY

AN EMPIRICAL ANALYSIS OF FIVE DESCRIPTIVE
MODELS FOR CASCADED INFERENCE

by

Joseph F. Funaro

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Thesis Director's Signature:

A handwritten signature in cursive script, reading "Wesley M. Du Charne", written over a horizontal line.

Houston, Texas

April, 1974

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to the members of his dissertation committee; Dr. Wesley Du Charme who served as chairman, and Drs. John Brelsford, William Howell, Paul Pfeiffer and David Schum. A very special debt is owed to Dr. Wesley Du Charme whose scholarship, patience, and wit helped "make the rocky path, smooth and the crooked way, straight."

The author would also like to thank the personnel of the Naval Aerospace Medical Institute for their invaluable assistance in data collection and in particular Dr. Norman E. Lane who wrote the computer program for data analyses.

Finally the author would like to acknowledge his indebtedness to the U. S. Navy, Bureau of Medicine and Surgery for funding his two academic years at Rice University, and who also provided general support for the dissertation research.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
INTRODUCTION	1
EXPERIMENTAL DESIGN	9
DATA ANALYSES	15
RESULTS	17
DISCUSSION	27
REFERENCES	36

INTRODUCTION

Most previous research on human inference has examined the ability of subjects to revise their opinions about the relative likeliness of two or more hypotheses on the basis of events known to have occurred. The major finding in most of these studies has been that intuitive opinion revisions are conservative in comparison to the optimal revisions specified by Bayes' theorem--that is, subjects revise their opinions in the proper direction but the magnitude of the revision is less than optimal (Peterson & Miller, 1965; Phillips & Edwards, 1966). A recent line of formal and empirical research has focused on probabilistic opinion revision when the occurrence of an event is not known with certainty (Dodson, 1961; Gettys & Wilke, 1969; Schum & Du Charme, 1971; Schum, Du Charme & De Pitts, 1973). This uncertainty occurs whenever some external source of data, or one's own observation, is less than perfectly reliable. In a medical context this is the task facing a physician who must revise his opinion that a patient has a specific disease given the report of a symptom from a laboratory test which is not completely reliable.

Inference in such situations can be described as cascaded or multi-stage in that more than one kind of information processing step is necessary. Since in many contexts reports are unreliable, one must distinguish between actual events D and $\bar{D}$ (in the case of binomial data) and reports of the occurrence of these events D^* and $\bar{D}^*$. Given two states or

hypotheses, H_1 and H_2 , which are mutually exclusive the diagnostic impact of D on H_1 and H_2 is specified by the likelihood ratio L_D where

$$L_D = \frac{P(D | H_1)}{P(D | H_2)} \quad (1)$$

A similar form holds for $L_{\bar{D}}$. When reports rather than events must be dealt with, however, an adjusted likelihood ratio (Λ) must be used. Schum and Du Charne (1971) developed formal expressions for determining Λ in certain well defined cases. Under the conditions of the present experiment Schum and Du Charne have shown that Λ can be defined as

$$\Lambda = \frac{P(D^* | H_1)}{P(D^* | H_2)} = \frac{P(D^* | D) P(D | H_1) + P(D^* | \bar{D}) P(\bar{D} | H_1)}{P(D^* | D) P(D | H_2) + P(D^* | \bar{D}) P(\bar{D} | H_2)} \quad (2)$$

Given Λ , the optimal model for opinion revision is Bayes' theorem. Posterior odds (Ω_1) of H_1 and H_2 are determined by:

$$\Omega_1 = \Lambda \cdot \Omega_0 \quad (3)$$

where Ω_0 is the prior odds of H_1 to H_2 . The cascaded inference in this situation involves combining observational reliability with event probability to obtain Λ and then combining Λ with prior odds to obtain posterior odds.

Several recent experiments examined human intuitive inference in situations which involved cascaded information processing. The major finding of these studies was that

intuitive cascaded inferences were generally excessive when compared to the calculations of the formally appropriate model. Youssef and Peterson (1973) compared subjects' cumulative posterior log odds revisions under cascaded and non-cascaded conditions both to each other and to the Bayesian values. The diagnosticity of the data and the reliability of the sources in their cascaded situation were such that if Λ was calculated in the optimal manner its value was equal to that of L in the noncascaded situation. Since Λ had the same value as L the optimal posterior odds were the same for both conditions.

When intuitive posterior odds under both conditions were compared to optimal odds they were slightly excessive for the lowest values of L and Λ , and became increasingly conservative for higher values. However, the posterior odds estimated under cascaded conditions were consistently greater than those estimated under noncascaded conditions with equivalent impact ratios. As a consequence cascaded posterior odds maintained their excessiveness to optimal odds over a wider range of impact ratios than did noncascaded odds. A later study (Youssef, 1973) required subjects to make information purchase decisions in lieu of posterior odds estimates and extended the generality of the findings to this indirect response mode.

One possible explanation of the excessiveness of cascaded to noncascaded posterior odds is that subjects may

be using the same process to revise their odds under both conditions but are overestimating Λ in the cascaded situation. Such an explanation might be described as a two-stage, two-process, model of intuitive cascaded inference. The first stage process, unique to cascaded inference produces estimates of Λ which exceed veridical values; the second stage process, common to both cascaded and noncascaded inference combines the impact ratio (Λ or L) with prior odds via an algorithm whose posterior odds are generally conservative in comparison to Bayes' theorem.

Studies by Snapper and Fryback (1971), and Schum, Du Charme and De Pitts (1973) provide strong evidence that subjective estimates of Λ are excessive compared to the veridical values of the Schum and Du Charme model. In addition, Snapper and Fryback found that subjects' estimates of Λ could be closely approximated by the formally inappropriate model, $\tilde{\Lambda} = r \cdot L$ (where r is the reliability of the reporting source¹ and L is the likelihood ratio of the datum). A post hoc analysis revealed that the Snapper and Fryback model closely approximates subjects' intuitive estimates of Λ in the Schum, Du Charme and De Pitts study.

The apparent ability of the Snapper and Fryback model to predict subjective estimates of Λ suggests a means of developing a descriptive model of intuitive cascaded inference and testing the two stage-two process hypothesis. If subjects utilize the same process for the second stage of cascaded

inference that they employ for single stage (noncascaded) inference, then it should be possible to predict subjective posterior odds by utilizing subjective values of Λ calculated by the Snapper and Fryback model in an algorithm which predicts intuitive single stage inference. One such single stage algorithm was developed by Phillips and Edwards (1966) who showed that subjective posterior odds (Ω_1) could be closely approximated by the model:

$$\Omega_1 = L^c \cdot \Omega_0 \quad (4)$$

where the value of the exponent, c , depended on the likelihood ratio of the datum.

A study by Gettys, Kelly and Peterson (1973) is directly relevant to a test of the two stage-two process model. The study compared subjects' intuitive cascaded odds revisions to the optimal odds generated by the odds form of a modified version of Bayes theorem (which Schaeffer and Borchering, 1973, have shown to be equivalent to the Schum and Du Charme model) and to a "best guess" model.

The "best guess" model predicts that the subject first infers what the posterior odds would be if the most likely event was true and then reduces the posterior odds by the amount he feels to be appropriate given the unreliability or uncertainty of the reporting source. The "best guess" model in its most general form does not specify how much the subject will reduce his odds given the unreliability of the reporting source. Gettys et al. hypothesized that the subject would

multiply the posterior odds by the reliability of the most likely datum. This version of the "best guess" strategy, called Model 1 by the authors and Hybrid Model 1 in this paper, is a two stage-two process model. The algorithm for the first stage process is Bayes' theorem ($\Omega_1 = L \cdot \Omega_0$). The algorithm for the second stage process specifies that the cascaded subjective posterior odds ($\tilde{\Omega}_1$) are equal to the posterior odds of the first stage times the reliability of the source or

$$\tilde{\Omega}_1 = \Omega_1 \cdot r \quad (5)$$

which can be rewritten

$$\tilde{\Omega}_1 = (L \cdot \Omega_0) r \quad (5a)$$

The predictions of Hybrid Model 1 are the same as would result from a descriptive model which used the Snapper and Fryback algorithm for the first stage process and Bayes' theorem for the second stage process. The posterior odds for this version, which will be called Hybrid Model 2, would be:

$$\tilde{\Omega}_1 = (L \cdot r) \Omega_0 \quad (6)$$

Both of these models can be described as "hybrid" in that they contain both a descriptive and an optimal algorithm. Most of the studies of intuitive single stage inference have found that such inferences are conservative in comparison to the values generated by Bayes' theorem (Slovic &

Lichtenstein, 1971). It therefore seems reasonable that more accurate predictions of subjective cascaded inference can be made by a "descriptive" model--that is, a model which employs a descriptive algorithm for each stage of cascaded inference. This paper suggests one such model which combines the Snapper and Fryback algorithm for subjective $\Lambda(\tilde{\Lambda})$ with the Phillips and Edwards algorithm for single stage inference. The mathematical description of the model is given by:

$$\tilde{\Omega}_1 = \tilde{\Lambda}^c \Omega_0 \quad (7)$$

where, $\tilde{\Lambda}^c = (r \cdot L)^c$; c is a fitted parameter dependent on the subject and the impact ratio; and Ω_0 = the prior odds. This model will be referred to as Descriptive Model 2 in that it follows the same sequence as Hybrid Model 2 but employs a descriptive algorithm for both stages. It should be readily apparent that there can also be a Descriptive Model 1 whose mathematical formulation is

$$\tilde{\Omega}_1 = (L^c \cdot \Omega_0)r \quad (8)$$

Although it is apparent that no distinction can be made between the predictions of Hybrid Models 1 and 2 such is not the case with their descriptive analogs. As we have seen the mathematical analog of Descriptive Model 1 is

$\tilde{\Omega}_1 = (L^c \cdot \Omega_0)r$ while the mathematical analog of Descriptive Model 2 is $\tilde{\Omega}_1 = (L \cdot r)^c \Omega_0$. If the prior

odds are set at one, the two models reduce to $L^c \cdot r$ and $(L \cdot r)^c$ respectively and obviously give rise to different predictions when $c \neq 1.0$.

The experiment by Youssef and Peterson described earlier suggests yet another model. They found that cascaded posterior odds were consistently excessive when compared to single stage posterior odds (when $\Lambda = L$). However, the magnitude of the difference was slight and in the absence of any information about variability in their study one may wonder whether the difference is a real one. The implication for the present study is that a subject's intuitive cascaded odds revisions can perhaps be approximated by his single stage odds revision when $\Lambda = L$. The model then can be stated as

$$\tilde{\Omega}_1 = \Omega_{ss} \quad (9)$$

where Ω_{ss} stands for a subject's single stage cascaded odds revisions in a condition where L is set equal to Λ . Such a model predicts that a subject's odds revisions in the cascaded condition will be the same as his single stage revisions in a condition where L is set equal to Λ . Since the predictions of this model are based on a subject's empirically determined single stage posterior odds the expression in (9) will be called the "Empirical" Model.

The objective of the present experiment was to compare subjects' intuitive cascaded posterior odds revisions with

the predictions of the various models described in Table 1. It can be seen that the predictions of the optimal and hybrid models can be made "a priori" using the experimental values of r and L for the cascaded condition. Descriptive Models 1 and 2 require that the values of c be determined empirically for each subject. Since both models imply that the subject uses the same process for combining impact ratio and prior odds under both cascaded and noncascaded conditions, a subject's values for c in the cascaded condition can be determined from his single stage inferences. However, since c is a function of the size of the impact ratio and each model utilizes a different impact ratio (L for Descriptive Model 1 and $L \cdot r$ for Descriptive Model 2) for the cascaded condition it is necessary for each subject to make single stage inferences based upon both impact ratios. Similarly the Empirical Model requires that the subject makes single stage inferences under a condition in which L equals Λ .

Experimental Design

Each inference task utilized a pair of symmetrical binomial hypotheses (i.e., $P(D | H_1) = P(\bar{D} | H_2)$). Each subject was required to make sequential odds revisions under three single stage conditions and one cascaded condition. Prior odds were equal to one for all inference tasks. The values of the likelihood ratios in the single stage tasks were those required to estimate the parameters used by the models to be tested. In order to permit the maximum discrimi-

Optimal Model

See equation 3

$$\Omega_1 = \Lambda \cdot \Omega_0$$

Hybrid 1 and 2

$$\begin{aligned} \tilde{\Omega}_1 &= (L \cdot \Omega_0) r \\ \tilde{\Omega}_1 &= \text{or} \\ &= (L \cdot r) \Omega_0 \end{aligned}$$

where r equals the reliability of the source and L equals the likelihood ratio of the datum in the cascaded inference

Descriptive 1

$$\tilde{\Omega}_1 = L^c \cdot \Omega_0 \cdot r$$

where r equals the reliability of the source, L equals the likelihood ratio of the datum in the cascaded inference and c is empirically determined from single stage inference where L single stage equals L cascaded

Descriptive 2

$$\tilde{\Omega}_1 = (L \cdot r)^c \cdot \Omega_0$$

where r equals the reliability of the source, L equals the likelihood ratio of the datum in the cascaded condition and c is empirically determined from single stage inference where L single stage equals L cascaded times r

Empirical

$$\tilde{\Omega}_1 = \Omega_{ss}$$

where Ω_{ss} is the subject's single stage posterior odds when L single stage equals the value of Λ for the cascaded inference

Table 1

Mathematical and verbal summary of the models

nation between the predictions of the models, three different L and r values were selected for the cascaded conditions. The experimental design is summarized in Table 2.

Apparatus

Each pair of binomial hypotheses was represented by two discs having different proportions of red and blue. A gray and white disc represented the reliability of the report under the cascaded condition where r equaled the proportion of the disc that was white. The numerical values of the proportion of each color were printed on each disc. The experimenter had identical discs. In addition, he had two balanced spinners, one was attached to the red-blue disc and the other to the reliability disc when appropriate. The experimenter's discs were concealed from the subjects by a screen. The data for each sequence were displayed by hanging appropriately colored (red or blue) poker chips on a peg board.

Subjects

Thirty subjects were randomly assigned to each of the three groups. All ninety were males in the Naval Aviation Training Program at Pensacola, Florida. The subjects performed in groups of about 15. Total experimental time for each group was approximately one hour.

Procedure

Phase One - Single Stage Inference:

Each group was required to make sequential single stage odds revisions for each pair of binomial hypotheses as

	Group 1	Group 2	Group 3
Single Stage	Condition 1	Condition 2	Condition 3
	$L_s = r \cdot L_c = 12$ Sequence Length 4-6	$L_s = r \cdot L_c = 7$ Sequence Length 4-6	$L_s = r \cdot L_c = 16$ Sequence Length 4-6
Inference	Condition 2	Condition 3	Condition 4
	$L_s = \Lambda = 1.36$ Sequence Length 9-11	$L_s = \Lambda = 1.98$ Sequence Length 9-11	$L_s = \Lambda = 4.00$ Sequence Length 9-11
Tasks	Condition 3	Condition 4	Condition 5
	$L_s = L_c = 20$ Sequence Length 4-6	$L_s = L_c = 10$ Sequence Length 4-6	$L_s = L_c = 19$ Sequence Length 4-6
Cascaded	Condition 1	Condition 2	Condition 3
	$L_c = 20; r = .60$ Sequence Length 9-11	$L_c = 10; r = .70$ Sequence Length 9-11	$L_c = 19; r = .83$ Sequence Length 9-11
Inference	Condition 2	Condition 3	Condition 4
Task	Condition 3	Condition 4	Condition 5

Table 2
Summary of Experimental Design*

*NOTE: L_s = likelihood ratio for single stage inference, L_c = likelihood ratio for cascaded inference, r = reliability for cascaded inference, Λ = adjusted likelihood ratio for cascaded inference for each group.

described in Table 2. The appropriate discs were displayed throughout each condition. Subjects were told that the experimenter would flip a fair coin and select one of the two discs. He would then attach a balanced spinner to its center and spin it a number of times. After each spin the experimenter reported the color that the spinner landed on by placing an appropriately colored poker chip on the sequence display board. After each datum was displayed the subjects recorded their posterior odds in individual response booklets. They first wrote down which disc they considered more likely on the basis of the report, then how many times more likely they considered that disc than the other. All numerical estimates were required to be of the form $X:1$. This procedure should always yield a posterior odds estimate greater than one. The same procedure was repeated for a total of five different sequences under each pair of binomial hypotheses and over the experiment half the sequences favored the predominantly red disc and half the predominantly blue. Immediately prior to the five experimental sequences, each group participated in one practice sequence of five data. The practice sequence was followed by five independent sequences, whose length varied as described in Table 2. The apparatus was arranged so that it appeared that the experimenter was actually selecting a disc and using a spinner to generate each sequence. He in fact used preprogrammed sequences which had been generated by a fair binomial process

prior to the experiment. As can be seen in Table 2 sequence length varied according to condition and group. The longer sequences were necessary in the less diagnostic conditions to obtain comparable data for all groups.

Phase Two - Cascaded Inference

After completing the single stage tasks, each group was required to make cascaded inferences under the conditions described in Table 2. The appropriate discs representing the binomial hypotheses and the reliability of the reporting source was displayed. The procedure within this condition was as follows. The experimenter flipped a coin to select one disc from the appropriate pair of red-blue discs and attached a spinner to its center. The other spinner was attached to the center of the appropriate reliability disc and both spinners spun. Subjects were informed that the experimenter would observe the outcomes of both spins to determine his report. If the reliability spinner landed on gray, the experimenter falsely reported the color on which the red-blue spinner stopped; otherwise he reported truthfully. The subjects were informed that this rule determined the report. On the basis of the experimenter's report subjects wrote down their posterior odds estimates about which red-blue disc had been selected. Once again the experimenter presented one practice sequence followed by five experimental sequences, all of which were preprogrammed. The subjects went through Phase One and Phase Two, performed an inter-

polated task, and then repeated both Phase One and Phase Two. The interpolated task lasted approximately a half hour and consisted of taking parts of a standard naval aviation selection test. Only the data gathered the second time through were used to test the models. This procedure was used in order to provide a reasonably stable data base on which to evaluate the models.

Data Analysis

In order to evaluate the models it was necessary to determine the predicted and subjective posterior cascaded odds for each datum in a sequence. For the optimal model posterior odds are generated by the equation $\Omega_1 = \Lambda \cdot \Omega_0$. The sequence of poker chips which reported the results of the series of spins from the selected disc can best be described by a binomial process in which a success (s) is defined as the spinner stopping on the predominant color of the disc, and a failure (f) as the spinner stopping on the other color.

When the adjusted likelihood ratio, Λ , is equally diagnostic given either of the binomial events (as it was in the present experiment) the optimal model can be rewritten as

$$\Omega_1 = \Lambda^{(s-f)} \Omega_0$$

where s-f is the number of successes minus the number of failures in a sequence of data. Since the prior odds were equal to one the optimal model reduces to $\Omega_1 = \Lambda^{(s-f)}$.

The s-f version of all other models (except Empirical) follow the same general algorithm--posterior odds are equal to the appropriate adjusted likelihood ratio $[(L \cdot r)$ for Hybrid; $(L^c \cdot r)$ for Descriptive 1; and $(L \cdot r)^c$ for Descriptive 2] raised to a power equal to s-f.

Since the predictions for most of the models are linear when log odds are plotted against s-f both predicted and subjective posterior cascaded odds were converted to $\log_{10}$. The predicted posterior cascaded log odds of the Optimal and Hybrid models were obtained for each s-f value by the equation

$$\log \Omega_1 = (s-f) \log \Lambda ,$$

and

$$\log \Omega_1 = (s-f) \log (L \cdot r)$$

Obviously no parameters need to be fitted from the data to obtain the predictions of these models.

The mean predicted log odds for Descriptive Model 1 were obtained by using the equation

$$\log \Omega_1 = (s-f) [c \log L + \log r] .$$

The values of L and r are those given in Condition 4. The quantity c is a fitted parameter obtained from the subject's data generated under Condition 3. The value of c was obtained for each subject by first finding his mean inferred log likelihood ratio (ILLR) across all sequences in Condition 3 where

$$\text{mean ILLR} = \sum_{n=1}^{N=5} \frac{\log \Omega_n - \log \Omega_{n-1}}{n}$$

The parameter c was then obtained by dividing the mean ILLR by the Bayesian log likelihood ratio (BLLR).

The mean predicted log odds for Descriptive Model 2 were obtained by using the equation

$$\log \Omega_1 = (s-f) [c \log (L \cdot r)]$$

where L and r are the values in Condition 4 and c was computed from the subject's Condition 1 data.

The mean log odds at each value of $s-f$ in Condition 2 were used as the predictions of the Empirical Model for each subject.

Finally the posterior cascaded odds estimated by each subject in Condition 4 were converted to log odds and the mean log odds were computed for each value of $s-f$ across all sequences. These Condition 4 log odds represent the criterion data against which all the model's predictions will be tested.

Results

The models were evaluated on the basis of two general criteria--goodness of fit and consistency of fit across subjects. A model's goodness of fit was evaluated both in terms of individual and group data, Figures 1, 2, and 3 represent group data. The mean predicted and obtained cascaded log odds were plotted as a function of $s-f$ for each group. The obtained mean log odds and associated 95 per cent confidence intervals are based on the data of all subjects

Figure 1

Predicted and estimated mean log odds for Group I as a function of S-F. Vertical lines represent 95% Confidence intervals

Figure 2

Predicted and estimated mean log odds for Group 2 as a function of S-F. Vertical lines represent 95 % confidence intervals.

Figure 3

Predicted and estimated mean log odds for Group 3 as a function of S-F. Vertical lines represent 95 % confidence intervals.

within a group. It can be seen that the size of the confidence intervals increases both over s-f values and over groups. The increased variability as s-f increases and the fact that variability increases from Group 1 to Group 2 to Group 3 can perhaps be explained by noting that sequence diagnosticity increases with s-f and the diagnostic impact of Λ increases from Group 1 to 2 to 3. The increase in variability then may be due to differentially operating response biases (see Du Charme, 1970), i.e., as more extreme odds estimates are called for some subjects begin exhibiting response biases and others do not thus increasing the variability of the group results.

Reviewing the data for Group 1 it can be seen that only the Empirical Model falls within the confidence intervals of the subjective cascaded odds and exhibits similar irregularities in the aggregation process. Although none of the predictions of the optimal model fall within the confidence intervals of the data, it is clearly superior to the three remaining models. The predictions of the Hybrid Model are clearly inferior to all other models both in its estimation of the adjusted likelihood ratio (its value at s-f = 1) and its aggregation process (reflected in the slope of its predictions).

For Groups 2 and 3 the results are less clear. Although the values of L and r used for the three groups were chosen in an attempt to provide maximum separation of the predictions of the models, the predictions of the Descriptive

and Empirical models could only be approximated on the basis of previous studies. Unfortunately these estimates, particularly for the values of the c parameter, were enough inaccurate that the predictions of the models for Groups 2 and 3 fall very close together.

For Group 2 only one model, the Hybrid, is clearly rejected. Once again both its estimates of the adjusted likelihood ratio and aggregation process are clearly excessive. The predictions of the remaining models are similar in regard to both their estimation of the subjective adjusted likelihood ratio and aggregation process. Likewise, no model's predictions succeed in falling within all the confidence intervals of the data.

Group 3 results reveal that once again the Hybrid Model is clearly rejected. The predictions of the remaining models are similar, but only one model, the Empirical, has all of its predictions within all the confidence intervals of the data. Descriptive Model 2 predictions fall within all confidence intervals of the data except one (at $s-f = 0$). The predictions of Descriptive Model 1 and the Optimal Model are nearly indistinguishable. Also for the first time subjective cascaded odds are exceeded by those of the Optimal Model.

Figures 4, 5, and 6 provide an indication of how well the predictions of each model fit the intuitive cascaded odds of the individual subjects within a group. The figures

Figure 4

Mean absolute deviation between predicted and obtained log odds in Group 1 for each model. Vertical lines represent 95% confidence intervals.

Figure 5

Mean absolute deviation between predicted and obtained log odds in Group 2 for each model. Vertical lines represent 95 % confidence intervals.

Figure 6

Mean absolute deviation between predicted and obtained log odds in Group 3 for each model. Vertical lines represent 95% confidence intervals.

present the group average, and associated 95 per cent confidence intervals, of the absolute mean deviation between model predicted log odds and log estimated cascaded odds for each subject. The values were obtained by first computing the absolute mean deviation between a model's prediction for a subject and his intuitive estimates across all values of s-f and then computing the grand mean and confidence intervals across subjects within a group.

From Figure 4 it can be seen that the goodness of a model's fit to the cascaded data of individual subjects exactly parallels its fit to the mean cascaded odds of the group. The Hybrid Model is clearly the worst, Empirical is the best, followed closely by Optimal. Descriptive 1 and 2 fall in between with the former providing a slightly better fit.

Figure 5 reveals the same relationship between group and individual predictions. Hybrid is clearly inferior and Empirical slightly better than the other three models.

Figure 6 shows that once again the Hybrid Model is clearly inferior to the others. Descriptive 2, Empirical and Descriptive 1 deviations are nearly the same and their goodness of fit for individual subjects parallels that for group data. The fit of the Optimal Model for both group and individual data falls closer to that of the latter three models than it does to the Hybrid Model.

Since all three groups contain an equal number of

subjects it seems reasonable to take the mean of the absolute deviations of individual predicted and obtained data for all five models over groups. The ranking of all five models in order of goodness of fit to individual data over groups results in the Empirical, Optimal, Descriptive 2, Descriptive 1 and Hybrid having overall absolute mean log deviations of .37, .57, .61, .66 and 1.96 respectively.

Table 3 presents the number of subjects "captured" by the models both within and over groups. A model "captured" a subject when the mean absolute deviation between its predictions and the subject's data (collapsed over all values of s-f) was smaller than that of the other four models. It is quite evident that the predictions of the Empirical Model most accurately fit the greatest number of subjects, capturing 51 out of 90 or 57%. The superiority of the Empirical Model is consistent over groups. Descriptive Model 2 and the Optimal Model nearly tie for second place, capturing 16 and 14 subjects respectively. As with all other analyses the Hybrid Model fares worst capturing only 2 subjects out of 90. It is interesting to note that the Optimal Model is significantly bettered by only one model, Empirical. Its descriptive ability is even more surprising when one considers that its predictions are made a priori and require no parameter estimations.

Discussion

It may be concluded from the results that the Empirical

	D1	D2	Hybrid	Optimal	Empirical
Group I	1	2	0	7	20
Group II	0	10	1	6	13
Group III	6	4	1	1	18
Total	7	16	2	14	51

Table 3

Number of subjects "captured" (see text) by each model

Model predicts sequential subjective cascaded odds (both group and individual data) better than the other four models. Given the range of data diagnosticity and source reliability in the present study it is not unreasonable to expect that the Empirical Model would fairly accurately predict subjective cascaded odds within a quite broad range of parameter values.

It is appropriate to ask what characteristics of the Empirical Model contribute to its superiority. One of the unique characteristics of the Empirical Model is that its predictions are not constrained by linearity as are those of the other models. To the extent then that subjects are non-linear² in both single stage and cascaded inference the Empirical Model has a potential advantage over any model which makes linear predictions. In their log versions the predictions of the other models increase by a constant amount (their predictions of the subjective adjusted likelihood ratio) for each increase in s-f value. It has been well documented in previous studies that subjective odds revisions do not exhibit such a constant relationship in regard to changes in s-f. Instead subjective odds revisions often display a primacy or inertia effect resulting in considerable non-linearity across s-f values (Peterson and Du Charme, 1967; Pitz, 1969). The fact that the intuitive cascaded odds of subjects in the present experiment appear to be non-linear (with the amount of nonlinearity varying across groups) suggests that at least part of the Empirical

model's superiority is due to its ability to capture non-linear aspects of the subjects' responses. Support for the above argument can be gained by looking at the Group 3 results. Figure 3 shows that descriptive models 1 and 2 and the Empirical Model make very similar predictions, but Figure 6 and Table 3 show the superior fit of the Empirical Model.

Although one reason for the superiority of the Empirical Model may be due to the fact that the general shape of its predicted function parallels that of the cascaded data, the question remains as to what processes the model captures. The Empirical Model essentially assumes that subjects have no trouble arriving at an accurate estimate of Λ and that the aggregation of Λ values in cascaded tasks is carried on in the same manner as the aggregation of L values in single stage tasks. Thus cascaded inference is no more difficult than single stage inference and any deviations from the optimal model noted in cascaded inference can be attributed to the aggregation process and not to difficulties in ascertaining Λ .

For example, it can be seen that the group mean posterior log odds for cascaded data bear the same general relationship to optimal odds found in many previous studies of single stage inference. That is, when the datum diagnosticity is low (Group 1) subjective posterior odds are excessive; when the datum diagnosticity is slightly higher (Group 2) subjective

odds are nearly optimal; and finally when datum diagnosticity becomes sufficiently large (Group 3) subjective posterior odds display the often found conservative relationship to optimal odds. It could therefore be argued that subjects accurately estimate Λ and that the deviation of their posterior odds from optimal is due to the same inappropriate aggregation process or response bias (Du Charme, 1970) found in single stage inference. One would then expect that subjective posterior odds from a single stage inference task in which L was set equal to Λ would provide an accurate prediction of cascaded odds, as indeed they do.

There are however, two observations which seem to contradict the assumption that subjects are able to accurately determine the veridical value of Λ . First, the earlier studies by Snapper and Fryback (1971) and Schum, Du Charme, and De Pitts (1973) found that in general subjective estimates of Λ were excessive and better described by the inappropriate algorithm $\Lambda = L \cdot r$ than by the optimal model. Secondly, if subjects are able to estimate Λ accurately but are inefficient in data aggregation, one would nonetheless expect to find that their subjective cascaded odds would be nearly optimal at $s-f = 1$ (when prior odds are one) since no aggregation is required. With the exception of Group 3 such was not the case in the present experiment.

These seemingly contradictory observations--that subjects

should be able to estimate Λ quite accurately (since subjective cascaded odds do not depart significantly from noncascaded odds where $L = \Lambda$) and that estimates of Λ in both the present and previous experiments reveal suboptimality--can be reconciled. It is quite possible that in an experiment such as the present one in which subjects are shown a number of samples of data generated in a cascaded condition (they were exposed to a total of 12 samples, counting practice sequences) they can estimate veridical Λ quite accurately by taking the ratio of the mean proportion of successes to the mean proportion of failures across all sequences.

For example, in the present experiment all the single stage and cascaded sequences were generated by a fair binomial process. For Group 1 the veridical value of Λ was 1.36. The ratios of the mean proportion of successes to the mean proportion of failures over all sequences in Condition 2 and Condition 4 were 1.41 and 1.44, respectively. For Group 2 in which veridical Λ was 1.98, Conditions 2 and 4 had ratios of 2.57 and 2.19, respectively. Group 3 which had a veridical Λ of 4.0 had ratios of 6.04 and 6.29. These values while excessive in comparison to the veridical values for Λ were much closer to the ILLR's of the subjects than the values predicted by the Snapper and Fryback algorithm which were 12, 7, and 16, respectively, for the three groups.

The ability of subjects to accurately estimate the population proportion from sample data has been well documented (Peterson and Beach, 1967). In addition, Peterson and Swensson (1968) found that subjects can use this information to revise their odds as to which of two symmetrical binomial populations is being sampled even when the populations are otherwise unspecified. If subjects estimate Λ in such a direct manner it is not surprising that posterior odds generated by subjects in a single stage task where L is equal to Λ provide such an accurate prediction of their cascaded odds. The fact that the values of subjective cascaded odds at $s-f = 1$ are not optimal does not necessarily mean that subjects do not know the true value of Λ . In the single stage task where they are given the likelihood ratio the values of their posterior odds at $s-f = 1$ are also not optimal.

It is important to note that the strategy of estimating veridical Λ from the ratio of the mean proportion of successes to the mean proportion of failures over sequences of reports does not take into account which hypothesis is in fact true. Under the conditions of the present experiment -- symmetrical binomial hypotheses having equal prior odds -- the accuracy of the estimate is not unduly affected by the lack of feedback.

There is then a distinct possibility that subjects approximate the veridical value of Λ by estimating the ratio of the proportion of successes to failures in a series of

reports from an unreliable source and thereby reduce the information processing requirement to that of a single stage inference task. If this assumption is true there are many real world situations in which the same or a highly similar strategy could be applied. In a medical context, a physician specializing in heart disorders could either consciously or subconsciously keep track of the number of times an unreliable cardiac test reported the presence of a symptom along with the subsequent confirmation or refutation of the diagnosis. As was seen in the introduction, Schum and Du Charme defined Λ as the ratio $\frac{P(D^*|H_1)}{P(D^*|H_2)}$. By keeping track of test reports and the ultimate outcome of each case the physician in the above example could develop an empirical estimate of Λ . Such an estimate should approach the veridical value of Λ as the number of cases approaches infinity. Such an "empirical" strategy lacks the flexibility, generality, and precision of the optimal model but it does reduce the mental effort. The empirical approach does have the advantage that gradual fluctuations in data diagnosticity or source reliability over time will be reflected in the ratio $\frac{P(D^*|H_1)}{P(D^*|H_2)}$ and thereby in the estimate of Λ .

To summarize, the results support the Empirical Model as the best predictor of intuitive cascaded odds. The superior fit of the Empirical Model to subjects' cascaded odds may result from two factors, first, that subjects use

the same process for aggregating evidence in both single stage and cascaded inference and secondly, that subjects estimate Λ by taking the ratio of the mean proportion of successes to the mean proportion of failures across all sequences.

If these two assumptions are true they suggest that the most accurate prediction of subjective cascaded odds would result from having subjects estimate posterior odds in a single stage inference task in which the data sequences are identical to those used in the cascaded task. The use of identical sequences should result in identical estimates of the impact ratio and also maximize the congruence between single stage and cascaded aggregation by keeping the factors which influence primacy and inertia effects constant. An experiment to test these assumptions would not be difficult to devise.

The possibility that subjects respond to factors in a cascaded inference situation which are irrelevant to the optimal model has implications for both future efforts to predict subjective opinion revision and attempts to compare intuitive and optimal inference. Experiments which take into account only those parameters and inputs required by the optimal model may lead to unnecessarily pessimistic conclusions about human inference capabilities in a complex real world situation. Obviously then any model which attempts to accurately describe the intuitive inference process must include all elements of the stimulus situation to which people respond.

REFERENCES

- Dodson, J. J. Simulation system design for a TEAS simulation research facility. Report No. 4, November 1961, AFCRL 1112, PRC R-194, Planning Research Corporation, Los Angeles, California.
- Du Charme, W. M. Response bias explanation of conservative human inference. Journal of Experimental Psychology, 1970, 85, 66-74.
- Gettys, C. F., Kelley, C. W., & Peterson, C. R. Best guess hypothesis in multi-stage inference. Organizational Behavior and Human Performance, 1973, 10, 364-373.
- Gettys, C. F., & Wilke, T. A. The application of Bayes' theorem when the true data state is uncertain. Organizational Behavior and Human Performance, 1969, 4, 125-141.
- Peterson, C. R., & Swensson, R. L. Intuitive statistical inferences about diffuse hypotheses. Organizational Behavior and Human Performance, 1968, 3, 1-11.
- Peterson, C. R., & Du Charme, W. M. A primacy effect in subjective probability revision. Journal of Experimental Psychology, 1967, 73, 61-65.
- Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.
- Peterson, C. R., & Miller, A. J. Sensitivity of subjective probability revision. Journal of Experimental Psychology, 1965, 70, 117-121.
- Phillips, L. D., & Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-354.
- Pitz, G. F. An inertia effect (resistance to change) in the revision of opinion. Canadian Journal of Psychology, 1969, 23, 24-33.
- Schaeffer, R. E., & Borcharding, K. A note on the consistency between two approaches to incorporate data from unreliable sources in Bayesian analysis. Organizational Behavior and Human Performance, 1973, 9, 504-508.
- Schum, D. A., & Du Charme, W. M. Comments on the relationship between the impact and the reliability of evidence. Organizational Behavior and Human Performance, 1971, 6, 111-131.

- Schum, D. A., Du Charme, W. M., & De Pitts, K. E. Research on human multi-stage probabilistic inference processes. Organizational Behavior and Human Performance, 1973, 10, 318-348.
- Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.
- Snapper, K. J., & Fryback, D. G. Inference based on unreliable reports. Journal of Experimental Psychology, 1971, 87, 401-404.
- Youssef, Z. I. The effects of cascaded inference on the subjective value of information. Organizational Behavior and Human Performance, 1973, 10, 359-363.
- Youssef, Z. I., & Peterson, C. R. Intuitive cascaded inference. Organizational Behavior and Human Performance, 1973, 10, 349-358.

FOOTNOTES

1. The symbol r will be used throughout this paper to indicate source reliability in the special case in which source reliability is the same for a report of either binomial event. For a more general discussion of reliability see Schum and Du Charme (1970).
2. The term "nonlinearity" is not used in this paper in its statistical sense but merely refers to the fact that the data points do not all fall on a straight line.