AD-A048 569 DECISIONS AND DESIGNS INC MCLEAN VA F/G 5/11
DEVELOPING THE TECHNOLOGY OF PROBABILISTIC INFERENCE: AGGREGATI--ETC(U)
AUG 77 L C EILS, D A SEAVER, W EDWARDS N00014-76-C-0074
UNCLASSIFIED NL

1 of 1
ADA048 569

END
DATE
FILMED
2-78
DDC

FG.

(12)

001855-3-T

USC

UNIVERSITY OF SOUTHERN CALIFORNIA

social science research institute

AD A048569

DDC FILE COPY

AD No. ____

RESEARCH REPORT

Developing the Technology of Probabilistic Inference: Aggregating by Averaging Reduces Conservatism

Lee C. Eils, III
David A. Seaver
Ward Edwards

Sponsored by:
Advanced Research Projects Agency
Department of Defense

Monitored by:
Engineering Psychology Programs
Office of Naval Research
Contract No. N00014-76-C-0074, ARPA

DDC
JAN 20 1978
F

Approved for Public Release;
Distribution Unlimited;
Reproduction in Whole or in Part is Permitted
for Any Use of the U.S. Government

August 1977

SSRI Research Report 77-3 ✓

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of the Advanced Research Projects Agency of the U.S. Government.

6

DEVELOPING THE TECHNOLOGY OF PROBABILISTIC INFERENCE: AGGREGATING BY AVERAGING REDUCES CONSERVATISM.

9 Technical rept. Oct 76 - Sep 77,

Research Report 77-3

Aug[redacted]77

11

12 31p.

DDC
JAN 20 1978
F

10
Lee C./Eils, III,
David A./Seaver
Ward/Edwards

Social Science Research Institute
University of Southern California

15 N00014-76-C-0074

Sponsored by

Defense Advanced Research Projects Agency

1473

390664

ACCESSION for
NTIS
DDC
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist.
A

Summary

A relatively large body of research indicates that people are conservative processors of probabilistic information. Recent attention has focussed on two possible explanations of this phenomenon. The misaggregation hypothesis depicts conservatism as an inability to properly combine the information in sequences of data. The other explanation suggests conservatism is the result of a response bias: the avoidance of extreme odds or probability judgments.

Two experiments explored the use of a specific response, average certainty, that was devised to thwart conservatism caused by either a response bias or a certain form of misaggregation. Use of appropriate instructions and response scales made the average certainty judgments good subjective assessments of the arithmetic mean likelihood ratio which could then be used in the appropriate form of Bayes' Theorem to calculate posterior odds. These judgments seemed unlikely to be affected by a response bias since extreme responses were not needed. In addition, research has suggested that people are more likely to aggregate information by averaging than by adding or multiplying, so misaggregation may be exhibited only in specific forms of aggregation and may not be present in averaging.

The results of Experiment I indicated that average certainty judgments were both more orderly and more veridical than cumulative certainty judgments of the type usually obtained in probabilistic inference tasks. The cumulative judgments were very conservative while the average certainty judgments were only slightly radical. Experiment II indicated that average certainty judgments and individual likelihood ratio judgments were

both more orderly and veridical than cumulative certainty judgments but that they did not differ significantly from each other in either orderliness or veridicality. A second factor, the diagnosticity level of the data was also found to influence the veridicality of obtained judgments. Regardless of the method of aggregation employed, estimates became more veridical as the data became more diagnostic. Since these studies were undertaken only to see if average certainty judgments are an effective way to reduce conservatism, they do not directly test what causes conservatism. However, some implications concerning the nature of conservatism are discussed, as are the implications for the technology of probabilistic inference.

Contents

Figures

Tables

Acknowledgment

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under Contract N00014-76-C-0074 and subcontract from Decisions and Designs, Inc.

Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.

I. Introduction

A relatively large body of literature (see reviews by DuCharme, 1969; Slovic and Lichtenstein, 1971) supports the assertion that people are conservative processors of probabilistic information. When presented with data, individuals typically revise their opinion to an extent less than prescribed by Bayes' Theorem, the formally appropriate model. In its simplest form with two hypotheses and conditionally independent data, Bayes' Theorem takes the form:

$$\Omega_n = \Omega_0 \prod_{i=1}^{n} L_i \qquad (1)$$

where Ω_0 is the prior odds, L_i is the likelihood ratio for the ith datum, and Ω_n is the revised (posterior) odds.

Three hypotheses have been advanced to explain conservative human inference (Edwards, 1968). The misaggregation hypothesis depicts conservatism as an inability to properly combine the information present in sequences of data, although each single datum is judged accurately. A second hypothesis--the misperception hypothesis--argues that people aggregate information properly (that is, according to Bayes' Theorem), but incorrectly diagnose the information in a single datum. The final explanation considers conservatism the result of a response bias: the avoidance of extreme odds or probability judgments.

Recent attention has focussed primarily on the misaggregation and response bias hypotheses, since Wheeler and Edwards (1975) demonstrated rather convincingly that misperception played little, if any, role in producing conservatism. DuCharme (1970) provides the main support for the response bias explanation. He found aggregated posterior odds judgments to be near veridical for non-extreme odds, while unaggregated likelihood ratio judgments were conservative when the true likelihood ratio was relatively extreme. Wheeler and Edwards refute some of DuCharme's findings, providing persuasive support for the misaggregation hypothesis. Still they conclude that

> "in all likelihood misperception, misaggregation, and response biases all contribute to conservatism. The real questions of importance then becomes finding the manner in which each phenomenon contributes to conservatism and the best way of avoiding or compensating for this non-optimal behavior." (p. 10)

One strategy that could lead to reduction or elimination of the bias against extreme responses is to remove the need for extreme responses: that is, to find some type of judgment that conveys the necessary information but for which the numerical response is not extreme. Consideration of these requirements for posterior odds judgments suggests some sort of average judgment. The responses required for average judgments have the advantage that the aggregated judgment falls within the range of the inputs to the judgment, as opposed to the usual posterior odds judgments in which the veridical response will involve multiplication of likelihood ratios, typically making the response much larger than the inputs.

Reviewing alternative forms of Bayes' Theorem (equation 1) suggests two possible types of judgments that convey all the information necessary and that are also averages of some sort. The geometric mean of the likelihood ratios is one possibility. However, geometric means seem to be a difficult judgment for people to make. In addition, if subjects substituted arithmetic means for the geometric means--a seemingly likely occurrence--the results would be biased away from conservatism since the arithmetic mean is always larger than the geometric mean.

Fortunately, a second possibility exists. Taking base ten logarithms in equation (1) yields

$$\log \Omega_n = \log \Omega_0 + \sum_{i=1}^{n} \log L_i \tag{2}$$

The judgment that would play a role in equation (2) equivalent to that played by the geometric mean likelihood ratio is the arithmetic mean log *likelihood ratio. If subjects assess arithmetic* mean log likelihood ratios, AMLL's, the posterior odds will be

$$\Omega_n = \Omega_0 10^{n(\mathrm{AMLL})} \tag{3}$$

A review of descriptive studies exploring how people process information lends additional support to the use of the arithmetic mean log likelihood ratio as a normative procedure for processing probabilistic information. People have been shown to use averaging rather than adding or multiplying as a method for combining information in a wide variety of contexts, for example in determining the overall value of products (Troutman and Shanteau, 1976), in deciding how well they would like a person described by personality trait adjectives (Andersen, 1965), and in predicting a criterion number on the basis of two cue numbers (Lichtenstein, Earle, and Slovic, 1975).

Probably the most relevant descriptive studies are those of Shanteau (1975) and Troutman and Shanteau (1977) in which they used a task very similar to that employed in conservatism experiments. In both studies, an averaging model provided a much better fit to the data than did a multiplying model, which is the appropriate model if subjects actually process information in a manner consistent with Bayes' Theorem.

Certainly these descriptive results should be considered in designing normative procedures. If indeed, as seems likely, people do tend to average information, then a normative procedure taking advantage of this tendency may produce better results than one which requires an alternative form of processing. Perhaps conservatism results from the specific manner in which subjects are required to aggregate information. That is, people may not be accustomed to using addition or multiplication as aggregation procedures for processing information.

The present experiments explore the possibility of using subjective judgments of the arithmetic mean log likelihood ratio to reduce conservatism in aggregated judgments of certainty. These aggregated judgments are also compared against unaggregated likelihood ratio judgments.

II. Experiment I

II.1. Method

II.1.1. Apparatus. The stimuli for Experiment I were 6.5 inch (16.51 cm) pick-up sticks painted yellow on one end and blue on the other. The length of yellow (or, because of symmetry, the length of blue) was the random variable. The sticks shown subjects were hypothetically drawn from two normally distributed populations differing only in mean length of yellow. One population (the predominantly yellow) had a mean of 4.05 inches (10.287 cm) of yellow and the other (the predominantly blue) a mean of 2.45 inches (6.223 cm). The populations shared a common standard deviation of 1.0 inches (2.54 cm).

Each stick was displayed on a white rectangular card. Eight sequences of four sticks were randomly selected for use as stimuli. Half the sequences were drawn from the predominantly blue population while the other half were drawn from the predominantly yellow population, so prior odds were always 1:1.

The population characteristics were displayed to the subjects by means of two random histograms, each representing a sample of 99 sticks from one of the two populations. The lengths displayed had been carefully chosen to accurately represent the populations. The displays were actual size and colors, and, on each, the population mean was displayed by a heavy black horizontal line at the appropriate position. Both displays were present throughout the experiment. All responses were made on sheets of paper which contained space at the top for the subject to check the more likely population along with four logarithmically spaced scales ranging from 1:1 odds (designated the "uncertainty" point) to odds of 10,000:1 and 1:10,000.

II.1.2. Subjects. Twenty naive subjects, run in groups of three or four, were drawn from an introductory undergraduate psychology course taught at the University of Southern California.

II.1.3. Procedure. Subjects were randomly assigned to one of two groups: one group made average certainty judgments, while the other judged cumulative certainty. For the group judging average certainty, the subjects were presented with the first stick and told to indicate, at the top of the scale, their choice of the more likely population, after which they were to designate, on the first scale column, the odds corresponding to their subjective certainty. As the second stick was placed alongside the first, subjects were told that their responses should represent a judgment as to the "average of, rather than the total of, their certainty". Each subject was told to use, as a reference in responding, the position of the point on the first scale column. Detailed examples emphasizing the process that should take place in arriving at an average certainty judgment were provided. Subjects were told that this procedure would continue with the addition of the third and fourth sticks, after which a new response sheet would be provided for a new sequence of four sticks. Each subject was also told that the same process was employed, but that direction changed toward the bottom half of the scale in the event that the subject changed his or her belief about which population was more likely to have produced the sticks in view.

The instructions given to each subject in the group judging cumulative certainty differed from those described above in one important way. As the second stick was placed alongside the first, the subject was asked to represent his or her cumulative certainty on the second scale column. This cumulative certainty was to be judged relative to the certainty after seeing the previous stick and was to be represented by moving distance (away from the "uncertainty" line) on the second scale column. Examples of this procedure were provided in the instructions.

Having read through one set of instructions with the subjects, the experimenter provided four sequences of three sticks each as examples. Subjects responded to the example stimuli and general feedback was provided. Then eight sequences of four sticks were presented and 32 point placements were made on eight sets of four scale columns.

II.2. Results and Discussion

The data were first subjected to a logarithmic transformation and the average certainty judgment responses in logarithmic form were each multiplied by n, the number of sticks on which the judgment was based. All analyses were performed on the log transformed responses and the dependent variable was the log posterior odds that were inferred from the subjects' responses to the sequences of sticks.

For each subject, a regression analysis of inferred log posterior odds on veridical log posterior odds was performed. The correlation coefficients and slopes from these analyses are presented in Table 1. Judgments in the average certainty condition were significantly more orderly ($t(18) = 3.975$ $p < .001$ on Fisher-z transformed correlation coefficients) as reflected in the mean correlation coefficients of .896 and .689 for the average and cumulative certainty response groups respectively. Posterior odds inferred from the averaging condition tended to be slightly radical (mean slope = 1.264), while the posterior odds obtained from the cumulative condition were extremely conservative (mean slope = 0.283). This difference was significant, $\underline{t}\ (18) = 6.115$, $p < .001$, in the predicted direction. Although average certainty judgments were slightly radical, they were much closer to veridical than the cumulative certainty judgments.

TABLE 1

Correlations and Slopes for Inferred Posterior Odds

	Average Certainty		Cumulative Certainty	
	r	b	r	b
	.882	1.068	.598	0.164
	.934	1.543	.809	0.371
	.922	0.951	.898	0.509
	.808	0.900	.864	0.356
	.925	0.501	.748	0.329
	.934	0.898	.415	0.330
	.871	1.899	.479	0.108
	.909	1.463	.601	0.147
	.882	2.110	.571	0.155
	.896	1.295	.906	0.358
mean	.896	1.264	.689	0.283
standard deviation	.038	0.496	.179	0.131

The results of this study were encouraging for the use of average certainty judgments as a means of assessing subjective certainty. Subjects apparently can make this type of judgment and, at least under the conditions of this experiment, make such judgments quite well. Since the feasibility of this type of judgment seems to have been established, further questions need to be pursued. Specifically, do these results generalize to other levels of data diagnosticity? And how does the veridicality of average certainty judgments compare with that of non-aggregated likelihood ratio judgments? These questions are addressed in Experiment II.

III. Experiment II

III.1. Method

The method used in Experiment II was the same as that used in Experiment I with two exceptions: three levels of data diagnosticity were used, and a third type of uncertainty judgment--single-datum likelihood ratios--was also examined. The two pairs of normal distributions used, in addition to the original pair had mean yellow lengths of 4.35 inches (11.149 cm) and 3.75 inches (9.525 cm) for the predominantly yellow populations, and 2.15 inches (5.461 cm) and 2.75 inches (6.985 cm) for the predonimantly blue populations. Thus, the three levels of data diagnosticity, $d' = (m_1 - m_2)/\sigma$, were 2.2, 1.6, and 1.0.

A three by three factorial design was created by crossing the three types of uncertainty judgments with the three levels of data diagnosticity. Ninety subjects, ten per cell, were randomly assigned to the experimental conditions and were run in groups of four or five. A few subjects were run individually to obtain equal cell sizes. The subjects were from an introductory psychology course at the University of Southern California in which participation in several experiments was required.

III.2. Results

A regression analysis was performed on the individual data of each subject. For subjects making average or cumulative certainty judgments, inferred log posterior odds were regressed on veridical log posterior odds, while for subjects making likelihood ratio judgments, subjective likelihood ratios were regressed on veridical likelihood ratios. The correlation coefficients and slopes from these analyses are presented in Table 2.

Results of analyses of variance performed on the slopes and Fisher-z transformed correlation coefficients are shown in Tables 3 and 4 respectively. Inspection of the cell means of the slopes, plotted in Figure 1(a), indicated little difference between the likelihood ratio judgments (mean = 1.359) and the average certainty judgments (mean = 1.366), both being slightly radical. The cumulative certainty judgments, however, were again extremely conservative (mean = .365). Slopes generally tended to decrease as d' increased.

The significant interaction was unexpected, but may be due to a problem on the first day of experimentation. Elimination of the five subjects run the first day (average certainty judgments at d' = 1.6) eliminated the interaction as illustrated in Figure 1(b). These five subjects all had higher slopes than subsequent subjects run in this condition. Perhaps these subjects did not correctly understand the nature of the judgments they were asked to make.

TABLE 2

Correlations and Slopes for Average Certainty, Cumulative Certainty, and Likelihood Ratio Judgments

d' = 1.0

	Average Certainty		Cumulative Certainty		Individual Likelihood Ratio	
	r	b	r	b	r	b
	.498	1.194	.893	0.834	.741	1.878
	.351	1.449	.613	0.420	.831	1.800
	.844	2.538	.852	0.565	.778	1.985
	.566	1.443	.640	0.282	.900	2.585
	.863	2.143	.913	0.726	.828	1.976
	.240	0.461	.870	0.385	.753	1.188
	.842	2.414	.838	0.518	.850	1.811
	.819	1.560	.752	0.407	.464	1.038
	.908	1.291	.729	0.507	.896	2.052
	.761	1.586	.712	0.310	.763	1.352
mean	.669	1.608	.781	.495	.780	1.767
standard deviation	.238	.619	.107	.176	.125	.459

TABLE 2
(continued)

Correlations and Slopes for Average Certainty,

Cumulative Certainty, and Likelihood Ratio Judgments

d' = 1.6

	Average Certainty		Cumulative Certainty		Individual Likelihood Ratio	
	r	b	r	b	r	b
	.822	2.583	.931	0.582	.572	0.623
	.932	2.606	.576	0.055	.797	1.335
	.928	1.788	.716	0.250	.855	1.110
	.773	1.580	.749	0.237	.642	0.877
	.912	2.051	.739	0.239	.811	1.780
	.864	0.965	.575	0.139	.901	2.120
	.939	0.667	.758	0.405	.904	1.703
	.914	0.877	.393	0.099	.819	1.034
	.860	0.948	.877	0.256	.751	-0.333
	.848	1.341	.646	0.160	.875	1.110
mean	.879	1.541	.696	.242	.793	1.136
standard deviation	.055	.705	.156	.155	.110	.687

TABLE 2
(continued)

Correlations and Slopes for Average Certainty,
Cumulative Certainty, and Likelihood Ratio Judgments
d' = 2.2

	Average Certainty		Cumulative Certainty		Individual Likelihood Ratio	
	r	b	r	b	r	b
	.776	0.785	.813	0.587	.841	1.542
	.830	0.643	.689	0.355	.901	1.156
	.813	0.915	.838	0.464	.929	1.036
	.896	0.964	.881	0.472	.927	1.328
	.813	1.359	.641	0.267	.959	1.374
	.889	1.107	.750	0.412	.938	1.335
	.701	1.197	.564	0.210	.528	0.601
	.950	0.869	.786	0.364	.799	1.033
	.676	0.770	.608	0.113	.890	1.155
	.847	0.914	.779	0.323	.806	1.195
mean	.819	.952	.735	.357	.852	1.176
standard deviation	.085	.215	.105	.138	.127	.257

TABLE 3

ANOVA of Slopes

Source	Sums of Squares	df	Mean Squares	p
Within Cells	15.731	81	0.194	
Aggregation Method	19.931	2	9.996	0.001
Diagnosticity Level	3.346	2	1.673	0.001
Aggregation x Diagnosticity	2.070	4	0.518	0.038

TABLE 4

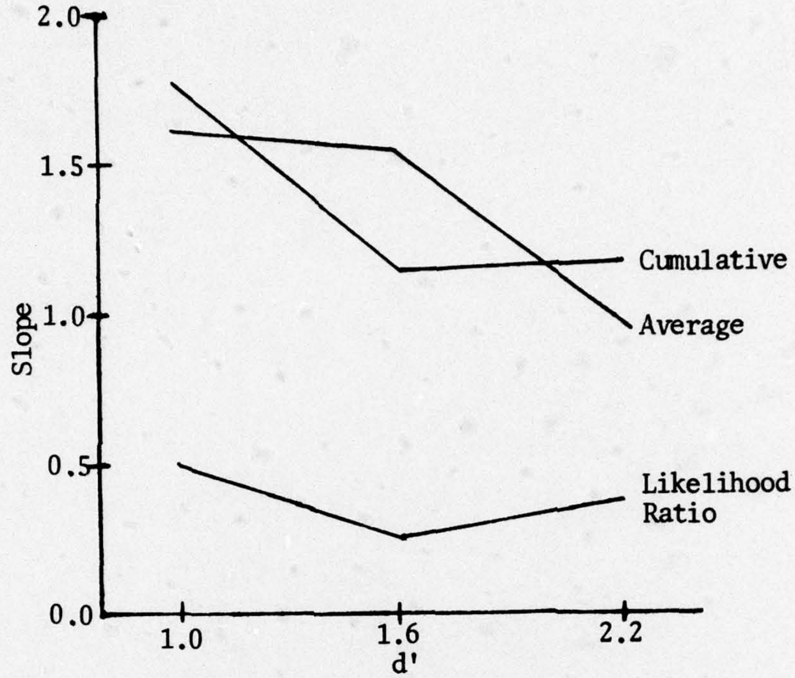
ANOVA of Correlations

Source	Sums of Squares	df	Mean Squares	p
Within Cells	8.224	81	0.102	
Aggregation Method	0.746	2	0.373	0.030
Diagnosticity Level	0.401	2	0.201	0.146
Aggregation x Diagnosticity	1.540	4	0.385	0.007

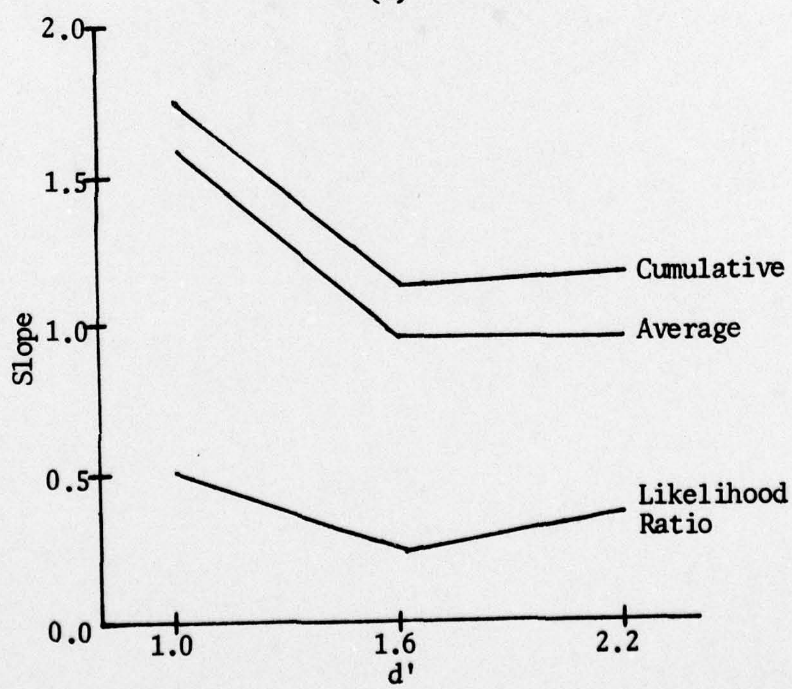
The average certainty and likelihood ratio judgments also appeared to be more orderly than the cumulative certainty judgments as measured by the correlations. Data diagnosticity did not significantly affect the correlations.

IV. Discussion

The results of these experiments suggest that aggregated judgments of



(a)



(b)

Figure 1: Mean Slopes

certainty need not be conservative. Past results have invariably shown cumulative judgments to be conservative, but experimental procedures have generally allowed subjective judgments to be veridical only if the aggregation method was multiplication. In some cases, proper use of the response scale could make the aggregation additive. For example, if the response scale presents logarithmically spaced odds, the normative aggregation method will be additive in distances on the scale.

However, in this study use of averaging as an aggregation method was demonstrated to be a viable method that could be taught to subjects and not result in conservatism. Specifically, subjects were asked to make arithmetic mean log likelihood ratio judgments, although of course, these precise words were never used. The posterior odds resulting from the judgments were very near veridical; while posterior odds, judged in the usual way by asking for cumulative certainty, were very conservative.

Either or both of two factors may account for the lack of conservatism in the average judgments. The aggregated responses necessary for the average judgments fall within the range of the single datum likelihood ratios that are inputs to the aggregation, and therefore, are not extreme responses. Thus, the bias against extreme responses may be reduced or eliminated. The second factor is simply the aggregation process that subjects need to employ to be veridical. In many situations people have been shown to use averaging rather than some other method of aggregating multiple pieces of information. In particular, the work by Shanteau (Shanteau, 1975; Troutman and Shanteau, 1977) has shown that people average the information in multiple samples of data on a task very similar to the one used in this study.

Since averaging appears to be the natural method of aggregation found in these descriptive studies, it seems reasonable that normative information processing based on averaging would outperform other methods of aggregation, a hypothesis consistent with the results of this study.

These findings have implications for the further development of the technology of inference. Because people typically have proved to be conservative processors of information, researchers have looked for methods of obtaining the desired information from people without the elicited judgments being affected by conservatism. Probabilistic information processing (PIP) systems were developed for this specific purpose (Edwards, Phillips, Hays, and Goodman, 1968). In a PIP system, people make only likelihood ratio judgments, a task this study and others (Wheeler and Edwards, 1975) show they can perform quite well. Bayes' Theorem is then used to combine the likelihood ratio judgments in the proper manner to produce posterior odds.

Since people are able to judge likelihood ratios quite accurately, why even consider average certainty judgments as an alternative? This study certainly did not show that average certainty judgments are superior to likelihood ratio judgments. The reason for considering an alternative to likelihood ratio judgments is that a problem may arise in applying PIP systems in real world contexts. The people assessing the likelihood ratios will typically have access to feedback about the posterior odds that are calculated from their likelihood ratios. Goodman (1973), in a reanalysis of data from five studies exploring methods of eliciting judgments about uncertain events, concludes that feedback about the implications of judgments makes them less extreme and is probably the most powerful variable controlling the extremeness of the judgments. Thus, even a PIP system may be susceptible to conservatism in real-world applications. This problem seems less likely to characterize judgments of average certainty due to the very nature of the elicited judgments. Should further research confirm feedback produced conservatism in PIP systems, average certainty judgments may prove to be a useful alternative to PIP.

V. References

Anderson, N.H., Averaging versus adding as a stimulus-combination rule in impression formation. Journal of Experimental Psychology, 1965, 70, 394-400.

Domas, P.A., Goodman, B.C., and Peterson, C.R. Bayes' Theorem: Response scales and feedback. Technical Report No. 037230-5-T, Engineering Psychology Laboratory, University of Michigan, September, 1972.

DuCharme, W.M. A review and analysis of the phenomenon of conservatism in human inference. Applied Mathematics and Systems Theory Technical Report No. 46-5, Rice University, December, 1969.

DuCharme, W.M. Response bias explanation of conservative human inference. Journal of Experimental Psychology, 1970, 85, 66-74.

Edwards, W. Conservatism in human information processing. In Kleinmuntz, B. (Ed.), Formal Representation of Human Judgments. New York: Wiley, 1968.

Edwards, W., Phillips, L.D., Hays, W.L., and Goodman, B.C. Probabilistic information processing systems: Design and evaluation. IEEE Transactions on Systems Science and Cybernetics, 1968, SSC-4, 248-265.

Goodman, B.C. Direct estimation procedures for eliciting judgments about uncertain events. Technical Report No. 011313-5-T, Engineering Psychology Laboratory, University of Michigan, November, 1973.

Lichtenstein, S., Earle, T.C., and Slovic, P. Cue utilization in a numerical prediction task. Journal of Experimental Psychology, 1975, 104, 77-85.

Phillips, L.D., and Edwards, W. Conservatism in a simple probability inference task. Journal of Experimental Psychology, 1966, 72, 346-357.

Shanteau, J. Averaging versus multiplying combination rules of inference judgment. Acta Psychologica, 1975, 39, 83-89.

Slovic, P., and Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Troutman, C.M., and Shanteau, J. Do consumers evaluate products by adding or averaging attribute information? Human Information Processing Report No. 75-4 (revised), Kansas State University, January, 1976.

Troutman, C.M. and Shanteau, J. Inferences based on nondiagnostic information. Organizational Behavior and Human Performance, 1977, 19, 43-55.

Wheeler, G.E., and Edwards, W. Misaggregation explains conservative inference about normally distributed populations. Research Report 75-11, Social Science Research Institute, University of Southern California, August, 1975.

Research Distribution List

Department of Defense

Assistant Director (Environment and Life Sciences)
Office of the Deputy Director of Defense Research and Engineering (Research and Advanced Technology)
Attention: Lt. Col. Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Office of the Assistant Secretary of Defense (Intelligence)
Attention: CDR Richard Schlaff
The Pentagon, Room 3E279
Washington, DC 20301

Director, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Program Management Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209
(two copies)

Administrator, Defense Documentation Center
Attention: DDC-TC
Cameron Station
Alexandria, VA 22314
(12 copies)

Department of the Navy

Office of the Chief of Naval Operations (OP-987)
Attention: Dr. Robert G. Smith
Washington, DC 20350

Director, Engineering Psychology Programs (Code 455)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
(three copies)

Assistant Chief for Technology (Code 200)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 436)
Attention: Dr. Bruce McDonald
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (ONR)
International Programs (Code 1021P)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye and Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106
(two copies)

Dr. M. Bertin
Office of Naval Research
Scientific Liaison Group
American Embassy – Room A-407
APO San Francisco 96503

Director, Naval Research Laboratory
Technical Information Division (Code 2627)
Washington, DC 20375
(six copies)

Director, Naval Research Laboratory (Code 2029)
Washington, DC 20375
(six copies)

Scientific Advisor
Office of the Deputy Chief of Staff
for Research, Development and Studies
Headquarters, U.S. Marine Corps
Arlington Annex, Columbia Pike
Arlington, VA 20380

Headquarters, Naval Material Command (Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Headquarters, Naval Material Command (Code 0344)
Attention: Mr. Arnold Rubinstein
Washington, DC 20360

Naval Medical Research and Development Command (Code 44)
Naval Medical Center
Attention: CDR Paul Nelson
Bethesda, MD 20014

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Naval Personnel Research and Development Center
Management Support Department (Code 210)
San Diego, CA 92152

Naval Personnel Research and Development Center (Code 305)
Attention: Dr. Charles Gettys
San Diego, CA 92152

Dr. Fred Muckler
Manned Systems Design, Code 311
Navy Personnel Research and Development
Center
San Diego, CA 92152

Human Factors Department (Code N215)
Naval Training Equipment Center
Orlando, FL 32813

Training Analysis and Evaluation Group
Naval Training Equipment Center
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Department of the Army

Technical Director, U.S. Army Institute for the Behavioral and Social Sciences
Attention: Dr. J.E. Uhlaner
1300 Wilson Boulevard
Arlington, VA 22209

Director, Individual Training and Performance Research Laboratory
U.S. Army Institute for the Behavioral and
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Director, Organization and Systems Research Laboratory
U.S. Army Institute for the Behavioral and
Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Department of the Air Force

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Robert G. Gough, Major, USAF
Associate Professor
Department of Economics, Geography and
Management
USAF Academy, CO 80840

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Aerospace Medical Division (Code RDH)
Attention: Lt. Col. John Courtright
Brooks AFB, TX 78235

Other Institutions

The Johns Hopkins University
Department of Psychology
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Director, Human Factors Wing
Defense and Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsville, Toronto
Ontario, Canada

Stanford University
Attention: Dr. R.A. Howard
Stanford, CA 94305

Montgomery College
Department of Psychology
Attention: Dr. Victor Fields
Rockville, MD 20850

General Research Corporation
Attention: Mr. George Pugh
7655 Old Springhouse Road
McLean, VA 22101

Oceanautics, Incorporated
Attention: Dr. W.S. Vaughan
3308 Dodge Park Road
Landover, MD 20785

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A.D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Catholic University
Attention: Dr. Bruce M. Ross
Washington, DC 20017

Stanford Research Institute
Decision Analysis Group
Attention: Dr. Allan C. Miller III
Menlo Park, CA 94025

Human Factors Research, Incorporated
Santa Barbara Research Park
Attention: Dr. Robert R. Mackie
6780 Cortona Drive
Goleta, CA 93017

University of Washington
Department of Psychology
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Eclectech Associates, Incorporated
Post Office Box 179
Attention: Mr. Alan J. Pesch
North Stonington, CT 06359

Hebrew University
Department of Psychology
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. T. Owen Jacobs
Post Office Box 3122
Ft. Leavenworth, KS 66027

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 001855-3-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Developing the Technology of Probabilistic Inference: Aggregating by Averaging Reduces Conservatism		5. TYPE OF REPORT & PERIOD COVERED Technical 10/76 - 9/77
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Lee C. Eils, III, David A. Seaver, and Ward Edwards		8. CONTRACT OR GRANT NUMBER(s) Prime Contract N00014-76-C-0074 Subcontract 76-0308-0715
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Virginia 22209		12. REPORT DATE August, 1977
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Decisions and Designs, Inc. Suite 100, 7900 Westpark Drive McLean, Virginia 22101 (under contract from Office of Naval Research)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Conservatism	Posterior odds
Probabilistic inference	Likelihood ratio
Bayes' Theorem	Arithmetic mean log likelihood ratio
Misaggregation	Probabilistic information processing system
Response bias	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A relatively large body of research indicates that people are conservative processors of probabilistic information. Recent attention has focussed on two possible explanations of this phenomenon. The misaggregation hypothesis depicts conservatism as an inability to properly combine the information in sequence of data. The other explanation suggests conservatism is the result of a response bias: the avoidance of extreme odds or probability judgments.

Two experiments explored the use of a specific response, average certainty

next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) cont that was devised to thwart conservatism caused by either a response bias or a certain form of misaggregation. Use of appropriate instructions and response scales made the average certainty judgments good subjective assessments of the arithmetic mean likelihood ratio which could then be used in the appropriate form of Bayes' Theorem to calculate posterior odds. These judgments seemed unlikely to be affected by a response bias since extreme responses were not needed. In addition, research has suggested that people are more likely to aggregate information by averaging than by adding or multiplying, so misaggregation may be exhibited only in specific forms of aggregation and may not be present in averaging.

The results of Experiment I indicated that average certainty judgments were both more orderly and more veridical than cumulative certainty judgments of the type usually obtained in probabilistic inference tasks. The cumulative judgments were very conservative while the average certainty judgments were only slightly radical. Experiment II indicated that average certainty judgments and individual likelihood ratio judgments were both more orderly and veridical than cumulative certainty judgments but that they did not differ significantly from each other in either orderliness or veridicality. A second factor, the diagnosticity level of the data was also found to influence the veridicality of obtained judgments. Regardless of the method of aggregation employed, estimates became more veridical as the data became more diagnostic. Since these studies were undertaken only to see if average certainty judgments are an effective way to reduce conservatism, they do not directly test what causes conservatism. However, some implications concerning the nature of conservatism are discussed, as are the implications for the technology of probabilistic inference.