LEVEL II

AD A072678

DECISION PROCESSES LABORATORY
UNIVERSITY OF OKLAHOMA

OFFICE OF NAVAL RESEARCH
CONTRACT NUMBER N00014-77-C-0615
WORK UNIT NUMBER NR197-040

12

The Availability Explanation

of Excessive Plausibility Assessments

Thomas Mehle, Charles Gettys, Carol Manning,

Suzanne Baca and Stanley Fisher

TR 30-7-79 July 1979

prepared for

Office of Naval Research
Engineering Psychology Programs
Contract Number N00014-77-C-0615
Work Unit Number NR 197-040

Decision Processes Laboratory
Department of Psychology
University of Oklahoma
Norman, Oklahoma 73019

DISCLAIMER NOTICE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
14
TR-30-7-79

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6
THE AVAILABILITY EXPLANATION OF EXCESSIVE PLAUSIBILITY ASSESSMENTS.

5. TYPE OF REPORT & PERIOD COVERED
9
Rept. for
Aug. [redacted]78-Aug. [redacted]79

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
10
Thomas Mehle, Charles Gettys, Carol Manning, Suzanne Baca [redacted] Stanley Fisher

8. CONTRACT OR GRANT NUMBER(s)
15
N00014-77-C-0615

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Decision Processes Laboratory
Department of Psychology
University of Oklahoma
Norman, Oklahoma 73019

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR 197-040

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Engineering Psychology Programs-Code 455
800 N. Quincy
Arlington, Virginia 22217

11
30

12. REPORT DATE
Jul[redacted]79

13. NUMBER OF PAGES
28

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
12
37p.

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)

1) decision theory
2) hypothesis generation
3) heuristics
4) excessive assessments

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

The assessment of hypotheses in hypothesis generation involves a comparison between those hypotheses that have been generated (specified) and those that are not generated (unspecified). This study investigated the "availability

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)

410918

VB

Block 20 (continued).

explanation" (Tversky and Kahneman, 1973) for subjects' overconfidence in estimating the probability of specified hypotheses. The conjecture is that subjects have difficulty retrieving unspecified hypotheses; a complete set of candidate unspecified hypotheses is unavailable during assessment. Therefore, the underpopulated set of unspecified hypotheses is regarded as less probable and the specified set is regarded as more probable. A control group in this study replicated previous findings of overconfidence for specified hypotheses. Two manipulations to increase the availability of unspecified hypotheses were investigated. One manipulation involved explicitly requesting subjects to populate the unspecified set. The other manipulation consisted of computer presentation of candidate unspecified hypotheses. Although in a normative sense, neither manipulation should have affected judgements, results indicated that assessment overconfidence for both experimental groups was reduced. These results support our conjecture that the availability heuristic is at least partially responsible for subjects' excessive behavior in evaluating specified hypotheses.

The availability explanation of excessive plausibility assessments

A necessary precursor to any decision analysis is an identification of possible hypotheses to be considered, a process we term "hypothesis generation." This process involves a partition of all possible hypotheses appropriate for the problem into two sets, the set of "specified" (generated) hypotheses and the complement of this set, the set of "unspecified" hypotheses. The result of a previous study, (Gettys, Fisher and Mehle, 1978) was that subjects were overconfident in assessing sets of specified hypotheses and underconfident in assessing sets of unspecified hypotheses. In this previous study, and in the current study, subjects estimated their feelings of certainty by judging the odds of three specified possible majors of an unknown undergraduate student at the University of Oklahoma and a fourth "catch-all" possibility corresponding to the alternative that the unknown student had some other major. The data for these problems were classes that the unknown student had taken. The veridical values were obtained by analyzing the computerized student master record file at the University of Oklahoma. A magnitude estimation procedure was used to obtain the subjects' responses.

In other contexts, the overconfidence bias has received considerable attention recently: Lichtenstein, Fischhoff and Phillips (1977) review several studies which address this issue. Kahneman and Tversky (in press) listed lack of expertise, insensitivity to the quality of data, oversensitivity to data consistency, conditionality (adopting unstated assumptions) and anchoring as contributors to the cverconfidence bias.

The purpose of the present study was to investigate a factor which may contribute to overconfidence in hypothesis-generation tasks, the "availability heuristic." We postulate that subjects may have underestimated the likelihood of the catch-all alternative in the Gettys et al. (1978) study simply because they had difficulty populating the catch-all alternative with hypotheses. Since some catch-all hypotheses would not be available and thus not evaluated when making certainty estimates, subjects would tend to underestimate the likelihood of catch-all sets.

This "availability" explanation of subjects' excessive odds estimates of specified hypotheses is related to, but not identical to, the availability heuristic described by Tversky and Kahneman (1973, 1974). Tversky and Kahneman reported evidence that subjects were influenced by the availability in memory of instances of an event when evaluating the probability of that event. In the context of hypothesis generation, subjects must be able to judge the likelihood of the set of all unspecified hypotheses in order to accurately assess the likelihood of the complement of this set, the set of specified hypotheses. If subjects simply cannot recall many of the hypotheses of the unspecified set, it stands to reason that their likelihood estimates for the unspecified set should be conservative.

The current study includes a partial replication of the Gettys et al. (1978) study and two additional manipulations to test the availability explanation.

The two manipulations were designed to increase the availability of catch-all alternatives. Our prediction was that increasing the availability of catch-all possibilities would increase catch-all assessments, reducing subjects' over-confidence in the specified sets. A Control group was presented problems in a computerized format, one datum per problem. The subjects' basic task was to estimate the odds for three specified hypotheses and the catch-all alternative. Subjects in one experimental condition, the "Exemplar" group, were presented the Control subjects' display plus five exemplar hypotheses. Subjects in the other experimental condition, the "Retrieval" group, were asked to generate candidate hypotheses for the catch-all before making the same type of odds estimates as subjects in the other groups.

We examined the two experimental manipulations partially for their applied implications. Although either could be implemented in an applied setting, the Retrieval group procedure of encouraging subjects to populate catch-all sets with possible hypotheses would be preferred over the Exemplar procedure if they were equally effective. The Retrieval manipulation is essentially only a change in instructions or training. The Exemplar procedure requires equipment to display the exemplar catch-all hypotheses during the hypothesis-generation task and the generation of extensive exemplar lists prior to the task. Particularly in nonrecurring situations, obtaining high-quality exemplars may be difficult or impossible.

Method

Subjects

A total of 48 subjects participated in this study. All were undergraduate students at the University of Oklahoma enrolled in the Introductory Psychology course. Subjects were randomly assigned to the three conditions, 16 subjects per condition. Half of the subjects in each condition were female and half were male.

Apparatus

The experimental sessions were under the control of an intelligent graphics terminal having color graphics capability. The computer was a Compucolor 8001, manufactured by the Intelligent Systems Corporation, Norcross, GA. Control and Exemplar group subjects entered odds estimates using the terminal's lightpen. Retrieval subjects entered possible hypotheses on the terminal's keyboard before entering odds estimates with the light pen. The odds estimates entered with the lightpens were assumed to be proportional to the probabilities of the hypotheses, or sets of hypotheses, given the data and could be converted to probability measures through a simple normalization.

Problem Generation

A data base consisting of 166,853 records was used to generate 30 problems for this study. The data base was created by accessing the computer master record file for nontransfer undergraduate students at the University of Oklahoma. The results of our analyses of this data base were frequencies which may be considered to be the actual population parameters. Classes were selected to have a reasonably large enrollment. Problems were selected so that the probability of the set of three specified hypotheses varied from fairly small to fairly large and so that the catch-all set of unspecified hypotheses was fairly rich.

Example Problem

Following is a description of three subjects' responses to an example problem to provide a concrete illustration of the procedure. The subjects' responses were to problem 24, which involved the datum: "Aviation 1113, Introduction to Aviation," a three-credit freshman-level course. This datum represents a class taken by an undergradute student having an unknown major. Subjects were asked to evaluate the relative li' hood of these four possibilities: Social Work, Psychology, Education and "all others," the catch-all alternative. The veridical probabilitis were, respectively, 0, 2.7, 6.6 and 90.7 percent. Subject 2, in the Control condition, gave magnitude estimation responses which, when converted to percent probabilities, were: 50.7 for Social Work, 14.2 for Psychology, 18.8 for Education and 16.2 for all others.

Subject 1 was assigned to the Exemplar condition and for this problem was shown a list of the following majors as possibilities in the catch-all set: Business, Journalism, University College Unclassified, Political Science and Nursing. Together, these five possibilities accounted for 56.4 percent of students who had enrolled in Aviation 1113. This subject's responses, converted to percent probabilities, were 31.3 percent for Social Work, 25.0 percent for Psychology, 20.0 percent for Education and 23.8 percent for all others.

Subject 3, a member of the Retrieval group, suggested the following set of majors as containing all possibilities having a probability greater than zero: Business, Journalism, Home Economics, Sociology and Chemistry. The veridical probabilty of this collection of five hypotheses is 42.0 percent. Subject 3's responses converted to probability percents were: 39.4 percent for Social Work, 9.8 percent for Psychology, 30.7 percent for Education and 20.1 percent for all others.

Procedure

Each session began with instructions presented on the terminal's CRT. In each task, the study was subject-paced. The Control and Exemplar group subjects generally required one hour to complete the instructions and the experimental session while the Retrieval group subjects required two hours. During the experimental session, each subject was presented 30 problems in a random order. Each problem contained three specified hypotheses concerning the possible major of an unknown University of Oklahoma undergraduate student and a fourth "catch-all" alternative that the unknown student had some other major. Also provided was a course that the unknown student had taken, described by the course number, department and title.

Instructions. The instructions were designed to provide graduated training in the experimental task. Subjects were first introduced to the operation of the light-pen, then were trained in the magnitude estimation procedure using a concrete problem involving estimation of the areas of rectangles and a more abstract problem involving prediction of the outcome of the next presidential election. The final phase of the instructions involved ten problems of the same type as those used in the actual experimental session.

Experimental Tasks. The display for the experimental task of the Control group subjects consisted of a boxed area at the top of the CRT containing the course number, department and title of the class the unknown student had taken. Below the box were four horizontal lines. The top-most three lines were labeled with three specified majors. The fourth line, labeled "All Others," corresponded to the catch-all alternative. Subjects made magnitude estimation responses by adjusting the length of a colored segment on the horizontal lines with a light pen. The horizontal lines were labeled with calibration markings at 0, 25, 50, 75 and 100, with 100 corresponding to the full length of the line. Thus, the subjects' modulus for the magnitude estimation procedure was 100, the length of the line identified with the most likely alternative. The specified majors for each problem were the same for all groups, but problem presentation order was randomized across subjects. Also, the order of the three specified hypotheses on the display was randomized for each problem in all conditions.

Exemplar group subjects saw virtually the same instructions and problems as the Control group subjects, except that the computer inserted the word "Including:" and a list of five candidate alternatives below the label "All Others" on the bottom line of the CRT display.

Unlike the Control and Exemplar group subjects, the Retrieval group subjects where shown two displays for each problem, rather than just one. Otherwise the instructions and experimental problems were identical to those for the other two groups. The first page display contained the data set off in a box at the top of the screen. Subjects were instructed to enter possible exemplars for the catch-all alternative until they believed their list covered virtually all possibilities in the catch-all having probabilities greater than zero. On the basis of a pilot study, the software was written to not accept more than five catch-all possibilities. (Subjects in the main study seldom entered even five possibilities. The mean number of possibilities entered by subjects in this condition was only 1.87.) For this subtask, the computer assisted subjects with spelling to insure that the majors would be correctly spelled for further processing. The second page displays for the Retrieval group were identical to the displays seen by Exemplar group subjects, except that page one responses were listed as candidate catch-all majors, replacing the computer-generated list supplied to Exemplar group subjects.

Results and Discussion

The probabilities subjects assigned to the catch-all alternatives were calculated from their magnitude estimates and were used as scores for an initial ANOVA. For this analysis, subjects' magnitude estimation responses for the three specified hypotheses and the catch-all alternative were normalized to probabilities and the probabilities assigned to the catch-all alternatives were used as scores. A conservative catch-all response corresponds to excessive assessments of the collection of specified hypotheses and vice versa. The factors for this analysis were the 30 problems, subjects, the three groups (Control, Exemplar and Retrieval) and a female/male blocking factor.

Overall, the pattern of excessive estimates for the specified hypotheses and conservative estimates for the catch-all hypotheses observed in the previous study (Gettys et al., 1978) was replicated here; also, both experimental manipulations reduced conservatism in a mean sense. The group means were: Control, 17.6 percent; Exemplar, 27.1 percent and Retrieval, 23.4 percent, compared to a veridical mean catch-all probability of 48.9 percent. The group main effect was significant, $F(2, 42) = 7.59$, $p < .01$. The male/female blocking factor was not significant. The main effect due to problems was significant, $F(29, 1218) = 23.0$, $p < .001$. The major difference between problems was the veridical probability of the catch-all alternative, and scores for this analysis were subjects' estimates of this probability. Subsequent analyses examine the significant problems effect and its interaction with the experimental manipulation, the groups effect, in more detail.

No interactions among the factors of this analysis achieved statistical significance, except the problem by group interaction, $F(58, 1218) = 2.77$, $p < .001$. This interaction suggests that the experimental manipulation did not have a simple additive effect on responses. An approach to investigating this significant interaction was to introduce an additional factor into the ANOVA. The "diagnosticity" factor was created by sorting problems into three groups on the basis of the veridical probabilities of the catch-all sets. These three categories were "low", "medium" and "high" diagnosticity, according to whether the veridical group probability of the catch-all sets was low, medium or high.

Table 1 shows the means obtained for subjects in each of the three conditions, Control, Exemplar and Retrieval, over the three diagnostic catagories of problems. The mean probability of the catch-all alternatives are contrasted with the veridical values.

(Insert Table 1 about here)

In general, subjects increased the magnitude of their responses as diagnosticity increased. The means for the diagnosticity categories were: low, 18.8 percent; medium, 24.1 percent and high, 25.3 percent. The diagnosticity main

Table 1

Mean Catch-All Probabilities

Expressed as Percents

	Group			
Diagnosticity	Control	Exemplar	Retrieval	Veridical
Low	16.6	21.1	18.7	24.9
Medium	18.2	29.1	24.9	49.5
High	18.0	31.1	26.6	72.3
Means	17.6	27.1	23.4	48.9

effect represented by these means was significant, $F(2, 84) = 49.73$, $p < .001$. Since the problems by group interaction was significant in the previous analysis, it should not be surprising that the diagnosticity by group interaction was significant in this analysis, $F(4, 84) = 7.58$, $p < .001$. The interaction of the blocking variable (male/female) with diagnosticity and the three-way interaction were not significant.

A more fine-grained analysis of the differential impact of the varied veridical probabilities of the catch-all sets on the three groups was undertaken using two approaches which yielded converging results. One approach was in the Bayesian tradition for examining the quality of probabilistic responses. Individual responses of each subject for each problem were transformed to log (base 10) odds, with the (posterior) odds being expressed as the ratio of the estimate for the set of three specified hypotheses divided by the likelihood estimate for the catch-all set of unspecified hypotheses. These transformed

----

(Insert Table 2 about here)

----

scores were compared to the veridical log odds in a correlational analysis for each group. Results are listed in Table 2. For these calculations, responses resulting in undefined (infinite) log odds were deleted. Tabled also are the number of responses deleted for this reason in each group.

This analysis sheds some light on the nature of the significant problems by

Table 2

Correlational Analysis of Log Odds Scores

	Correlation		Regression[a]			
Group	r	r^2	Slope	Intercept	Number Deleted	Number of Pairs
Control	.094	.008	.100	.796	1	479
Exemplar	.223	.050	.251	.498	16	464
Retrieval	.206	.042	.199	.584	11	469

[a] The Slope and Intercept have the following interpretation:
Subject's Log Odds = Veridical Log Odds x Slope + Intercept.

group and diagnosticity by group interactions noted earlier. The variability in the Control group responses is nearly unrelated to the variability in the veridical values. The slope of the regression line for the control group is nearly flat, .100. Both experimental manipulations reduced the conservatism bias in responses, but not as an additive constant; subjects in both experimental groups were more inclined to vary their estimates somewhat in accord with variations in the population parameters. In comparison to the Control group, the square of the Pearson r was over five times as large for both the Exemplar and Retrieval groups, with the Exemplar group showing somewhat of an advantage. There was an increase in the slope of the regression lines for both experimental groups also. By way of reference, the regression line slope would be 1.0 if subjects were perfectly calibrated. The regression lines for the three groups are plotted on the same graph for comparison in Figure 1.

As might be expected from the low correlations obtained, the scatter plots for these regression lines are fairly uninformative. Another approach to illustrating the differences between groups was to consolidate the scattered problem means into diagnosticity means, making use of the additional factor introduced for the second ANOVA. Figure 2 is a graph of these means.

(Insert Figure 1 and 2 about here.)

To examine the diagnosticity factor means in terms of log odds, the problem

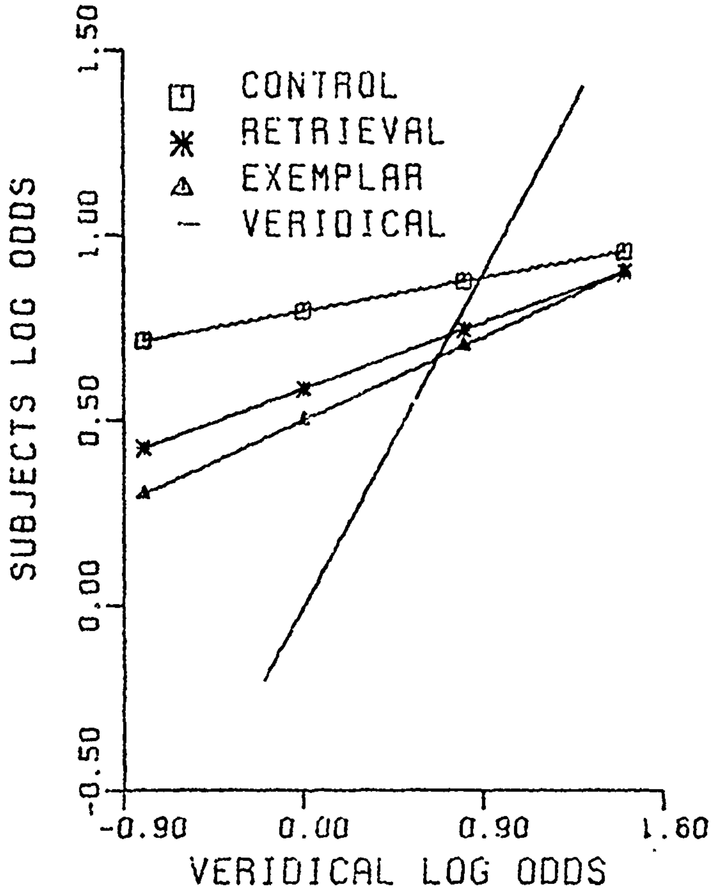

Figure 1. Regression lines for the Control, Exemplar and Retrieval groups contrasted with the veridical line. The scores used in the regression analysis were calculated as the log (base 10) of the ratio of assessments for specified hypotheses divided by assessments for catch-all sets. The symbols do not represent significant points on the lines; they were plotted only to distinguish among the regression lines. The solid line represents the performance of an optimal subject producing veridical responses. Each regression line summarizes approximately 480 scores.

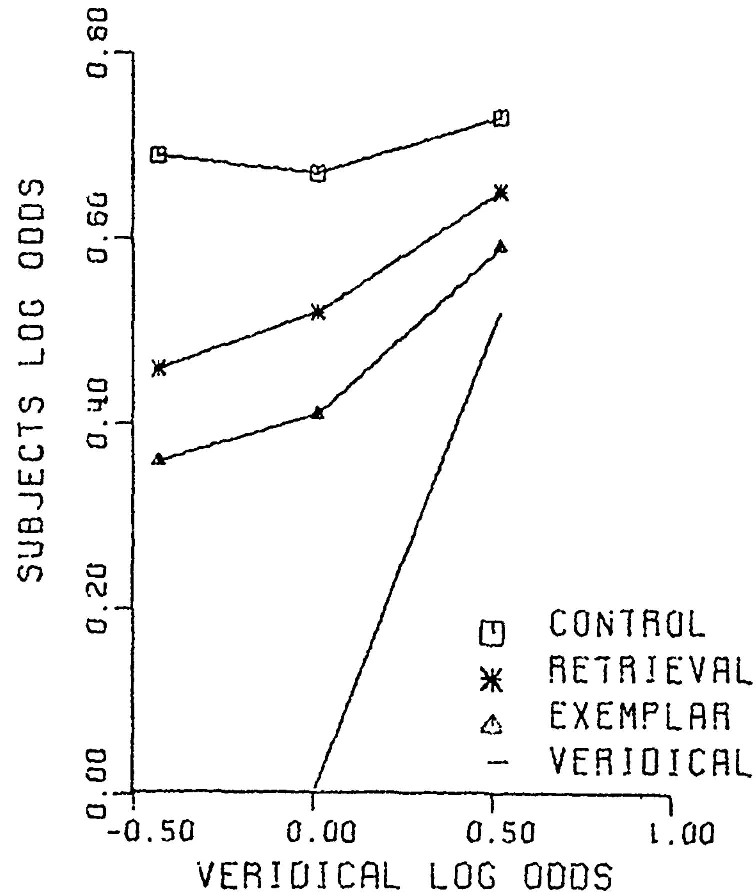

Figure 2. Group means for each level of the diagnosticity factor, expressed as log (base 10) odds. Each level of the diagnosticity factor included ten distinct problems.

mean catch-all probability for each group was transformed to log odds. These transformed means were averaged within the three diagnosticity categories to obtain the points plotted in Figure 2. The pattern of decreased overconfidence for both experimental groups is in evidence in the second figure also. There is a "fanning" tendency across the diagnosticity factors, with the Exemplar group's superiority to the other groups maintained over all three diagnosticity levels.

An alternative approach to examining the trials by group interaction in more detail was carried out by calculating Brier scores for each of the three groups and examining a partition of these scores (Murphy, 1973). The Brier score is a member of a class of measures of probabilistic estimates called "proper scoring rules." The principle application of proper scoring rules in psychology has been as feedback mechanisms in the training of probabilistic assessors, see Pickhardt and Wallace (1974) and Lichtenstein and Fischhoff (Note 1) for examples. Our motivation for investigating the Murphy partition was to examine the effect of the experimental manipulations on each component. The names of the components and their relations to the Brier score are: Brier score = Uncertainty + Resolution - Reliability. See Murphy (1973) and Lichtenstein and Fischhoff (1977) for discussions of the interpretations of these components. Results of the calculations are shown in Table 3.

(Insert Table 3 about here.)

Table 3

Proper Scoring Rules Analysis of Subjects'
Assessments of Specified versus Unspecified Sets

Group	Partition[a] Uncertainty	Reliability	Resolution	Brier	Confidence[b]
Control	.500	.222	.002	.720	- .66
Exemplar	.500	.115	.006	.608	- .47
Retrieval	.500	.145	.008	.637	- .53

[a] The Brier score and reliability component each have a range { 0, 2 }, with smaller scores being preferred. Smaller scores are preferable for the uncertainty component also; this component has a range of { 0, .5 }. The resolution component has a range of { 0, .5 }, and larger scores are preferred. The Brier score is Uncertainty + Reliability - Resolution.

[b] The confidence score is not a component of the Brier score. The preferred score is 0 and the range of possible values is {-1, 1 }.

This analysis was done in terms of problem means for each group. The scores describe how well the problem means for each group characte... +' › population parameters. Murphy's (1973) approach was modified to calculate the scores shown in Figure 3. Murphy used vectors having all zero entries except for a "1" representing the state of the world which obtained. We were able to employ vectors having entries corresponding to the population parameters. Our guess is that the effect of this modification is to reduce the variability in computed scores. However, as noted by Lichtenstein and Fischhoff (1977), the distribution of the Brier score and its partitions are unknown at the present time. Murphy (1974) discussed a very related issue in the context of another scoring rule. Our analysis was in terms of two-state vectors (specified set, unspecified (catch-all) set) and the interval size was set to ten percent.

The uncertainty component was the maximum of .5 for each group. The difference between the theoretical maximum of .5 and the computed scores was in the fifth decimal place. Since this component is a property of the environment (Murphy, 1973) and each group was presented the same collection of 30 problems, the uncertainty score should not vary across groups. The magnitude of the uncertainty score was interpreted to indicate that we had achieved a modicum of success in our attempt to choose problems having catch-all probabilities which varied over a large number of values, with neither large nor small values favored.

Compared to the Control group, the reliability component decreased (improved) for both experimental groups. The reliability scores are clearly the component most influencing the differences among groups in total Brier scores. The reliability component is related to calibration as discussed in the context of the regression analysis. The Retrieval group's reliability score being the best of the three is in agreement with the regression analysis.

The resolu··n scores were so nearly identical that differences between them may be attributed to chance. However, both experimental groups had larger (better) resolution scores than the Control group.

Also listed in Table 3 are the confidence scores, a metric suggested by Lichtenstein and Fishhoff (1977), which is related to the reliability component, but which is not part of the Brier score. All three groups exhibited negative confidence scores, indicating excessiveness in specified set estimates (conservatism in unspecified set estimates). The ordering among the three groups is the same as suggested by the overall group mean catch-all responses of Table 1.

(Insert Table 4 about here.)

To further examine the nature of the availability heuristic in hypothesis generation, an analysis of the hypotheses suggested by Retrieval groups was carried out. Table 4 is a summary of this analysis. Subjects in the Retrieval condition were instructed to respond with every possible Major in the catch-all alternative having a probability greater than zero. Table 4 documents the difficulty subjects encountered on this subtask. Although the overall mean catch-all probability was actually 48.87 percent, the mean veridical

Table 4

Analysis of Retrieval Condition Catch-All Responses

Subject Number	Mean Number of Hypotheses in Subjects' Catch-All Sets	Mean Actual Per-Cent Probability of Subjects' Catch-All Sets
3	2.27	6.15
4	2.47	8.59
5	2.13	7.75
9	3.40	10.10
12	1.97	8.35
17	.97	3.91
22	2.00	8.56
24	.93	3.78
25	.60	1.63
26	.97	4.57
27	2.13	5.95
30	2.27	7.74
37	2.67	6.00
40	1.13	4.08
41	2.67	8.17
47	1.33	4.70
Means	1.87	6.25
Values for an Optimal Subject	23.13	48.87

probability of the sets of catch-all hypotheses subjects generated was only 6.25 percent.

One explanation for the very low probability of catch-all sets generated by subjects may be that, while the average number of hypotheses actually contained in the catch-all sets was 23.13, subjects were limited by the software to entering no more than five possibilities. However, subjects were usually satisfied with sets of possible catch-all hypotheses numbering far less than five. The average number of hypotheses in subjects' catch all sets was only 1.87. Apparently subjects could access in memory only approximately eight percent of the catch-all possibilities in this admittedly difficult task. We believe that this result is compelling evidence that most catch-all hypotheses were not available to subjects in this task.

To examine the effect of this heuristic from another perspective, an additional correlational analysis was undertaken. This analysis was done by problem for each subject in the Retrieval group. The actual probability of the catch-all set the subject generated was substituted for the veridical probability of the entire catch-all set. With this exception, the calculations were carried out in the same manner as those summarized in Table 2. The Pearson correlation coefficient calculated was .289; the square of this correlation was .084. The regression line had a slope of .201 and an intercept of .331. Eleven data pairs were deleted because the subject's log odds were undefined and and additional 87 data pairs, were deleted because the veridical log odds were undefined (i.e. the subject entered no catch-all possibilities). As a result of these deletions, the correlation statistics were descriptive of 382 total scores. This simple manipulation nearly doubled the correlation squared, from

.042 to .084, providing additional evidence for the operation of an availability heuristic.

Conclusions

The major conclusion of this study is that our "availability explanation" conjecture was supported by the data. Two independent manipulations designed to increase the availability of hypotheses in the catch-all alternative each served to decrease subjects' overconfidence in specified hypotheses, resulting in more veridical estimates overall. This change in subjects' probabilistic estimates was obtained for either of two manipulations which have no effect on the veridical probabilities.

It is clear that either experimenter-supplied exemplars for the catch-all, or subject-generated exemplars reduce the bias of plausibility estimates. If the Exemplar manipulation had involved populating the catch-all alternative with more than five hypotheses, this bias might have been reduced still further. The study did not address the extent to which availability, as we have defined it, explains the totality of the observed nonoptimal performance. It may be that other factors contribute to overconfidence in hypothesis generation tasks, for example, those mentioned by Tversky and Kahneman (in press) to explain this bias in other contexts. However, increasing the availability of catch-all hypotheses does decrease this bias. Either experimental procedure could be implemented in practical hypothesis generation to increase the quality of subsequent decision analyses.

Reference Note

1. Lichtenstein, S. and Fischhoff, B. _Training for Calibration_. Manuscript submitted for publication, 1979.

References

Gettys, C.G., Fisher, S.D. and Mehle, T. Hypothesis Generation and Plausibility Assessment. Annual Report TR 15-10-78, Decision Processes Laboratory, University of Oklahoma. October 1978. AD A060786.

Kahneman, D. and Tversky, A. Intuitive prediction: Biases and corrective procedures. Management Science. In press.

Lichtenstein, S. and Fischhoff, B. Do those who know more also know more about how much they know? Organizational Behavior and Human Performance, 1977, 20, 159-183.

Lichtenstein, S., Fischhoff, B. and Phillips, L.D. Calibration of probabilities: The state of the art. In H. Jungermann and G. deZeeuw (Eds.) Decision Making and Change in Human Affairs. Dordrecht-Holland: D. Reidel, 1977.

Murphy, A.H. A new vector partition of the probability score. Journal of Applied Meteorology, 1973, 12, 595-600.

Murphy, A.H. A sample skill score for probability forecasts. Monthly Weather Review, 1974, 102, 48-55.

Pickhardt, R.C. and Wallace, J.B. A study of the perfomance of subjective probability assessors. _Decision Sciences_, 1974, _5_, 347-363.

Tversky, A. and Kahneman, D. Availability: A heuristic for judging frequency and probability. _Cognitive Psychology_, 1973, _5_, 207-232.

Tversky, A. and Kahneman, D. Judgment under uncertainty: Heuristics and biases. _Science_, 1974, _185_, 1124-1131.

Footnote

This study was supported by the Office of Naval Research.

The authors are grateful for the assistance of Jeff Casey in analyzing the data and Chuck Rice for programming which provided us with a computer file summarizing four years of the student master records at the University of Oklahoma.

Requests for reprints should be sent to Charles Gettys, Department of Psychology, Room 740, Dale Hall Tower, University of Oklahoma, Norman, OK 73019.

OFFICE OF NAVAL RESEARCH, CODE 455
TECHNICAL REPORTS DISTRIBUTION LIST

CDR Paul R. Chatelier
Military Assistant for Training and Personnel Technology
Office of the Deputy Under Secretary of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Director
Engineering Psychology Programs
Code 455
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217 (5 cys)

Director
Analysis and Support Division
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Naval Analysis Programs
Code 431
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Operations Research Programs
Code 434
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Statistics and Probability Program
Code 436
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Information Systems Program
Code 437
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Special Assistant for Marine Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Commanding Officer
ONR Branch Office
ATTN: Dr. J. Lester
Building 114, Section D
666 Summer Street
Boston, Massachusetts 02210

Commanding Officer
ONR Branch Office
ATTN: Dr. C. Davis
536 South Clark Street
Chicago, Illinois 60605

Commanding Officer
ONR Branch Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, California 91106

Office of Naval Research
Scientific Liaison Group
American Embassy, Room A-407
APO San Francisco, California 96503

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375 (6 cys)

ONR, Code 455, Technical Reports Distribution List

Dr. Robert G. Smith
Office of the Chief of Naval Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Naval Training Equipment Center
ATTN: Technical Library
Orlando, Florida 32813

Human Factors Department
Code N215
Naval Training Equipment Center
Orlando, Florida 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation Group
Naval Training Equipment Center
Code N-00T
Orlando, Florida 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, California 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, California 93940

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Mr. Arnold Rubinstein
Naval Material Command
NAVMAT 98T24
Washington, D.C. 20360

Bureau of Naval Personnel
Special Assistant for Research Liaison
PERS-OR
Washington, D.C. 20370

Navy Personnel Research and Development Center
Manned Systems Design, Code 311
San Diego, California 92152

Navy Personnel Research and Development Center
Code 305
San Diego, California 92152

Navy Personnel Research and Development Center
Management Support Department
Code 210
San Diego, California 92152

CDR P. M. Curran
Human Factors Engineering Division
Naval Air Development Center
Warminster, Pennsylvania 18974

Mr. Ronald A. Erickson
Human Factors Branch
Code 3175
Naval Weapons Center
China Lake, California 93555

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, Maryland 21402

Mr. J. Barber
HQS, Department of the Army
DAPE-PBR
Washington, D.C. 20546

ONR Code 455, Technical Reports Distribution List

Dr. Joseph Zeidner
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Director, Organizations and
Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Dr. Edgar M. Johnson
Organizations and Systems Research
Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Technical Director
U.S. Army Human Engineering Labs
Aberdeen Proving Ground, Maryland 21005

U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

Dr. Donald A. Topmiller
Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, Ohio 45433

Air University Library
Maxwell Air Force Base, Alabama 36112

Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, Virginia 22314 (12 cys)

Dr. Stephen J. Andriole
Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

Dr. Judith Daly
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

Professor Douglas E. Hunter
Defense Intelligence School
Washington, D.C. 20374

Dr. Richard S. Heuer, Jr.
ORPA/AMERS
Washington, D.C. 20505

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

Dr. Meredith P. Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L. Street, N.W.
Washington, D.C. 20037

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

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Room 201
Boulder, Colorado 80309

ONR Code 455, Technical Reports Distribution List

Dr. Ronald Howard
Department of Engineering-Economic Systems
Stanford University
Stanford, California 94305

Dr. William Howell
Department of Psychology
Rice University
Houston, Texas 77001

Journal Supplement Abstract Service
American Psychological Association
1200 17th Street, N.W.
Washington, D.C. 20036 (3 cys)

Dr. Clinton Kelly
Decisions and Designs, Inc.
8400 Westpark Drive, Suite 600
P.O. Box 907
McLean, Virginia 22101

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, California 93017

Dr. Gary McClelland
Institute of Behavioral Sciences
University of Colorado
Boulder, Colorado 80309

Human Resources Research Office
300 N. Washington Street
Alexandria, Virginia 22314

Dr. Miley Merkhofer
Stanford Research Institute
Decision Analysis Group
Menlo Park, California 94025

Dr. Terence R. Mitchell
University of Washington
Seattle, Washington 98195

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, Virginia 22202

Professor Judea Pearl
Engineering Systems Department
University of California-Los Angeles
405 Hilgard Avenue
Los Angeles, California 90024

Professor Howard Raiffa
Graduate School of Business Administration
Harvard University
Soldiers Field Road
Boston, Massachusetts 02163

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, Oregon 97401

Dr. J. A. Swets
Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, California 94305

Dr. W. S. Vaughan
Oceanautics, Inc.
422 6th Street
Annapolis, Maryland 21403

Dr. Gershon Weltman
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, California 91364

Dr. David Dianich
Chairman, Dept. of Business and Economics
Salisbury State College
Salisbury, Maryland 21801

ONR Code 455, Technical Reports Distribution List

North East London Polytechnic
The Charles Myers Library
Livingstone Road
Stratford
London E15 21J
UNITED KINGDOM

Professor Dr. Carl Graf Hoyos
Institute for Psychology
Technical University
8000 Munich
Arcisstr 21
FEDERAL REPUBLIC OF GERMANY

Director, Human Factors Wing
Defense & Civil Institute of
Environment Medicine
Post Office Box 2000
Downsville, Toronto, Ontario
CANADA

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF
UNITED KINGDOM

Dr. J. Baal Schem
Acting Director
Tel - Aviv University
Interdisciplinary Center for Technological
Analysis and Forecasting
Ramat-Aviv, Tel-Aviv
Israel

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, California 94305