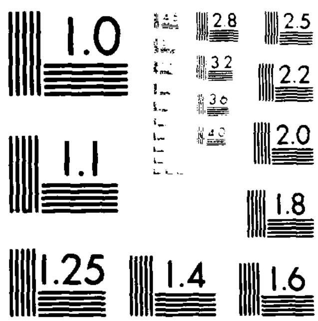
AD-A120 715 ACT GENERATION PERFORMANCE: THE EFFECTS OF INCENTIVE 1/1
(U) OKLAHOMA UNIV NORMAN DECISION PROCESSES LAB
R M PLISKE ET AL. 15 AUG 82 TR-15-8-82 N00014-80-C-0639

UNCLASSIFIED F/G 5/10 NL

END
DATE
FILMED
12-82
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AD A120715

LEVEL II

(11)

DTIC FILE COPY

DECISION PROCESSES LABORATORY
UNIVERSITY OF OKLAHOMA

DTIC
ELECTE
OCT 26 1982
S D
B

OFFICE OF NAVAL RESEARCH
CONTRACT NUMBER N00014-80-C-0639
WORK UNIT NUMBER NR197-066
REPRODUCTION IN WHOLE OR IN PART IS
PERMITTED FOR ANY PURPOSE OF THE
UNITED STATES GOVERNMENT
APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

82 10 26 010

ACT GENERATION PERFORMANCE:
THE EFFECTS OF INCENTIVE

REBECCA M. PLISKE, CHARLES F. GETTYS,
CAROL A. MANNING, AND JEFF T. CASEY

TR 15-8-82 AUGUST 1982

PREPARED FOR

OFFICE OF NAVAL RESEARCH
ENGINEERING PSYCHOLOGY PROGRAMS
CONTRACT NUMBER N00014-80-C-0639
WORK UNIT NUMBER NR 197-066

ACCESSION for
NTIS White Section
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

DTIC COPY INSPECTED 2

DECISION PROCESSES LABORATORY
DEPARTMENT OF PSCYHOLOGY
UNIVERSITY OF OKLAHOMA
NORMAN, OK 73019

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR 15-8-82	2. GOVT ACCESSION NO. AD-A120 715	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Act generation: The effect of performance incentives		5. TYPE OF REPORT & PERIOD COVERED technical Sept. 1981-August 1982 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. M. Pliske, C. F. Gettys, C. A. Manning, and J. T. Casey		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0639
9. PERFORMING ORGANIZATION NAME AND ADDRESS Decision Processes Laboratory Department of Psychology, Univ. of Oklahoma Norman, OK 73019		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 197-066
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Engineering Psychology Programs, Code 455 800 N. Quincy St., Arlington, VA 22217		12. REPORT DATE 15 August 1982 13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		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)

Act generation
Incentive effects
Predecision processes
Utility estimation

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

Two experiments explored the generalizability of earlier research which indicated that human act generation performance was impoverished. Subjects were given a realistic decision problem and were asked to generate actions which could be taken to solve the problem. Subjects in two incentive conditions were offered monetary rewards for generating additional actions. Subjects in one condition were rewarded for the sheer quantity of actions

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)

20. (continued)
produced and subjects in the other condition were rewarded for the quality of the actions produced. In a second experiment, both expert and naive subjects judged the quality of the actions produced by subjects in the first experiment. The results replicate earlier research in that most subjects generated relatively few actions and they also failed to generate important actions as rated by both expert and naive judges. There were no significant differences between the performance of subjects in the incentive conditions and subjects in the control condition. Thus, even when subjects are given substantial monetary incentives to generate additional actions, their act generation performance is impoverished. Differences in the act generation performance of the "quantity" and "quality" incentive conditions are discussed.

Act generation performance: The effects of incentive

Most research on human decision making has focused on the choice stage of the decision making process where the decision maker selects one of several alternative courses of action. Much of this research compared human performance with optimal decision models. There has recently been a growing interest in studying the cognitive processes involved in decision making (Wallsten, 1980). Researchers are also beginning to examine previously unexplored aspects of decision making called predecision processes.

Predecision processes are the cognitive processes which occur prior to the choice stage where the decision maker actually selects a course of action. Many activities must occur before a decision maker selects a course of action. Carroll (1980) states that we "must study decision makers from the point at which they recognize a decision could be made" (p. 70). He goes on to discuss the importance of problem representation in decision making tasks. Corbin (1980) examines a different aspect of predecision processes. She discusses various predecision processes which determine when and if choices get made. For example, a decision maker may delay the choice stage by continuing to examine alternative actions.

The most important process involved in predecision behavior is problem structuring. After decision makers recognize there is a decision problem they want to resolve, they generate the various options for action available to them. We refer to this process as act generation. An evaluation of the generated acts involves hypothesis and outcome generation. That is, the decision maker should identify the relevant states of the world which will influence the possible outcomes of the acts (i.e., they must form hypotheses about the future state of the world) and then consider the possible consequences that may result from choosing a particular course of action (i.e., they should generate outcomes). Once the structure of the problem has been elaborated, the decision maker may also consider the probability of states of the world and the utility of outcomes. These latter activities and the actual choice itself have been the traditional focus of study.

The elements of decision problem structuring - acts, states of the world, and outcomes- are inherent in decision tasks, and hence should be considered by a motivated decision maker who is faced with an important problem. We do not claim that all decision makers explicitly and consciously generate these entities for all decisions. Their behavior may vary according to the importance of the decision problem and their familiarity with it. Decision analysis is sometimes used on important problems by a sophisticated decision maker. Decision analysis is the art of making the implicit problem structure explicit. It pays great attention to the formal structure of the decision problem because this structure is used to develop the decision algorithm and to decompose the problem into manageable subproblems.

Research on hypothesis generation (Gettys and Fisher. 1979; Manning, Gettys, Nicewander. Fisher, & Mehle, 1980; Mehle, Gettys, Manning, Baca, & Fisher. 1981; Mehle, Note 1) suggests that hypothesis generation performance is impoverished. Subjects not only fail to generate important hypotheses. they also grossly over-estimate the completeness of their hypothesis sets. Inferior performance has been found with both college students and "expert" populations (Mehle, Note 1; Gettys, Mehle, Baca, Fisher. & Manning, Note 2).

The first study which examined act generation in a decision making context (Pitz, Sachs, & Heerboth, 1[illegible]) investigated different methods for eliciting acts from subjects and found that the serial presentation of the problem objectives improved act generation performance. They observed that subjects failed to generate complete sets of acts, but could not make any strong conclusions about act generation performance *per se* because their subjects were given a very limited time to respond.

Gettys, Manning, & Casey (Note 3) assessed the adequacy of act generation performance by giving subjects realistic decision problems and asking them to respond with any act occurring to them. Subjects' act generation performance was evaluated using several different techniques. A group decision tree was constructed by combining all the acts suggested by the subjects into a hierarchical tree structure. The major or generic ideas formed the "limbs" of the tree and the variations of the generic ideas formed the "branches." Individual subjects' decision trees were then compared to the group tree. It was found that individual subjects failed to generate important limbs and branches of the group decision tree.

In addition to obtaining data on the quantity of acts generated, Gettys et al. devised a methodology to estimate the quality of act generation performance. They obtained utility estimates from a different group of subjects for the various acts generated by subjects in the act generation experiment. These utility estimates were used to calculate the "opportunity loss" or potential cost which would result from subjects' failure to generate important limbs and branches. The opportunity loss analysis indicated that failure to generate a complete decision tree could be quite costly.

The research on predecision processes by Gettys and his colleagues suggests that future efforts to improve decision making might profit more from concentrating on the predecision processes of hypothesis and act generation than by concentrating exclusively on the choice process. If decision makers fail to generate the important hypotheses and actions, then their choice is likely to be suboptimal and the effort expended to choose the "optimal" action from an impoverished list of action alternatives may be wasted. Furthermore, some work on computer-assisted decision making has already begun to explore ways to improve act generation. This research makes the tacit assumption that decision makers are so deficient in problem structuring that a major effort is needed to aid them (eg., Leal and Pearl, 1977; Pearl, Leal. & Saleh. Note 4). Because the Gettys et al. (Note 3) study is the first study demonstrating that act generation performance is impoverished, it is important to establish the generality of this result before recommending a major redirection of effort in decision research. Thus, the present study attempts to replicate and extend our earlier work on act generation.

Several factors should be examined to establish the generality of the Gettys et al. results. Anyone who has worked with introductory psychology students as subjects knows that some of these students are not highly motivated. Although most of the subjects in the Gettys et al. study seemed to find the decision problems interesting and many subjects reported that working on a computer was "fun," we could not rule out a simple motivational explanation for the observed deficiency in act generation performance. Therefore, Experiment 1 in the present study includes two different incentive conditions to determine the extent to which act generation performance might be improved by substantial incentives to perform well.

Another area in which the generality of the Gettys et al. conclusion should be explored involves the techniques used to assess the quality of subjects' act generation performance. Previous research on act generation had simply counted the number of acts generated (e.g., Pitz et al., 1980; see also Lorge, Fox, Davitz & Brenner, 1958). Gettys et al. introduced new techniques to assess the quality as well as quantity of the acts generated. Experiment 2 in the present study explored the generalizability of the Gettys et al. findings by using a different utility estimation procedure and by obtaining utility estimates from expert as well as student subjects.

Three different instructional conditions were included in Experiment 1 to explore a motivational explanation for the observed deficiency in act generation performance. The Control condition was essentially a replication of the Gettys et al. study except for a small, but important change in the instructions. In the Gettys et al. study, we went out of our way to encourage subjects to record **any** idea which came to mind. That is, we did not want them to censor or evaluate the actions which they generated. These instructions prompted subjects to produce some worthless ideas (e.g., "Nuke the university."). To reduce the number of worthless solutions generated, the subjects in Experiment 1 were encouraged to think of as many alternative actions as possible, but were told that "worthless or counterproductive" actions would "not help their score."

In addition to the Control condition, two incentive conditions were employed. The Quantity condition received the same instructions as the Control condition except they were also told they would receive a monetary reward for each action they generated. We included another incentive group which was told their payment would be based on the **quality** of the solutions they generated. We included the Quality condition to determine whether subjects who were specifically instructed to generate high-quality solutions (and were paid for doing so) would produce more high quality solutions than subjects paid for the quantity of their solutions or control subjects who were not paid at all.

We predicted that subjects in both incentive conditions would produce more acts than subjects in the Control condition, but that all three groups would display the impoverished act generation behavior found previously to some degree. Based on our own intuitions, we expected subjects in the Quality condition to produce fewer low utility acts and a greater number of high utility acts than subjects in the Quantity and Control conditions. However, a study by Johnson, Parrott & Stratton (1968) suggested a different outcome for Experiment 1.

Johnson et al. (1968) manipulated quantity and quality instructions for several different tasks (e.g., generation of titles for a story plot, generation of titles for a table of data, etc.). They found the somewhat paradoxical effect that subjects in the "quantity" instructional condition actually produced higher quality solutions than subjects in the "quality" instruction condition. Because the tasks employed by Johnson et al. were so different from the decision problem we used in Experiment 1, we did not necessarily expect to replicate their findings. Furthermore, the Johnson et al. finding may have occurred because the experimenters' definition of quality may have differed from the subjects' definition.

Subjects in all three conditions in Experiment 1 were asked to assess their own act generation performance. Subjects were asked to estimate the number of acts (in addition to the ones they had already listed) which might be taken to solve the problem. They were also asked to estimate how many of these "ungenerated" acts were **good** acts which should definitely be considered as possible solutions to the problem. Based on our previous research on hypothesis generation (see Gettys, Manning, Mehle, & Fisher. Note 5. for a review). we predicted that subjects would greatly overestimate the completeness of their set of actions.

Method

Subjects

Subjects were 63 introductory psychology students from the University of Oklahoma. All subjects participated in the experiment for course credit. Subjects were told when they signed up for the experiment that they would have to be able to type 20 words per minute to be in the experiment. Data from three subjects (one from each condition) were not analyzed because their post-experimental questionnaires indicated that they ran out of time before completing the experimental task. The data reported below are based on 20 subjects per condition.

Incentive

The amount of incentive used was set so excellent performance would be quite profitable. Based on our previous research (Gettys et al., Note 3), we chose incentive amounts so that our best subjects would have earned \$15 to \$20 if we had paid them 50 cents per action. Subjects in the Quantity condition of the present study were paid \$.50 for each reasonable idea they generated. Subjects in the Quantity condition were paid between one cent and \$1.00 for the actions they generated. The actual amount received for each act was determined by the experimenter after the subject had finished the entire experimental session. The amount paid to subjects in the Quality condition was in direct proportion to the utility of the action as estimated by subjects in the Gettys et al., (Note 3) study. That is, if an act was estimated be worth 100 utiles, then the subject was paid \$1.00 for that act; if the act was estimated to be worth 76 utiles, then the subject was paid \$.76 for that act.

Problem

The decision problem used will be referred to as the "Parking" problem. Subjects were asked to think of themselves as members of a student committee whose task was to think of alternative solutions to the parking problem at the University of Oklahoma. Our previous research (Gettys et al., Note 3) indicated that subjects find this problem realistic and meaningful. It is also an attractive problem because it is a specific example of a more general type of problem faced by many decision makers. That is, the Parking problem is a specific example of a generic shortage problem in which a decision maker is faced with a shortage of a valuable commodity (i.e., parking places).

Instructions

Three instructional conditions were included. Subjects in the Control group were instructed to generate all *possible solutions to* the problem. The Quantity instruction group received the same instructions as the Control group, except they were also told they would receive 50 cents for each reasonable solution they generated. In contrast, subjects in the Quality instruction group were specifically instructed to try to think of the best solutions to the problem. They were told they would be paid between one cent and one dollar for their solutions depending on the quality of the solutions as rated by other introductory psychology students. Subjects in all three conditions were cautioned not to suggest "minor variations of the same solution" or "frivolous or counterproductive solutions." They were also told that they could spend as much time as they wanted to complete the problem; the average subject spent 45 minutes on the Parking problem.

Procedure

Subjects were seated in front of a CRT terminal which was controlled by a microcomputer. Subjects interacted with the computer by reading textual material displayed on the CRT screen and typing in their responses via the keyboard. Subjects' interaction with the computer was simplified by providing them with a "menu." The menu allowed them to: 1) Review the instructions, 2) Display the problem on the screen, 3) Enter a new action, 4) Review acts already generated (to avoid duplication), 5) Get help on the menu, and 6) Terminate the experiment.

Subjects were given three chances to pass a short typing test for which they had to type approximately 20 words per minute. Subjects unable to pass the test were not eligible to participate in the experiment.

Subjects were given a practice problem in which they were asked to think of actions one might take if one ran out of gas on the highway and had no money. The experimenter helped the subjects during the practice problem to insure they understood the instructions and could interact with the computer without any difficulty. After completing the practice problem, subjects were shown exemplar acts generated by a highly creative subject for this problem to encourage them to think of as many acts as possible.

Following the practice problem, the instructions for the Parking problem appeared on the screen. At this point, the instructional manipulation described above was introduced and the experimenter questioned the subjects in the Quantity and Quality conditions to make sure they understood how they would be paid for their performance. After reading the text describing the parking problem on the University of Oklahoma's campus, subjects began entering their actions. To enter an action they had to select the appropriate key from the menu described above (an "E" in this case). After they typed in their act, they were asked for a brief justification of the act. This procedure was used to aid the experimenters in the subsequent classification of responses by allowing the specification of how the act might accomplish the University's objectives. After subjects entered their justification, the computer returned to the menu. When subjects chose to terminate the experiment, they were asked to make sure they had entered all possible actions they could think of which could be taken to solve the problem and were given an opportunity to resume entering actions.

After subjects indicated they had thought of all the actions they could, they were asked to make two estimates. First, they were asked to "estimate the number of reasonable alternative solutions to this problem which you were unable to generate. That is, in addition to the options you have already entered, how many more options are there for solving this problem?" Next. they were asked to estimate how many of these "ungenerated" actions were "very good options which should definitely be considered as possible solutions to the problem."

Subjects were asked to fill out a post-experimental questionnaire before they left the experiment. The questionnaire included open-ended questions which asked subjects to describe how they generated their solutions. We also included several structured questions to which they responded "yes" or "no". These questions attempted to determine how familiar subjects were with the parking problem on campus and why subjects stopped generating solutions (e.g., Did you run out of time? Did you find yourself thinking of more and more bad ideas?).

Results and Discussion

Number of acts generated

On the average, subjects in the Control condition generated 9.3 acts, subjects in the Quantity condition generated 9.7 acts, and subjects in the Quality condition generated 8.1 acts. There was no significant difference between the mean number of acts generated by the different conditions ($F < 1$). The absence of large differences in the number of acts generated by subjects in the three conditions in the present study indicates that a substantial monetary incentive does not improve act generation performance in terms of the quantity of the acts generated.

Furthermore, these results suggest that the impoverished act generation performance found in the Gettys et al. (Note 3) study was probably not due to motivational factors. There was only a small difference between the number of acts generated by the subjects in the present study, who (when collapsed across condition) generated an average of 9.0 acts, and the number of acts generated by the subjects in the Gettys et al. study, who generated an average of 11.2 acts.

The larger number of acts per subject in the Gettys et al. study is readily accounted for by the change in instructions. Recall that subjects in the present study were cautioned not to include frivolous or counterproductive actions, whereas subjects in the Gettys et al. study were not. Twenty per cent of the acts generated by subjects in the Gettys et al. study were classified as "bad acts," whereas less than five per cent of the acts in the present study fell into that category. Subjects in the Gettys et al. study generated about 9.0 "good" acts and subjects in the present study generated about 8.6 "good" acts. Thus, it appears that subjects generated approximately equal numbers of reasonable acts in both experiments regardless of whether or not they received an incentive.

Large individual differences in the number of acts generated by the subjects were observed. These individual differences in the three conditions are displayed in the histograms in Figure 1. The largest number of acts was

Figure 1

Histograms showing the number of acts generated by subjects in the three conditions.

Middle of Interval	Control	Quantity	Quality
		Condition	
4	***	***	**
6	**	****	******
8	********	**	****
10	***	**	***
12		**	****
14	**	****	
16	*	*	*
18		**	
34	*		

Each * is one subject

generated by a subject in the Control condition who generated 34 acts. Three subjects, one in each condition, generated three acts which comprised the smallest number of acts generated by any one subject. As in the Gettys et al. study, most of the individual subjects generated only a small number of solutions to the Parking problem.

Classification of the acts

Many of the acts generated by the subjects were conceptually quite similar. For example, one subject suggested that the University build a parking lot next to the library, whereas another subject suggested building a parking lot next to the student union. The acts generated by the subjects were classified into a 40 category hierarchy so that these minor differences would not obscure the results and so we could examine the "breadth" as well as the "depth" of their performance. The categories were organized hierarchically into the decision tree shown in Table 1. Generic acts, or "limbs", are designated by the number to the left of the decimal point; major variations of this idea, or "branches", are designated by the number to the right of the decimal point. The two acts described above would both be classified as variants of category 1.3 "Build new parking lots on University land."

The decision tree shown in Table 1 was designed so that it would include the major ideas the subjects expressed in the actions they generated. It was developed from a tree constructed for our previous research (Manning, Note 6) which was based on a cluster analysis of subjects' similarity judgments of actions proposed to solve the Parking problem. Three of the authors modified the "cluster analysis" tree by adding several categories. This modification

Table 1

The Decision Tree Used to Classify Acts.

1.0 Create more parking spaces by building new facilities.
- 1.1 Build a highrise parking structure.
- 1.2 Build underground parking.
- 1.3 Build new parking lots on university land.
- 1.4 Expand existing surface lots.
- 1.5 Tear down old buildings to create space to build parking spaces.
- 1.6 Buy land to build additional parking lots.
- 1.7 Take the parking problem into account in future planning of expansion of the university.
- 1.8 Build additional remote lots and run buses to campus.

2.0 Obtain more parking spaces without actually building new facilities.
- 2.1 Use space more effectively (e.g., decrease the width of spaces).
- 2.2 Request the use of areas around campus (e.g., church lots) for additional parking.
- 2.3 Use city streets near campus for university parking.

3.0 Alternative forms of transportation--Group.
- 3.1 Encourage people to carpool.
- 3.2 Force certain people (e.g., commuters, faculty) to carpool.
- 3.3 Encourage people to use the C.A.R.T. system on campus.
- 3.4 Improve the C.A.R.T. system on campus.
- 3.5 Expand the C.A.R.T. system to include other areas of Norman.
- 3.6 Work with the near-by communities to form a mass transit system.

4.0 Alternative forms of transportation--Individual.
- 4.1 Encourage use of bicycles and/or motorcycles.
- 4.2 Make individual transportation safer.
- 4.3 Encourage other forms of individual transportation (e.g.,walking).

5.0 Change current university policies regarding parking.
- 5.1 Eliminate parking priorities.
- 5.2 Allow students to park in restricted areas (e.g., faculty/staff lots. during certain hours (e.g., after 6:00 p.m.).
- 5.3 Set time restrictions (e.g., 2 hour parking) on more lots.
- 5.4 Enforce existing parking regulations more strictly.
- 5.5 Make certain people (e.g., commuters) park in certain places.
- 5.6 Limit the number of cars on campus by not letting certain people (e.g., freshmen) have cars on campus.
- 5.7 Distribute a limited number of parking stickers.
- 5.8 Assign a specific space for each driver.
- 5.9 Outlaw cars on campus for everyone.
- 5.10 Allow certain people (e.g., those who have even number license plates) to only park on certain days of the week.
- 5.11 Increase the price of parking stickers.

6.0 Reduce the number of people who need to park.
- 6.1 Offer more correspondence courses.
- 6.2 Establish branch campuses of the university.
- 6.3 Reschedule activities and/or classes to change demand.
- 6.4 Provide housing or improve existing housing so people can walk.
- 6.5 Reduce the student population. For example, limit enrollment.
- 6.6 Have someone drop students & faculty/staff off and pick them up.

7.0 Indirect strategies for solving the problem.
- 7.1 Appeals to good judgment.
- 7.2 Ways to make money to solve the problem.
- 7.3 Suggestions for ways to come up with solutions.

8.0 "Flaky" acts (e.g., issue everyone a set of wings).

was necessary because the subjects in this experiment generated several additional actions that were not included in Manning's cluster analysis.

Two raters who were naive with respect to the details of the experiment independently classified the acts into the 40 category tree. The raters were not told from which condition the acts were obtained. The percentage of agreement for the two raters was 84 percent. Some of the disagreement was due to lack of knowledge on the part of the raters. For example: Was the location which a subject specified on-campus or off-campus? Other acts were difficult to classify because the subject had not clearly specified the action they had in mind. The acts for which the raters disagreed were discussed with several of the authors until everyone agreed on an appropriate classification.

Performance differences due to incentive

The frequencies with which subjects in the three conditions generated acts in the different limbs of the decision tree are shown in Table 2. A Chi-square analysis indicates there are reliable differences between conditions in the frequency with which acts were generated in the different limbs [$X^2(df=14)=45.3$, $p<.005$]. These differences are primarily in limbs 1, 5, 7 and 8.

Table 2

Frequency of the acts generated for the different limbs

Limb	Condition: Control	Quantity	Quality
1. Build new facilities	44	67	40
2. Create more parking without building new facilities	10	9	8
3. Alternate group transportation	28	33	29
4. Alternate individual transportation	12	7	6
5. Change parking policies	78	48	45
6. Reduce number of people who need to park	8	9	9
7. Indirect strategies for problem	1	6	18
8. "Flaky" acts	5	15	7

Subjects in the Quantity condition suggested more acts which were classified into limb 1 ("Build more parking spaces by building new facilities") than subjects in the Quality and Control conditions. Subjects in the Control condition generated more acts in limb 5 ("Change current university policies regarding parking") than subjects in the other conditions. It is not clear why subjects in the Control condition produced more acts in limb 5 and the subjects in the Quantity condition produced more acts in limb 1. Although these differences do not appear to be of great psychological interest, they do serve to demonstrate that the present experiment is of adequate power to detect differences between the conditions.

The differences in frequency with which subjects generated acts classified into limbs 7 and 8 are of some psychological interest. Subjects in the Quality condition generated more acts which were classified in the "Indirect strategies for solving the problem" limb. Most of these acts were classified in branch 7.2 "Ways to make money to solve the problem." This difference was probably due to the modified instructions which the Quality condition received. Whereas the Quantity and Control groups were told not to worry about how to finance their actions, this disclaimer was not included in the instructions for the Quality condition because cost is one of the factors which influences the "quality" of an action.

The Quantity condition generated more actions classified in limb 8 which included all actions with negative utility. Although subjects in all three conditions were cautioned not to enter frivolous or counter-productive actions, a few such "flaky" actions were suggested. The Quantity condition may have produced more of these actions than the other conditions because they felt they had nothing to loose and perhaps 50 cents to gain by entering additional actions.

The differences in the frequencies with which subjects in the different conditions generated actions in limbs 7 and 8 indicates that there were significant effects due to the instructional manipulation. However, because we had no **a priori** expectations concerning the expected frequencies with which the different conditions would generate acts classified into the various limbs and branches of the decision tree, the explanations of these differences are admittedly **post hoc**. The reader is encouraged to examine the frequency data presented in Table 2 and formulate his or her opinion.

Breadth versus depth

In order to determine whether the instructional conditions differed in terms of taking a "breadth" versus "depth" approach to the Parking problem. the number of limbs for which at least one act was generated and the number of branches generated for each limb was computed for each subject for each condition. If subjects generated a large number of limbs, this would suggest a "breadth" approach to the problem whereby subjects tried to generate a wide variety of solutions to the problem. If subjects generated a large number of branches for each limb, this would suggest a "depth" approach to the problem whereby subjects tried to generate several variations of the same generic idea.

The mean number of limbs for the Control condition was 3.6; the mean number of limbs for the Quantity condition was 3.7; the mean number of limbs for the Quality condition was 3.9. These differences were not statistically significant ($F < 1$). The mean number of branches per limb for subjects in the Control condition was 1.9; the mean number of branches per limb for subjects in the Quantity condition was 2.0; the mean number of branches per limb for subjects in the Quality condition was 1.6. The difference between these means approached significance, $F(2,45)=2.76$, $p<.10$.

These analyses suggest that on the average subjects in all three conditions generated actions in three or four of the seven major limbs (acts with negative utility were excluded from these analyses) and generated approximately two acts for each limb that occurred to them. There do not appear to be any large differences between the instructional conditions in terms of taking a "breadth" or "depth" approach to solving the problem. The fact that subjects in the Quality condition had a tendency to generate fewer branches per limb than subjects in the other conditions suggests that subjects in the Quality condition were less likely to suggest multiple variations of the same general ideas (i.e., they were less likely to use a "depth" approach). However, they did not generate a greater number of limbs as compared to subjects in the other conditions, so we cannot conclude they were more likely to take a "breadth" approach.

Conclusions from Experiment 1

The performance of the incentive groups in Experiment 1 was roughly comparable to the control group in the analyses presented above which all involve the number of acts generated. These results suggest that the impoverished act generation performance described by Gettys et al. (Note 3) cannot be attributed to a lack of motivation on the part of the subjects.

In addition to examining the number of acts generated by subjects given a realistic decision problem to solve, Gettys et al. (Note 3) developed a technique to examine the quality of act generation performance. The primary purpose of Experiment 2 is to further examine the procedures that were used to assess the quality of performance. The assessment of quality involves utility estimation. In Experiment 2 we explore a variant of the utility estimation procedures used by Gettys et al. to further examine the generality and robustness of our conclusion that human act generation is impoverished.

EXPERIMENT 2

In order to assess the quality of acts generated by the subjects in Experiment 1, it was necessary to obtain utility estimates for the 40 categories included in the decision tree shown in Table 1. One of the purposes of the present study was to test the generalizability of the Gettys et al. (Note 3) findings with respect to the utility estimation procedures used. Therefore, before the utility estimation procedures used in the present study are introduced, we will briefly summarize the methodology used by Gettys et al.

Gettys et al. performed two experiments (in addition to their act generation experiment) to obtain utility estimates for the acts proposed by their subjects. First, a preliminary screening experiment (their Experiment 2) was conducted to reduce the vast number of acts to a more manageable number by discarding negative or low utility acts. A group of introductory psychology students was shown the acts which had been generated by another group of subjects (their Experiment 1). The subjects in the preliminary screening experiment were asked to respond "yes" or "no" to the question of whether taking the action specified would leave the University better off than they currently were with regard to the Parking problem. If the majority of the subjects thought an act was of positive utility (i.e., they responded "yes"), then the act was included in the utility estimation experiment (their Experiment 3). If the majority of the subjects thought an act was of 0 or negative utility (i.e., they responded "no"), then that act was given a utility value of 0 for all subsequent analyses.

Gettys et al. conducted a third experiment to obtain utility estimates for the acts which had been screened in their second experiment. Introductory psychology students were first shown two anchor acts. The act "Do nothing about the parking problem" was assigned a value of zero and the act "Improve the Campus Area Rapid Transit System" was assigned a value of 100. Subjects made direct utility estimates by assigning a number to the act to be rated which represented their evaluation of that act as compared with the two anchor acts. If subjects believed an act had a higher utility value than the anchor act assigned 100 utility points, they were free to assign it a larger number. If they believed an act had a lower utility than the anchor act assigned 0 utility points, they could assign it a negative number. The median utility estimate made by the subjects was then used to assess the quality of the act generation performance.

Unfortunately, the procedure used in the Gettys et al. study produced data with a fairly large degree of between-subject variability. The distribution of utility estimates for a given act typically was a single-peaked function, but these distributions had a high variance and outlying estimates. Some of this variability was almost undoubtedly due to disagreement among the subjects about the utility of various acts, but some of the variability may have been due to the measurement procedure used. Although median utility estimates were chosen to reduce the influence of outlying scores, we were not entirely satisfied with the estimation procedure. To insure that the impoverished act generation performance reported in the Gettys et al. study was not an artifact produced by the utility estimation procedure, we used a different utility estimation procedure in the present study.

Experiment 2 examines two of the possible sources of the observed variability in the utility estimates obtained in the Gettys et al. study. The variability could be a result of the particular utility estimation procedure used or it could be due to the fact that the individual subjects had very different values in mind when making their utility estimates. To explore the first possibility we used a multi-stage utility estimation procedure. To explore the second possibility we obtained utility estimates from expert subjects as well as student subjects. Presumably, if expert and student subjects are in general agreement about utilities, then it is reasonable to assume that our utility measurement procedure tapped the underlying utility structure shared by most individuals in the University community to some degree.

The multi-stage utility estimation procedure involved four stages. First, subjects were asked to sort the acts into five categories identified with semantic labels such as "Suggestions which are worthwhile and should be considered for implementation." Second, subjects ordered the acts in each pile from best to worst. Third, subjects placed anchor acts on a utility number line which ranged from 0 to 100 utiles. Fourth, subjects placed the remaining acts on the number line. By having subjects work their way through these various stages and restricting the range of the utility estimates, we hoped to reduce the amount of variability in their responses.

The variability in the utility estimates used in the Gettys et al. study may simply have reflected the diverse viewpoints of the subjects. Some students in the Gettys et al. utility estimation experiment may not have been very knowledgeable about the parking situation on campus and did not have a clear set of criteria to judge the quality of the actions. We decided to include a sample of University administrators, who were designated as experts on the parking situation on campus, in the present study to see if they would produce utility estimates which were in general agreement with those of introductory psychology students.

The utility estimates employed in the present study

The utility estimates obtained in Experiment 2 are global estimates of the relative value of the different acts. The large number of unique acts which were generated by subjects in Experiment 1 made it impossible to use a more sophisticated type of utility estimation procedure such as multi-attribute utility scaling or conjoint measurement (von Winterfeldt, Note 7). Furthermore, some researchers argue that global utility estimates are superior to a decomposition approach even when the number of alternative actions is small enough to allow for a more complicated procedure (cf. Slovic, Fischhoff, & Lichtenstein, 1979).

When decision analysis is employed, the decision maker who structures the decision problem by generating actions, hypotheses about the state of the world, and outcomes, also estimates the utility of the generated outcomes. It was not possible to have each subject in Experiment 1 estimate the utility for the various alternative actions because the entire act pool was not constructed until all the subjects had completed the experiment. Therefore, we had to rely on a group utility estimate whereby an independent group of subjects estimated the utility of the actions.

As discussed above, some variability is to be expected in a group utility estimate because of the different values held by the individuals in the group. In fact, the validity of any type of group utility estimate has been debated (e.g., Keeney and Kirkwood, Note 8). For the purpose of the present study, an estimate of the general consensus regarding the relative value of the various actions for all members of the University community was desirable. Therefore, we propose that if consistency can be found both within and between two different samples of people within the University community, then we can conclude that the utility estimates obtained are reliable reflections of the general consensus regarding the relative value of the various actions.

Method

Expert Subjects

Four expert subjects were referred to us by the office of the Vice President of Administrative Affairs at the University of Oklahoma in response to a request made for a list of individuals who were knowledgeable about the parking problem on campus. They included the Director of Architectural and Engineering Services, the Manager of Parking and Traffic, the head of the Environmental Safety Committee for the University of Oklahoma Police Department, and the Project Coordinator for the Physical Plant. In addition, one of these expert subjects referred us to a faculty member in the Regional and City Planning Department who also agreed to participate in the study.

Student subjects

The student subjects were selected from the same population as used in Experiment 1. Ten introductory psychology students participated in the experiment for course credit.

Materials

The 40 categories which were used to classify the actions generated by subjects in Experiment 1 were typed on 3" X 5" index cards. An identification number was printed on the back of each card to aid the experimenter in recording the subjects' responses. Five additional cards were used for the category sort. These cards read as follows:

1) Suggestions which, if implemented, would leave us worse off than we are now and, thus, should definitely not be considered.

2) Suggestions which are not very good and should not be considered for implementation.

3) Suggestions which are reasonable but probably are not worth considering for implementation.

4) Suggestions which are worthwhile and should be considered for implementation.

5) Suggestions which are very worthwhile and should definitely be considered for implementation in the near future.

A large number line (approximately 4.5 feet long) was constructed from posterboard. It was scaled in units of 10 from 0 to 100 with the 0 end labeled "worst" and the 100 end labeled "best."

Expert subject procedure

The expert subjects were interviewed in their offices. The first part of the interview involved a series of open-ended questions. Subjects were asked to describe their experience in working on the parking problem on campus. They were then asked what solutions they thought the University should take to solve the problem. Finally, they were asked to describe what characteristics they considered to be important in designing a "good" solution to the problem (e.g., cost effectiveness, environmental impact, etc.). This portion of the interview was recorded on a tape recorder and was subsequently transcribed for future use.

The utility estimates were obtained in the second part of the interview. Subjects were told that the actions described on the 40 index cards represented different categories of actions suggested by introductory psychology students. They were given a copy of the textual description of the Parking problem which the original subjects had read. The subjects for the utility estimation study were asked to base their evaluation of the acts described on the cards on the following factors: cost effectiveness, potential benefits for the entire University community, esthetic and environmental impact, and political ramifications for both the University and the surrounding city.

The utility estimation procedure involved the following stages:

Stage 1) Subjects were asked to sort the 40 cards into five separate piles designated by the category headings described in the materials section above. They were encouraged to read through all the cards before sorting them into piles. However, some subjects insisted on sorting the first time through the deck of cards. Subjects were allowed to reclassify a card into a different pile if they wanted to.

Stage 2) Subjects were asked to order the cards in each of the five piles from best to worst. Subjects were to place the best solution on the top of the pile and the worst solution on the bottom of the pile. Subjects were allowed to reclassify a card into a different pile at this time if they wanted to.

Stage 3) The number line was shown to the subject and the properties of an interval scale were briefly explained. Subjects were instructed to put their best act (the act on top of the best pile) on the number line at the 100 utile position and to put their worst act (the act on the bottom of the worst pile) on the number line at the 0 utile position. The subject was then asked to take the worst act from the bottom of the remaining piles and place these acts on the number line while attempting to maintain the interval properties of the utility scale. The experimenter questioned the subject to make sure they understood how to use the scale (e.g., Do you think that the difference between acts A and B is in fact equal to the difference between acts C and D?). Subjects were encouraged to adjust these "anchor" acts until they were satisfied with their values.

Stage 4) In the final stage of the procedure, subjects were asked to place the remaining acts on the scale. They started with the best pile and placed these acts on the scale between 100 and wherever they had placed the highest anchor (which was the position of the worst act for the best category). The subjects continued to place the acts from the remaining piles on the scale between the appropriate anchors until all acts had been assigned utility values. The experimenter recorded the utility assigned to each act as it was placed on the scale.

Student utility estimation procedure

The student utility estimation procedure was identical to the procedure described above except for two modifications. First, student subjects were run in our laboratory, whereas expert subjects were run in their own offices. Second, we did not conduct the first part of the interview (the open ended questioning session) with the student subjects.

Results and Discussion

Expert interviews

We obtained a great deal of information about the parking situation on campus from the first part of our interviews with the expert subjects. Although they had rather diverse backgrounds and held quite different positions within the University, they were all very knowledgeable about the parking problem. Each expert had a slightly different opinion on which characteristics make up a "good" solution to the problem. For example, one expert was very concerned with the environmental impact the action would have, whereas another expert was more concerned with how the various actions would affect the various people in the university community.

In response to our request to describe the various solutions they thought the University should take to solve the problem. we found that even expert subjects failed to mention solutions which they subsequently rated as being of high utility. For example, one expert subject was very impressed with the action "Reschedule activities and/or classes to spread demand more evenly over time." He seemed somewhat surprised that such a good idea had escaped him. The informal nature of the interview did not allow for a controlled test of whether these expert subjects were indeed better act generators than our student subjects. The purpose of the interview was to obtain expert utility judgments, not to measure expert act generation behavior. However, our earlier work on hypothesis generation (Mehle. Note 1; Gettys et al., Note 2) suggests that experts, as well as student subjects, display impoverished decision structuring behavior.

Utility estimation

The median utility estimates for each of the 40 categories are shown in Table 3 for both expert and student subjects. The median utility estimates were rank ordered to facilitate the comparison of the two samples of subjects. These rankings are also shown in Table 3. A small number (e.g., a rank of 1.5) represents a high utility act and a large number (e.g., a rank of 39.5) represents a low utility act.

There was reasonable agreement within and between the two samples. The inter-subject correlations for the student sample ranged from .025-.715 (median = .470). The inter-subject correlations for the expert sample ranged from .244- .693 (median = .548). The correlation between the median utility estimates for the two samples was .678; the correlation between the rank orderings was also .678. The substantial amount of agreement within and between the two samples indicates that both sets of utility estimates provide approximate indices of the quality of the 40 categories of actions.

Inspection of Table 3 indicates some interesting similarities and differences between the two sets of utility estimates. For example, the act rated highest by the students was "Allow students to park in restricted areas during certain hours" (act classification 5.2). However, the consensus expressed by the experts regarding this act was that students were already allowed to park in many restricted areas and that changing the times when they are allowed to park in these additional areas would not be a very good idea. Two very similar acts received the highest rating by the experts. These acts (classifications 3.3 and 3.4) suggested that the Campus Area Rapid

Table 3

Median Utility Estimates and Rank Orderings for the Limbs and Branches of the Decision Tree Shown in Table 1

Decision Tree Classification	Experts Median	Experts Rank	Students Median	Students Rank
1.0				
1.1	50.0	20.0	62.5	27.0
1.2	30.0	28.5	71.5	11.0
1.3	75.0	11.5	68.8	14.0
1.4	62.0	17.0	78.0	5.0
1.5	34.0	24.5	46.0	24.0
1.6	83.0	3.5	51.5	22.0
1.7	79.0	7.0	75.0	10.0
1.8	83.0	3.5	79.0	4.0
2.0				
2.1	80.0	5.5	52.0	21.0
2.2	42.5	21.0	77.0	6.5
2.3	3.0	38.0	42.0	26.0
3.0				
3.1	77.0	9.0	84.5	2.5
3.2	18.0	33.0	19.0	33.0
3.3	95.0	1.5	76.0	8.5
3.4	95.0	1.5	76.0	8.5
3.5	78.0	8.0	84.5	2.5
3.6	67.0	16.0	70.0	13.0
4.0				
4.1	80.0	5.5	71.0	12.0
4.2	72.0	13.0	77.0	6.5
4.3	75.0	11.5	58.5	18.0
5.0				
5.1	13.0	34.0	13.3	37.0
5.2	33.5	26.0	89.5	1.0
5.3	32.0	27.0	58.0	19.0
5.4	59.0	18.0	35.0	27.0
5.5	34.0	24.5	65.0	16.0
5.6	20.0	32.0	19.5	32.0
5.7	4.0	36.5	42.5	25.0
5.8	27.0	30.0	17.5	35.0
5.9	1.0	39.5	0.0	40.0
5.10	4.0	36.5	11.0	38.0
5.11	70.0	15.0	20.8	31.0
6.0				
6.1	21.0	31.0	21.0	30.0
6.2	1.0	39.5	12.5	37.0
6.3	71.0	14.0	30.0	28.0
6.4	36.0	22.5	18.5	34.0
6.5	7.0	35.0	2.0	39.0
6.6	36.0	22.5	25.5	29.0
7.0				
7.1	51.0	19.0	55.0	20.0
7.2	30.0	28.5	50.5	23.0
7.3	76.0	10.0	67.5	15.0

Transit (CART) system be improved and that people should be encouraged to ride it. The students also rated these acts relatively highly (the median rank was 8.5). Both experts and students agreed that "Outlaw cars on campus for everyone" (act classification 5.9) and "Reduce the number of students" (act classification 6.5) were low utility acts. The reader is encouraged to examine Table 3 for additional comparisons of the two samples of subjects.

Assessing the quality of act generation performance

The various actions proposed for the decision problem used in the present study are not mutually exclusive. That is, the University could decide to implement multiple actions in an attempt to solve the Parking problem. A decision maker faced with this problem should identify a number of high quality solutions which could be taken to solve the problem. Gettys et al. (Note 3) developed a technique that allows one to assess to what extent subjects are able to generate a number of high quality actions. This technique is briefly summarized below. For a more complete discussion, the interested reader is referred to their paper.

Because we are unable to state unequivocally how many actions a subject should generate or could be implemented, our technique examines the best actions subjects generated in sets of various sizes. First, we calculate the utility of the set consisting of the best action. Next, the set containing the two best actions is obtained, and the utilities for these two best actions are summed to obtain the utility for set size two, etc. By establishing the utility for sets of various sizes, we avoid having to state exactly how many actions the subject should generate, while at the same time examining the breadth of their performance. Calculating the utility of the best actions for various set sizes is mathematically equivalent to first ordering the actions in order of decreasing utility, and then calculating the **cumulative utility** function from these ordered utilities. Due to the distributive law of mathematics, each point on the cumulative utility curve corresponds to the utility of the set of the best actions of that size. Therefore, the cumulative utility function has a straightforward and simple interpretation as a global measure of subject performance.

A cumulative utility function was created for each subject which sums the utilities for the ten highest utility acts generated by the subject. The analysis for subjects in the present study was based on the 40 category tree shown in Table 1. If a subject generated acts which were essentially multiple variations of a particular class of acts (e.g., "Build a parking lot next to the cafeteria" and "Build a parking lot next to the recreation center"), they were only given credit for one of these acts for this particular analysis. Thus, to calculate the cumulative utility function for a subject, the utility for their highest utility act was added to the utility for their second highest utility act, which was then added to their third highest utility act, etc. The cumulative utility function resulting from this process increases monotonically and eventually is asymptotic.

A similar cumulative utility function was obtained for the entire group of subjects by combining the data from all the subjects in all three conditions. That is, the highest utility act generated by the entire group of subjects was added to the second highest utility act generated by the entire group of subjects, etc. This group function was then used as a lower-bound estimate of optimal act generation performance. This criterion is a lower-bound estimate in the sense that adding more subjects to the group should increase the cumulative utility function. (See Gettys et al., Note 2 for more information.)

The difference between the first point on the optimal performance (i.e., the group) function and the first point on the average individual function reflects opportunity loss due to the failure to generate the highest utility act. The differences among the remaining points on the two functions reflects the completeness of the subjects' performance. That is, did they generate a large number of high utility acts? Thus, the difference between the two functions provides a useful summary of performance.

Cumulative utility functions were plotted for subjects in each of the three experimental conditions using both expert and student median utility estimates. Figure 2 compares the average individual performance for each of the three conditions to a lower-bound estimate of optimal performance using the expert median utility estimates. Figure 3 makes the same comparison using the student median utility estimates.

An examination of the cumulative utility functions for the different conditions indicates there is essentially no difference between conditions in terms of cumulative utility of the acts generated using either the expert or student median utility estimates. Although the functions for the different conditions separate somewhat at the larger act set sizes these differences are not statistically significant at the endpoints of the functions ($F < 1$). Therefore, subsequent analyses and discussion will combine the data from the three conditions.

An inspection of the first points (a set size of 1) of the functions shown in Figures 2 and 3 indicates that there is little difference between the average individual performance for the different conditions and the lower bound estimate of optimal performance. This suggests that most subjects were able to generate at least one high utility action. This conclusion may be seen more clearly if one considers the percentage of students generating the acts which were given the most utility points by the expert and student subjects in Experiment 2.

The best act generated by 15 per cent of the subjects was one of the two best acts (classifications 3.3 and 3.4) as rated by the experts. These acts suggested that the University improve the Campus Area Rapid Transit (CART) system and encourage people to ride it. Furthermore, the best act generated by 35 per cent of the subjects was one of the two "second best" acts as rated by the experts. These acts (classifications 1.6 and 1.8) suggested that the University buy additional land to build parking lots and that they build additional remote parking lots which would be linked to campus by a mass transit system. Thus, 50 per cent of the subjects were able to generate an act that was one of the top four acts according to the experts in our sample.

In terms of the student utility estimates, the best act generated by 20 per cent of the subjects was the act the students rated as the best act. This act (classification 5.2) suggested that the University should allow students to park in restricted areas (e.g., faculty lots) after certain hours. The best act generated by 55 per cent of the subjects was one of the two acts rated "second best" by the students. These acts (classifications 3.1 and 3.5) suggested that the University encourage people to carpool and that the

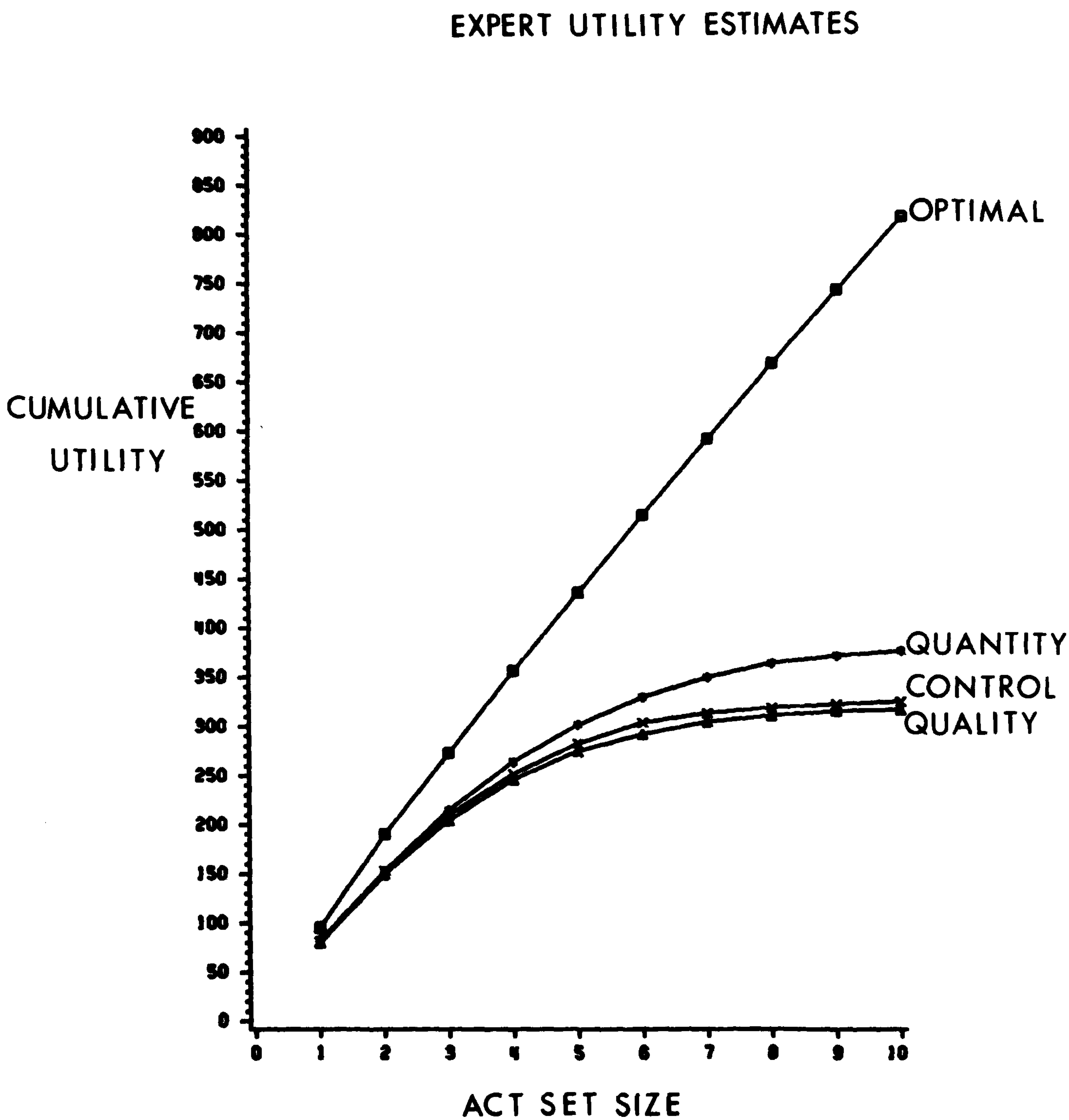


Figure 2. Cumulative utility functions, based on expert median utility estimates, for the three instructional conditions compared to a lower-bound estimate of optimal performance.

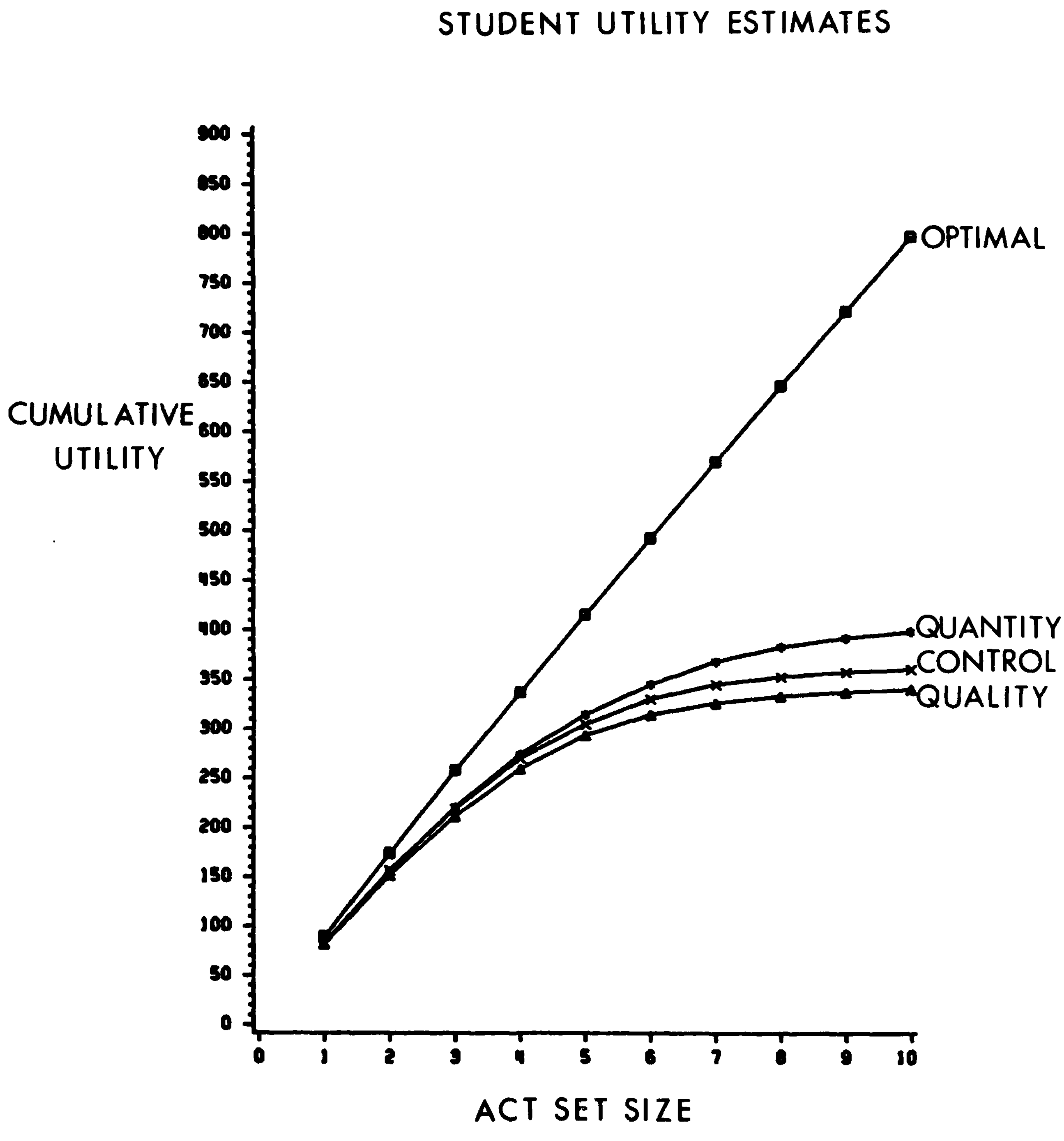


Figure 3. Cumulative utility functions, based on student median utility estimates, for the three instructional conditions compared to a lower-bound estimate of optimal performance.

University should expand the CART system to include more off-campus stops. Thus, 75 per cent of the subjects were able to generate an act in the set of the top three acts as rated by an independent sample of students.

Subjects' performance in terms of the utility of the highest utility act appears to be better when the student median utility estimates are used than when the expert median utility estimates are used. This is what one might expect as the subjects did not have access to the expert's utility structure, but were using their own utilities which were not in complete agreement with those of the experts. Even though the subjects in the Control condition and the Quantity condition were asked to respond with all worthwhile actions, it is reasonable to assume that their utilities were used to some extent to decide if an action was worth entering.

The relatively small difference between the average individual functions and the optimal performance function at the first point of the functions is of considerable interest. Whereas subjects do not always generate the best action, their best action is often one of the better actions generated by the group. Thus, the best of the actions that they generate usually is worthwhile, and probably should be implemented. However, this is not true for all act generators. Some subjects were unable to generate even one high utility act. In fact, the best act generated by 17 per cent of the subjects was 20 utiles or more below the highest utility act as rated by the experts. The best act generated by 7 per cent of the subjects was 20 utiles or more below the the highest utility act as rated by the students.

Completeness of subjects act generation performance

The difference between the average individual cumulative utility curves and the optimal utility curve is a measure of the completeness of the average individual's decision tree. As can be seen in Figures 2 and 3, the average individual cumulative utility functions for the different conditions asymptote as act set size increases, but the optimal cumulative utility function continues to increase. The comparison of interest is the proportion of the possible utility subjects were able to earn for various set sizes, as compared to the lower-bound estimate provided by the group cumulative utility function. This measure is computed by dividing the cumulative utility earned by the average individual by the corresponding value for the entire group (which is the lower-bourd estimate of optimal performance) for each set size. If this proportion were 1.0, then the performance would be optimal. Similarly, if the proportion were 0.0 then the subjects performance would have been valueless. These data are presented in Figure 4.

An inspection of Figure 4 indicates that, on the average, subjects only earned a small proportion of the total possible utility points. The average subject generated a few actions which were of fairly high utility. For example, when the student median utility estimates are used, the two highest utility acts generated by subjects earned close to 90 per cent of the possible utility points. However, subjects' performance begins to deteriorate after this point. On the average, their five highest utility acts only earn them about two thirds of the possible utility points and their ten highest utility acts earn them less than 50 per cent of the possible utility points. We conclude from these results that the **breadth or completeness** of subjects' performance in this regard is lacking. The goal of the task is to generate **all or most** of the viable action alternatives. In this respect, the subject's

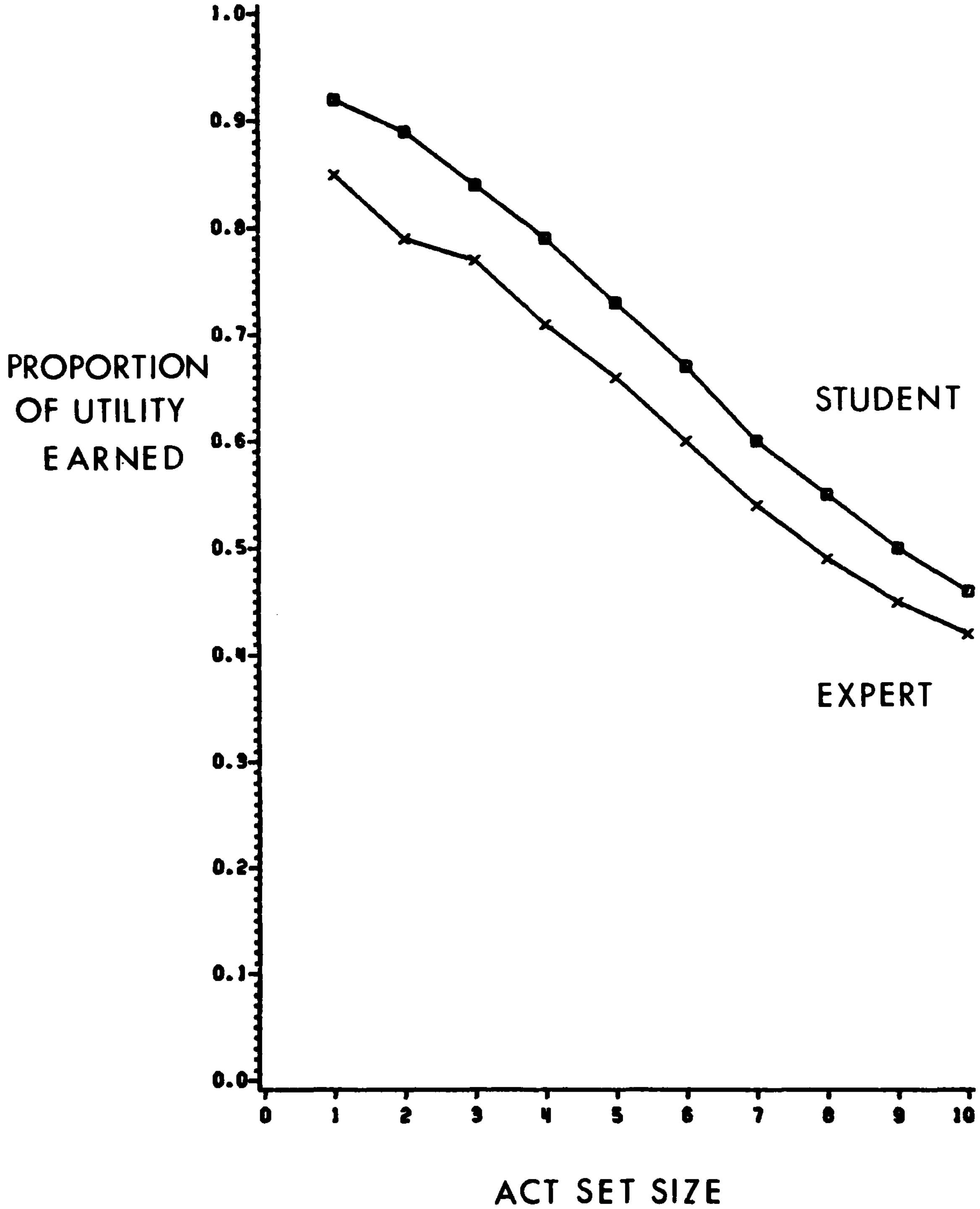


Figure 4. The proportion of the total possible utility points subjects earned plotted as a function of act set size for both expert and student median utility estimates.

performance is impoverished. Even though the typical subject has two or three high utility actions, their set of actions is inferior to that of the lower-bound estimate of possible performance provided by the group.

Subjects assessment of their own performance

The results of the "post-generation" estimates of the number of acts remaining to be generated were quite interesting. Subjects in all three conditions thought they had done quite well. When asked to estimate the number of alternative solutions that they were unable to generate, the median response from the Control condition was 5; the median response from the Quantity condition was 4.5; and the median response from the Quality condition was 6. When asked to estimate how many of these "ungenerated" alternatives were good solutions which should be considered when solving the problem, the median response from the Control condition was 2.5; the median response from the Quantity condition was 3; and the median response from the Quality condition was 3.

There are, in theory, an infinite number of acts which could have been generated to solve the problem. If we rule out minor variations of the same idea, then there are at least 30 or 40 unique acts of positive utility which could have been suggested. It seems that subjects in all three conditions overestimated the completeness of their performance. This finding is consistent with our previous work on hypothesis generation (Mehle et al., 1981) which reports a sequence of studies in which subjects grossly overestimate the completeness of their hypothesis sets while simultaneously generating few hypotheses. In fact, these subjects apparently believed that their performance had made them "fat and happy" when they *should have felt* "thin and worried". Mehle et al. suggested that the same memory processes used in hypothesis generation are also used in assessment of completeness of performance, leading to the paradoxical result that individuals who do poorly think that they have done well. The present result, while incidental to the main purpose of the experiment, suggests that the same phenomenon is present in act generation, a process that almost certainly employs many of the same cognitive mechanisms as hypothesis generation.

Post-experimental questionnaires

The post-experimental questionnaires provided interesting data about some of the possible processes involved in the generation of alternative solutions to the problems. We asked subjects to describe their strategy for generating solutions to the Parking problem. The most frequently reported strategy was the reliance on past experience with similar problems. Other strategies mentioned include: using previously generated solutions to generate new solutions, using external sources of information such as actions other people had mentioned to them or actions they had read in the paper, and trying to think of the problem as their own personal problem, as when they are trying to find a parking space in the morning.

We also asked subjects how they decided they had generated enough solutions to the problem. Subjects must have some type of "stopping rule" which they use to determine when to give up on generating solutions to a problem. Over two-thirds of the subjects reported that they stopped because their memory was "empty;" that is, they simply could not think of any additional solutions. A little over half of the subjects reported that they stopped because they kept coming up with more and more bad ideas. It may be that subjects find themselves generating more and more bad solutions. When the proportion of bad solutions generated approaches 1.0. they may conclude that their memories are "empty".

Summary and Conclusions

The primary purpose of the present experiments was to explore the generality of of the Gettys et al. (Note 3) conclusion that the sets of potential actions generated by subjects were impoverished. Two possible limitations of the previous experiment were investigated: the motivation of the subjects and the methodology used to characterize the quality of performance. We found no major differences in the results when an incentive was introduced or when a different utility estimation procedure was used. Subjects in the present study. like subjects in the Gettys et al. (Note 3) study, failed to generate complete sets of actions. The present study also had subjects estimate the completeness of their own performance. These estimates indicate that subjects overestimated the completeness of their own action sets.

The results of the present investigation indicate that most subjects usually generate at least one reasonable action and sometimes generate several good actions. This level of performance is satisfactory for solving trivial problems where the decision maker often chooses any action that seems satisfactory. but we contend that this quality of performance is not satisfactory for important problems where the opportunity loss of not picking the best action may have serious consequences in human life or extra expense. Furthermore, subjects in all three conditions overestimated the completeness and quality of their act generation performance. This suggests that people are unable to evaluate the quality of the acts they produce. If this is the case, then we cannot assume that subjects will be able to generate at least one quality solution when instructed to do so. In fact, Johnson et al. (1968) specifically instructed their "quality" instruction group to come up with the one best solution. They found that the subjects in their "quantity" condition, who were instructed to generate many solutions, generated higher quality solutions than subjects in their "quality" condition. It seems that when subjects are instructed to produce quality ideas they may become overly critical of their own ideas and fail to consider "objectively" very good ideas which should have been considered.

Proponents of brainstorming (Osborn, 1953) have recognized the potentially stifling effect of evaluation and have encouraged people to delay evaluation of ideas until all possible ideas have been generated. The results of the present study are consistent both with the empirical work of Johnson et al. and the philosophy which guides the brainstorming approach to problem solving. If people are encouraged to generate all possible actions to a decision problem. then at least a few of the actions generated should be acceptable solutions. The more solutions generated, the more likely it is that at least one solution will be worthwhile. Unfortunately. our research indicates that people quit generating solutions long before they have exhausted the set of possible solutions to the problem.

The present investigation increased the generality of the Gettys et al. (Note 3) finding that act generation performance is impoverished by ruling out simple motivational and methodological explanations for the observed result. The results of this study and the results of our previous work on hypothesis generation (see Gettys et al., Note 5 for review) strongly suggest that efforts to improve decision making should concentrate on the predecision process of problem structuring. If a decision maker fails to identify the important actions that could be taken to solve a decision problem and fails to consider the hypotheses about the states of nature that will affect the outcome of taking a particular action, then their subsequent decision is bound to be suboptimal. Additional research needs to focus on how we can improve decision structuring behavior.

Reference Notes

1. Mehle, T. **Hypothesis generation in an automobile malfunction inference task** (Tech. Rept. TR 25-2-80). Decision Processes Laboratory: University of Oklahoma, Norman, OK., February, 1980.

2. Gettys, C., Mehle. T., Baca. S., Fisher. S., & Manning, C. **A memory retrieval aid for hypothesis generation** (Tech. Rept. TR 27-2-79). Decision Processes Laboratory, University of Oklahoma, Norman, Ok., July, 1979.

3. Gettys, C., Manning, C., & Casey. J. **An evaluation of human act generation performance** (Tech. Rept. TR 15-8-81). Decision Processes Laboratory: University of Oklahoma, Norman, Ok., August 1981.

4. Pearl, J., Leal. A., & Saleh. J. **Goddess: A goal-directed decision structuring system.** Cognitive Systems Laboratory, School of Engineering. University of California, Los Angeles (UCLA-ENG-CLS-8034), June, 1980.

5. Gettys, C., Manning, C., Mehle, T.. & Fisher, S. **Hypothesis generation: A final report of three years of research** (Tech. Rept. TR 15-10-80). Decision Processes Laboratory, University of Oklahoma, Norman, Ok., October, 1980.

6. Manning. C. **Describing the representation of decision problems: An application of multidimensional scaling and cluster analysis** (Tech. Rept. TR 15-12-81). Decision Processes Laboratory. University of Oklahoma. Norman, Ok., December, 1981.

7. Winterfeldt, D. **An overview, integration, and evaluation of utility theory for decision analysis** (Tech. Rept. 001597-4-T). Social Science Research Institute, University of Southern California, Los Angeles, Ca., August, 1975.

8. Keeney. R. & Kirkwood, R. **Group decision making using cardinal social welfare functions** (Tech. Rept. 83). Operations Research Center, Massachusetts Institute of Technology. Cambridge, Ma.. October. 1973.

References

Carroll. J. Analyzing decision behavior: The magician's audience. In T. S. Wallsten (Ed.). **Cognitive Processes in Choice and Decision Behavior.** N. J.: Lawrence Erlbaum Associates, 1980.

Corbin, R. Decisions that might not get made. In T. S. Wallsten (Ed.). **Cognitive Processes in Choice and Decision Behavior.** N.J.: Lawrence Erlbaum Associates, 1980.

Gettys, C. & Fisher. S. Hypothesis plausibility and hypothesis generation. **Organizational Behavior and Human Performance,** 1979, 24, 93-110.

Johnson, D., Parrott, G., & Stratton, R. Production and judgments of solutions to five problems. **Journal of Educational Psychology,** 1968. 59,(6, Pt. 2).

Leal, A. & Pearl, J. An interactive program for the conversational elicitation of decision structures. **IEEE Transactions on Systems, Man, and Cybernetics,** 1977, 7, 368-378.

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

Manning, C, Gettys, C., Nicewander, A., Fisher, S., & Mehle, T. Predicting individual differences in generation of hypotheses. **Psychological Reports,** 1980, 47, 1199-1214.

Mehle. T., Gettys, C., Manning, C., Baca, S., & Fisher. S. The availability explanation of excessive plausibility assessments. **Acta Psychologica,** 1981, 49, 127-140.

Osborn. A. **Applied imagination: Principles and procedures of creative thinking.** New York: Charles Scribner's Sons, 1953.

Pitz, G., Sachs, N., & Heerboth, J. Procedures for eliciting choices in the analysis of individual decisions. **Organizational Behavior and Human Performance,** 1980. 26, 396-408.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. **Annual Review of Psychology,** 1977, 28, 1-39.

Wallsten, T. S., (Ed.), **Cognitive processes in choice and decision behavior.** Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

OFFICE OF NAVAL RESEARCH

TECHNICAL REPORTS DISTRIBUTION LIST

Engineering Psychology Programs
Code 442
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
(2 cys)

Mathematics Research Programs
Code 411 MA
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics and Probability Program
Code 411-S & P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Systems Program
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

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

ATTN: CDR K. Hull

Physiology & Neuro Biology Programs
Code 441
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

ONR Eastern/Central Regional Office
Building 114, Section D
666 Summer Street
Boston, MA 02210

ATTN: Dr. J. Lester
Commanding Officer

ONR Western Regional Office
1030 East Green Street
Pasadena, CA 91106

ATTN: Dr. E. Gloye
Commanding Officer

Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375

ATTN: Director

Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D.C. 20375

ATTN: Dr. Michael Melich

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

ATTN: Dr. Robert G. Smith

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

Training Analysis and Evaluation Group
Naval Training Equipment Center
Code N-OOT
Orlando, FL 32813

ATTN: Dr. Alfred F. Smode

Combat Control Systems
Naval Underwater Systems Center
Newport, RI 02840

ATTN: Dr. Jerry Lamb

Opertions Research Department
Naval Postgraduate School
Monterey, CA 93940

ATTN: Dr. Gary Poock

Human Engineering Branch
Code 8231
Naval Ocean Systems Center
San Diego, CA 92152

ATTN: Mr. Warren Lewis

Scientific Adviser
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

ATTN: Dr. A.L. Slafkosky

Naval Material Command
NAVMAT 0722 - RM. 508
800 North Quincy Street
Arlington, VA 22217

Naval Air Systems Command
Human Factors Programs
NAVAIR 340F
Washington, D.C. 20361

ATTN: Commander

Command, Control & Communications
Code D 100A, Development Center
Marine Corps Development &
Education Command
Quantico, VA 22134

ATTN: Mr. Fred M. Griffee

Naval Medical R & D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

ATTN: CDR Robert Biersner

Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

ATTN: Dr. Arthur Bachrach

Naval Health Research Center
San Diego, CA 92152

ATTN: CDR Thomas Berghage

Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 0340

ATTN: Dr. George Moeller

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

CINCLANT FLT HQS
Code 04E1
Norfolk, VA 23511

ATTN: Dr. James McGrath

Navy Personnel Research
and Development Center
Command and Support Systems
San Diego, CA 92152

ATTN: Dr. Robert Blanchard

Human Factors Engineering
Division
Naval Air Development Center
Warminster, PA 18974

ATTN: CDR J. Funaro

Human Factors Engineering Division
Naval Air Development
Warminster, PA 18974

ATTN: Dr. Julie Hopson

Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

ATTN: Mr. Jeffrey Grossman

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

Naval Postgraduate School
Code 55MP
Monterey, CA 93940

ATTN: CDR C. Hutchins

Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and
Systems Research Laboratory
U.S. Army Research Institute
5002 Eisenhower Avenue
Alexandria, VA 22333

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

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Chief Scientist
AFHRL/CCN
Brooks AFB, TX 78235

ATTN: Dr. Earl Alluisi

Admiralty Marine Technology
Establishment
Applied Psychology Unit
Teddington, Meddlesex TW11 OLN
ENGLAND

ATTN: Dr. Kenneth Gardner

Director, Human Factors Wing
Defense & Civil Institute of
Environmental Medicine
Post Office Box 2000
Downsview, Ontario M3M 3B9
CANADA

Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF
ENGLAND

ATTN: Dr. A.D. Baddeley

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
(12 cys)

Defense Advanced Research Projects
System Sciences Office
1400 Wilson Blvd.
Arlington, VA 22209

ATTN: Dr. Judith Daly

Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, CA 93017

ATTN: Dr. Robert R. Mackie

Institute for Defense Analyses
400 Army-Navy Drive
Arlington, VA 22202

ATTN: Dr. Jesse Orlansky

Massachusetts Institute of Technology
Department of Mechanical Engineering
Camridge, MA 02139

ATTN: DR. T.B. Sheridan

Decision Research
1201 Oak Street
Eugene, OR 97401

ATTN: Dr. Paul Slovic

Virginia Polytechnic Institute
and State University
Department of Industrial Engineering
Blacksburg, VA 24061

ATTN: Dr. Harry Snyder

Stanford University
Department of Psychology
Stanford, CA 94305

ATTN: Dr. Amos Tversky

NAS-National Research Council
2101 Constitution Ave., N.W.
Washington, D.C. 20418

ATTN: Dr. Robert T. Hennessy

Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

ATTN: Dr. M.G. Samet

Virginia Polytechnic Institute
and State University
Human Factors Laboratory
130 Whittemore Hall
Blacksburg, VA 24061

ATTN: Dr. Robert Williges

University of Southern California
Director, Social Science Research
Institute
Los Angeles, CA 90007

ATTN: Dr. Ward Edwards

University of Colorado
Institute of Behavioral Science
Room 201
Boulder, CO 80309

ATTN: Dr. Kenneth Hammond

Catholic University
Department of Psychology
Washington, D.C. 20064

ATTN: Dr. James H. Howard, Jr.

Rice University
Department of Psychology
Houston, TX 77001

ATTN: Dr. William Howell

University of Illinois
Department of Psychology
Urbana, IL 61801

ATTN: Dr. Christopher Wickens

Bolt Beranek & Newman, Inc.
Information Sciences Division
50 Moulton Street
Cambridge, MA 02238

ATTN: Dr. Richard W. Pew

University of Chicago
Graduate School of Business
1101 E. 58th Street
Chicago, IL 60637

ATTN: Dr. Hillel Einhorn

Duke University
Graduate School of Business
Administration
Durham, NC 27706

ATTN: Dr. John Payne

Decision Research
1201 Oak Street
Eugene, OR 97401

ATTN: Dr. Baruch Fischhoff

University of Virginia
School of Engineering and
Applied Science
Charlottesville, VA 22901

ATTN: Dr. Andrew P. Sage

University of Wisconsin
Department of Psychology
Madison, WI 53706

ATTN: Dr. Lola Lopes

Alphatech, Inc.
3 New England Industrial Park
Burlington, MA 01803

ATTN: Mr. Joseph G. Wohl

Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church VA 22043

ATTN: Dr. Rex Brown

Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

ATTN: Dr. Wayne Zachary

Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

ATTN: Dr. Marvin Cohen

LME

-8