J. MICHAEL SHAUGHNESSY

MISCONCEPTIONS OF PROBABILTY: AN EXPERIMENT WITH A SMALL-GROUP, ACTIVITY-BASED, MODEL BUILDING APPROACH TO INTRODUCTORY PROBABILITY AT THE COLLEGE LEVEL

1. INTRODUCTION

Anyone who has recently had the opportunity to teach elementary probability and statistics to college students is probably aware of some misconceptions that these students possess prior to, and possibly after, a formal course in probability. Students often believe, for example, that when tossing three coins, the probability of getting 3 heads is 1/4, since the outcomes are 3H, 2H and 1T, 1H and 2T, and 3T. Outcomes are frequently judged to be equally likely by the beginning probability student. Students also tend to feel that The probability of a tail is greater on the seventh toss after a run of six heads. This belief, known as the Gambler's Fallacy, arises because people expect even short runs of coin flips to reflect the fairness of a coin. One author [1] has referred to this phenomenon as the "Law of Small Numbers."

Students also have misconceptions about probability due to their inexperience with the rapid growth and decay of combinatorial expressions. It has been found [2] that college students who are unfamiliar with counting techniques greatly underestimate the number of possible batting orders of nine players on a baseball team. Combinatorial growth tends to be underestimated by students who are inexperienced with counting principles. Similarly, college students tend to believe that a group of 180 or more people are needed so that the probability that at least two people from the group have the same birthday exceeds 1/2. Factors affecting combinatorial decay tend to be overestimated by college students who are inexperienced with counting principles [2].

Some misconceptions of probability may be of a mathematical sort, the result of a person's inexperience with the mathematical laws of probability. Thus, it may be possible to clear up such misunderstandings by familiarizing a person with the concepts of sample space, simple probability, counting principles, independent events, and uniform and non-uniform probability distributions. However, there is considerable evidence that suggests that misconceptions about probability are sometimes of a psychological sort, and that mere exposure to the laws of probability may not be sufficient to overcome some misconceptions of probability. Cohen [3, 4], Cohen and Hansel [5], Edwards [6], Fischbein *et al.* [7–9], and Kahneman and Tversky [1, 10–13] are among those authors who have investigated peoples' understanding of probability from a psychological point of view. The studies of Daniel Kahneman

Educational Studies in Mathematics 8 (1977) 295–316.

and Amos Tversky provided motivation and a psychological basis for the author's own investigations. Thus it is fitting to describe some of their work in more detail.

2. THE USE OF HEURISTICS IN ESTIMATING THE LIKELIHOOD OF PROBABILISTIC EVENTS

Kahneman and Tversky claim that people who are combinatorially naive utilize certain strategies when they are asked to estimate the likelihood of complex probabilistic events. These strategies are devices by which a person can reduce a complex probability problem to a simpler problem. Kahneman and Tversky call these strategies *heuristics.* According to Kahneman and Tversky, the use of heuristic principles in estimating the likelihood of a probabilistic event may be a helpful simplifying device and may, indeed, lead a person to a reasonable estimate for the probability of an event. However, Kahneman and Tversky also propose that the use of heuristics may lead to bias and systematic error in probability estimates.

Two specific heuristics, or strategies, which Kahneman and Tversky characterized in their research are called the *representativeness heuristic* and the *availability heuristic.*

According to the *representativeness heuristic,* people tend to make decisions about the likelihood of an event based upon how similar (i.e. representative) the event is to the distribution from which it was drawn, or upon how similar the event is to the process by which the sample space is generated [7]. For example, a long string of heads does not appear to be representative of the random process of tossing a coin. People who were employing the representativeness heuristic would tend to believe that tails will be more likely to occur than heads on a subsequent toss since the process is random and the number of heads and tails should 'even out.' In a similar way, a person who judges that the probability is 1/2 that the outcome 2H and 2T will occur from tossing four coins may be using representativeness to make his estimate. The event 2H and 2T appears to be representative of the distribution of heads in the population of outcomes for flipping one coin once.

Kahneman and Tversky [10] summarize the pervasiveness of the representativeness heuristic by indicating that representativeness can be shown to account for fallacies in prediction that arise from:

(i) insensitivity to prior probabilities and disregard for population proportions,
(ii) insensitivity to the effects of sample size on predictive accuracy,
(iii) unwarranted confidence in a prediction that is based upon invalid input data,
(iv) misconceptions of chance, such as the Gamblers' Fallacy,
(v) misconceptions about the tendency for data to regress to the mean.

Availability may also be used in assessing the likelihood of an event. According to the *availability heuristic,* people tend to make decisions about the likelihood of an event based upon the ease with which instances of that event can be constructed or called to mind [9]. For example, when asked whether there are more distinct three-person committees or more distinct nine-person committees that can be formed from a group of twelve members, people who rely upon the availability heuristic tend to guess that there are more possible three-person committees. It is easier to construct instances of three-person committees than nine-person committees. Similarly, if a person is asked to estimate the local divorce rate or to estimate the probability of being involved in an automobile accident, the frequency of his/her personal contact with these events (perhaps through friends or relatives) may influence the probability estimates. Kahneman and Tversky [9] claim that the availability heuristic causes systematic bias in probability estimates because people tend to believe that those outcomes which can easily be brought to mind will also be more likely to actually occur.

In summary, Kahneman and Tversky suggest that combinatorially naive people often do not apply the theory of mathematical probability in estimating the likelihood of events. Instead, people tend to employ heuristic principles such as representativeness or availability when they are asked to make probability estimates.

Most of the subjects involved in this series of studies by Kahneman and Tversky [1, 7–9] were combinatorially naive college students who had no prior training in probability or statistics. It is not altogether surprising that these students were found to rely upon such heuristics as representativeness and availability when predicting the likelihood of events. However, Kahneman and Tversky [1] also found that psychologists who had a substantial background in probability and statistics were subject to the same types of misconceptions about probability as the combinatorially naive college students. Evedently, exposure to the theory of probability and statistics is not necessarily sufficient to overcome the systematic biases that arė induced by the heuristics of availability and representativeness. In fact, Kahneman and Tversky found such strong and widespread evidence for the use of representativeness and availability in estimating the likelihood of events that they suggest that misconceptions of probability and statistics that arise from the use of these heuristics may be very difficult to overcome.

3. THE PROBLEM

The investigations of Kahneman and Tversky, Edwards, Cohen and Hansel, and others pose a challenge to mathematical education. The evidence from these studies indicates that students enter a course in elementary probability and statistics with a set of misconceptions about probability and statistics, and with a set of heuristic principles that could propagate and maintain their misconceptions about probability and statistics. Prior to any formal course work in probability, students have had experience in and have dealt almost exclusively with 'subjective probability.' Suddenly, students are confronted in a formal course with a completely mathematized model of what Carnap [14] calls 'statistical probability.' The challenging problem for mathematical education that arises is this: *How should elementary probability and statistics be taught so as to maximize the students' chances of overcoming their misconceptions of probability and statistics*? In particular, is there an effective way of teaching elementary probability so that students would learn to rely upon probability theory in making estimates for the likelihood of events rather than relying upon heuristic principles which may bias probability estimates? The author does not presume to have a complete answer to this problem. He would, however like to suggest that a conventional lecture approach to the teaching of elementary probability and statistics may not be the best way to overcome students' misconceptions about probability. The author would like to present some experimental evidence that suggests that a small-group, activity-based, model building approach to elementary probability and statistics can help undergraduates to overcome some of their misconceptions about probability, and can reduce reliance upon heuristics such as availability and representativeness.

4. AN EXPERIMENT

During the 1976–1976 academic year an experimental activity-based course in elementary probability and statistics was developed by the author at Michigan State University. Comparisons were made between groups of college students who took this experimental course and groups that took a lecture-based course in elementary probability. The comparisons were made to test the relative effectiveness of a lecture-based approach in overcoming certain misconceptions that college students have about probability. The misconceptions that were investigated were those that arise from reliance upon the heuristics of representativeness and availability. A detailed report of the statistical results of the experiment and comparisons made between the groups can be found in [2]. It was found that the experimental course was more effective in overcoming some misconceptions that are attributable to the use of representativeness and availability. In this paper we shall describe the experimental course and discuss the results of the subjects' responses to several of the items that were used to test for misconceptions of probability.

The experimental activity-based course was constructed as an alternative to the lecture method for an undergraduate course in finite mathematics. A series of nine activities in probability, combinatorics, game theory, expected value, and elementary statistics were developed by the author. Students in the experimental course worked together in class on the activities in small groups of four or five members. Each activity required the groups to perform experiments, gather data, organize and analyze the data, and finally reach some conclusions which could be stated in the form of a mathematical principle or mathematical model. The students were strongly encouraged to co-operate with one another, to solve problems as a group rather than individually, and to help all the members of their group to understand the concepts and problems of each activity. The groups were changed often so that everyone had a chance to work with everyone else during the course.

The role of the instructor in the experimental course was that of organizer, diagnostician, devil's advocate, and critic. During each activity the instructor circulated among the groups, clarifying students' questions and assisting groups which had stalled on a particular problem. Instructor assistance usually came in the form of questions put to the groups, questions that were intended to lead the group from a point which they already understood step-by-step back up to the source of their original question. This technique of 'answering' a question with another question was used in order to encourage the groups to work problems out for themselves and to keep the investigation on each activity guided, but also as open-ended as possible.

Several texts were used to supplement and re-inforce the in-class activities. The texts used in the experimental course were *Statistics by Example: Exploring Data* and *Weighing Chances* [15]; *Fifty Challenging Problems in Probability* [16]; and *How to Lie With Statistics* [27]. Reading assignments and problem assignments were selected from these texts for work outside class. Each student was also required to write ten careful critiques of instances that he (she) found of misuses of statistics. The misuses could be from advertisements, newspaper or magazine articles, or any other source that the student could uncover.

The lecture-based course was taught from a text called *Finite Mathematics* [Weiss and Yoseloff, 18]. The mathematics content of both the experimental and lecture courses was quite similar, although the order of the topics was different. The experimental course introduced some elementary statistics in place of a segment that was devoted to linear programming in the lecture course. A summary of the order of the topics and the time spent on the topics in each course is given in Table 4.1.

A small-group, problem-solving and model-building approach was undertaken in the experimental groups for two reasons. First, perhaps the transition for

TABLE 4.1
Order and duration of topics

Experimental course		Lecture Course	
(1st) Probability Models (2nd) Counting Principles	4 ½ weeks	(1st) Counting Principles (2nd) Probability Models	4 ½ weeks
(3rd) Game Theory	2 weeks	(3rd) Linear Programming	3 weeks
(4th) Statistics	3 weeks	(4th) Game Theory	2 weeks

students from preconceptions and misconceptions of probability to mathematizations of probabililtic laws can be facilitated if students are encouraged to experience elementary probability and statistics as a *process* of describing observed experimental phenomena more and more accurately, rather than as a *system* of rules, axioms, and counting techniques that must be learned and applied to problems. Second, the process of building models gets students involved in a part of applied mathematics that is sorely neglected in lower level mathematics courses. Pollak [19], Thompson [20], Klamkin [21], Fitzgerald [22], and Freudenthal [23] have all made pleas for teaching mathematics in such a way that students can build their own mathematical models and think problems through for themselves.

5. BACKGROUND OF THE SUBJECTS

In the Spring term of 1976, students at Michigan State University registered into seven sections of a finite mathematics course. Four sections were randomly selected for this experiment. The sections were randomly assigned to either the experimental activity-based course or to the lecture-based course in finite mathematics. The subjects consisted of 80 college undergraduate students, 48 men and 32 women. Personal background information was gathered from a form filled out by each student. A summary of class level, major field of study, previous college mathematics courses, and prior exposure to probability and statistics is given in Tables 5.1 and 5.2. The course in finite mathematics at Michigan State University was created primarily to serve the needs of students who were majoring in business, agriculture, or biology. The subjects in this study were primarily freshmen business or accounting majors [Table 5.2]. There is a prerequisite course in College Algebra which all of the subjects had successfully completed. In addition, a substantial number of the subjects (51) indicated that they had taken at least one of two secondary-school level remedial algebra courses prior to enrolling in the College Algebra [Table 5.1].

TABLE 5.1
Previous mathematics course work

Group	N	Math 101	Math 102	Previous Probability
L_1	26	1	14	2
L_2	14	2	6	1
E_1	20	5	14	2
E_2	20	1	17	2
TOTAL	80	9	51	7

L = lecture-based E = Activity-based

TABLE 5.2
Class level and major

Group	Freshman	Upperclassmen	Business majors	Accounting majors	Other majors[a]
L_1	19	7	7	9	9
L_2	9	5	3	3	8
E_1	9	11	8	4	8
E_2	14	6	10	1	10
TOTAL	51	29	28	17	35

[a] Primarily agriculture majors or as yet undecided about major field.

These courses, Mathematics 101 and 102, correspond respectively to a first year and third year high school algebra course. The subjects in this study, therefore, definitely did not have strong backgrounds in mathematics. Exposure to probability prior to the course was minimal within the groups, as only seven students in the sample indicated that they had had any previous work in probability.

6. THE EXPERIMENTAL COURSE

The nine activities that were carried out in small groups of four or five students in the experimental sections provided the focal point for the experimental course. The day-by-day occurences within one of the experimental groups were recorded in a journal kept by the experimenter. This section describes the activities and discusses some of the observations made by the experimenter while the activities were being performed in small groups by the students in one of the experimental sections.

Activity 1. This activity begins by asking the members of each small group to guess the probability of getting various number of heads in tossing six coins. The groups then performed the experiment 48 times, recording the number of heads. Experimental probabilities for the outcomes 6 heads, 5 heads, . . ., 0 heads were calculated from the data using the relative frequency model.

The groups had difficulty setting up a theoretical mathematical model for this experiment. The students did not have a clear concept of what a mathematical model involved during this first activity. Furthermore, the groups could not agree among themselves how to list the outcomes. Some students felt that only the number of heads should determine 'an outcome'. For these students there were seven outcomes, from 0 heads to 6 heads. There were other students who felt that the *position* of the heads among the six coins changed the outcome. The issue was debated hotly in the small groups. Eventually the first approach to a model was abandoned, for it suggested assigning a probability of 1/7 to each of the outcomes, and the experimental data had indicated that it was unlikely that the outcomes '6 heads' and '3 heads' were equally likely to occur. (The outcomes 6 heads occured only once in the pooled small-group experimental data.) The second approach to a model was adopted, but it soon became apparent to the students that there were a large number of outcomes to list. The first attempts by the groups at listing the outcomes as sequences of heads and tails failed because they had not yet developed a systematic way of enumerating the outcomes. Gradually, the groups discovered that if they held values of some of the coins fixed while changing others, the list of outcomes became more manageable. The groups, of course, had no introduction to counting principles at this time in the course.

When the 64 outcomes in the model had been listed, theoretical probabilities for the number of heads were calculated. Many students were surprised that their guesses for the probabilities were so far off. Over half the students had guessed that the probability that three heads would occur was at least 1/2. The probability of three heads based upon their experimental data was about .30, and was 20/64 based upon their mathematical model. They were also surprised that the probability of 6 heads was so small. Only a few students had estimated the probability of 6 heads to be below 10%.

Activity 2. In the second activity, a model for tossing three tacks, listing the outcomes, and assigning probabilities to the outcomes was developed by the students. The groups first had to find an estimate for the probability P(U) that a thumbtack lands point up. The range of values for P(U) obtained by the students in 72 trials was from 0.48 to 0.76. As a result of the wide range of outcomes for P(U), a discussion arose concerning the factors that may have affected the outcomes – the way the tack was dropped, the height from which it was dropped, the surface on which it landed, etc. The class decided to rerun the experiment to estimate P(U) while attempting to control for as many nuisance variables as possible. The class decided that P(U) was approximately 2/3 after careful 'quality controls' were introduced into the experiment of tossing the tack. There was agreement among the subjects that the outcomes point up (U) and point down (D), tack on its side, were *not equally likely to occur,* based upon the experimental evidence. However, when the subjects constructed a mathematical model for tossing three of these tacks, each outcome was assigned a probability of 1/8. The groups had actually performed the experiment of tossing the three tacks, and had experimental evidence that the 8 outcomes were *not* equally likely to occur, since UUU, UDU, and UUD occurred much more frequently than did DDD, DDU, or DUD. Nevertheless, the subjects persisted in their belief that the eight outcomes should be equally likely. The subjects wrote in their logs that their was probably something wrong with the tacks. '*Theoretically,* U and D should be equally likely, even though *experimentally* they were not,' was written in many of the students reports on this activity. This feeling among the students that every probability model was really a uniform model was difficult to overcome. Manifestations of this belief persisted throughout the experimental course. The instructor assisted the groups in discovering a model for the non-uniform case by means of a series of questions.

Instructor:	Suppose that three tacks were tossed on the table 1200 times. You have decided that P(U) = 2/3. in how many of those 1200 tosses would you expect to find the *first* tack land upright?
Student:	800, because that is 2/3 of 1200.
Instructor:	Now, of those 800, in how many cases would expect to see the second tack land down?
Student:	1/3 of the 800.

In this manner, the model of multiplying probabilities of independent outcomes was slowly elicited from the groups.

Activity 3. This activity on modeling the outcomes for tossing three dice was similar to the coin and tack experiments. The difficulties with 'equally likely vs. unequally outcomes' or 'the best way to model the outcomes' (as ordered triples or as the sum of the three faces) that appeared in the first two activities were not as troublesome for the students in activity 3. The subjects were not happy about having to list the 216 outcomes in order to make a mathematical model. However, many of them discovered patterns during the process of making their list of outcomes for this experiment, or noticed the symmetry of the frequency distribution. This simplified the job of listing the outcomes. At the conclusion of this tedious listing process, the subjects were demanding that we investigate some 'easier' way of counting the outcomes for an experiment.

Activity 4. The fourth activity was constructed to lead subjects to discover several counting principles for themselves. The instructor gave a brief talk on counting, and led the students to a point where they could state the sequential principle in their own words, the multiplication principle.

The activity began with a series of questions on 'spelling problems'. How many distinct words can be written from the letters L A Z K using each letter once, if every arrangement of the letters spells a 'word' in our language? The outcomes for the first few problems were listed longhand. Eventually the students began to see that the sequential counting principle could help them list the outcomes for the number of words that could be spelled from the letters L Z A K E. It was $5 \times 4 \times 3 \times 2 \times 1 = 120$, using each letter once. It took the groups a long time to discover what to do when some of the letters occurred more than once. If the letters were L Z A K L, or L Z A L L, the first conjecture made in each group was that only 1/2 (respectively 1/3) of the 120 possibilities would actually be distinct. The instructor encouraged the subjects to list the outcomes. When only 20 outcomes for words from L Z A L L could be found, the groups began searching for an alternative explanation. They finally discovered that although the first L in L Z A L L accounted for three redundancies per word, the second L still accounted for two additional redundancies per word. Thus, they first divided the total number of 120 arrangements of the five letters by 3, and then reduced the remaining 40 by one-half, for the number of redundancies from the next L. After many examples of this sort, the groups produced the formula $n!/(k_1!k_2! \ldots k_r!)$, where n is the total number of letters in the word, and k_i is the number of repetitions of the ith letter. The subjects were elated when they discovered this formula. The room was filled with triumphant smiles.

Other counting formulae were developed as special cases of this formula, and a wide variety of counting problems were then presented to the groups in which the principles they discovered were applied in different ways. The process of counting the number of groups of x people that could be chosen from a group of y people was seen to be equivalent to the process of counting the number of distinct words that could be spelled with x C's and $(y - x)$ N's, where C stands for 'chosen' and N for 'not chosen'.

Activity 5. In this activity three games were played by pairs of students in order to provide an introduction to two-person game theory. In the first game one or two fingers were thrown by the players. Player 1 received payoffs of \$10 or \$30 when different numbers of fingers were shown. Player two received \$20 whenever there was a match. After playing the game 20 times, the results in the small groups indicated that there was a tendency for player two, the matcher, to win. The subjects generally attributed the success of the winner to his ability to 'psyche out' the other player and guess what the other player would do. None of the students indicated that they thought the game was rigged in favor of the matcher. The instructor suggested that the game be simulated with random choices to make it difficult for either player to pick up a pattern in the other player's strategy. Every pair of students simulated an equally likely game, making their choices on a 50:50 schedule with coin tosses. It did not occur to the subjects that perhaps a 50:50 schedule was not in the best interest of both players.

The advantages of carefully alternating among the choices in a game became more apparent to the subjects when they played the second game. This 4 x 4 two-person game had black and red cards from an ordinary deck as the entries in the payoff matrix. Player-black tried to make a choice that would result in obtaining points equal to the face value of a black card in the matrix, player-red did the same but for red cards. Choices were made independently and then the entry in the chosen row and column tallied for the winning color. There were so many more choices in this game than in the first game that the students began to develop strategies for playing. Rows or columns with too many of the opponents' entries were disdained, or altogether avoided. There was a tendency to pick the 'safer' rows (or columns) which had two cards of each color. High payoff cards like nines and tens that were imbedded in a row that otherwise contained all opponent's cards were only occasionally gambled upon. The beginnings of some naive 'mixed strategies' were used by the students in this game.

The last game was a 4 x 4 game that contained a saddle point. pairs of students decided upon a strategy in this game. Of the 8 pairs of students who played this game, 5 pairs chose the saddle point, and the other three pairs picked one of the two co-ordinates of the saddle point. The students had no game theory prior to this activity. At the end of the activity the subjects were already displaying some intuition for both 'mixed' and 'pure' strategies in their choices while playing a two-person game.

Activity 6. This activity on expected value consisted primarily of working out solutions to problems and games in order to calculate the long-run payoffs. A brief lecture on expected value of a 2 x 2 game was given by the instructor. The optimal strategy for playing a 2 x 2 two-person game was simulated for several different games. The subjects were surprised at how close the mean payoff in 25 plays came to the theoretical payoff calculated from the optimal mixed (or pure) strategy.

Activity 7. This was the first of two activities on the effects of sample size upon measure of central tendency and variability. The subjects guessed the number of cards that they would have to turn over to have at least a 50% chance of getting at least one ace from an ordinary well-shuffled deck of cards. The guesses were mostly from 12 to 15 cards. Experimental data on the number of cards necessary to obtain an ace were gathered for sample sizes of 10, 20, and 100 trials. The median was used by the subjects as an estimate for the number of cards necessary to have at least a 50% chance of getting an ace. Medians for sample size 10 ranged from 4 to 13, for sample size 200 from 5 to 9, and the median for sample size 100 was 7. Techniques learned in earlier activities (2 and 4) were used to calculate the theoretical value at 9.

The guesses made by the subjects indicated that they were more aware of the deceptive nature of the probability of disjunctive events than they had been at the beginning of the course. Only one student guessed that it would take 26 cards to have at least a 50% chance. A pre-test question administered to all the subjects prior to the course asked for an estimate of the number of people needed so that there was at least a 50% chance that two people had the same birthday (classical birthday problem). 62 out of the 80 subjects responded that it would take 183 or more people. The tendency to use 50% as a representative multiplier [10] of the total population had practically disappeared in the experimental group by the time activity 7 was carried out, about 6 1/2 weeks into the experimental course.

Activity 8. Means and standard deviations for sets of two-digit numbers of various sample sizes were calculated in this activity. The samples of size 5 yielded means from 31 to 71, while samples of size 25 had means from 43 to 52. The parameters were calculated for random numbers chosen from the set 00 to 99. The standard deviations calculated for samples of size 5 and 25 resulted in a similar 'narrowing' of the range of observations in the samples of larger size. The subjects hypothesized that measures of central tendency and variability are rather unstable for small samples, and may not be very accurate indicators of the true population parameters.

Activity 9. The students were presented with a challenge which was much less structured than the first eight activities. The problem was to design and carry out on experiment to test the truth of the statement 'Pulse rates go up when taken by a member of the opposite sex.'* The design of the experiment was set up by the experimental class during an open class discussion. The class decided

* Courtesy of Professor William Fitzgerald, Michigan State University.

to carry out the experiment on themselves. Pulses were taken on the temple or the neck in order to maximuze the chance of raising the pulse rates. Each person in the class took his (her) own pulse first. The pulse-by-self outcome was used by the students as a basis of comparison with pulse rates found by members of the same sex and members of the opposite sex. The data were organized into 2 x 2 contingency tables of the forms below.

TABLE A

	up	not up
males		
females		

TABLE B

	up	not up
same sex		
opposite sex		

The students used chi-square statistics (learned in a reading assignment from *Statistics by Example* [15]) to test for the independence of males vs. females or same sex vs. opposite sex with respect to raising the pulse rates. They found no significant differences. However, they had a lot of fun conducting the experiment.

This activity was performed at the end of the course after the students had already analyzed many articles in newspapers and magazines for misuses of statistics. Written reports on this activity contained the following sorts of suggestions and criticisms of the pulse-rate experiment.

1. We all knew each other and that may have biased the results. It would have been interesting to have done this activity at the beginning of the course to see if there were any differences.

2. It would be better to have one fixed person of each sex take everyone's pulse. This should be done by a very handsome man and a very beautiful woman and they should be expert at taking pulses. We are not very good at taking pulses and this may have biased the results.

3. Knowing who is taking your pulse might affect the pulse rate. Thus the design of this experiment might not really help answer the original question. The subjects should be blindfolded so that any bias that might occur from knowing the pulse-taker could be controlled. (On the other hand, being blindfolded might make your pulse go up independently of having your pulse taken by a person of the opposite sex, so this might bias the experiment also.)

The subjects in the experimental class realized that their experiment admitted many sources of bias due to uncontrolled nuisance variables.

At the end of the experimental course the students in the experimental sections were asked to respond to a questionnaire. They were asked to comment on working in small groups, keeping a log of all their reports on activities and problems, the texts, what they liked about the course and what they disliked. Responses were generally in the form of a letter to the instructor. There was a strong feeling among the students that the log kept by each student was essential to the course. It provided a study guide, a reference book, and a tremendous sense of accomplishment for the students in the experimental course. It was also universally agreed that working in groups was an excellent way to learn mathematics. Interaction and co-operation in solving mathematics problems were new experiences for these students. Their comments on the evaluation forms indicated that they thoroughly enjoyed working in small groups. Several students did mention that a few of their group members had a tendency to rely on other people's work and to not contribute to the group very much. However, most of the students were very active and co-operated well in the groups. Several evaluations mentioned that the activities really helped to 'prove' the theory that was being learned in the course.

Overall attitude towards the experimental course was very positive. Almost every evaluation indicated that the students had enjoyed the class. Several students wrote that they were 'amazed to think that they had actually enjoyed a mathematics class.' Initial frustration at not having the answers or rules or formulas provided for them by the instructor had disappeared for most of the students by the end of the course.

7. RESULTS ON SOME REPRESENTATIVENESS AND AVAILABILITY ITEMS

The 80 subjects were pre-tested and post-tested on instruments developed by the author. The instruments tested for knowledge of some probability concepts and for reliance upon representativeness and availability in estimating the likelihood of events. Many items were similar to or the same as items used by Kahneman and Tversky in their research [10, 12]. The results on these items provided some measure of the subjects' use of heuristics vs. use of probability theory to estimate probability, both before and after exposure to probability via one of the two courses, lecture-based or activity-based. The subjects were asked to *supply a reason* for each of their responses. In this way it was possible to gain some insight into the thinking process of the students as they answered the questions.

A thorough analysis of the results of the experiment can be found in [2]. The experimental activity-based classes were more successful at overcoming reliance upon representativeness ($p < 0.05$, df $= 2$) and tended to be more successful at overcoming reliance upon availability ($p < 0.19$, df $= 2$). The results of several of the 'representativeness' and 'availability' items are reported below.

Personal background information indicated that the subjects for this experiment were weak in mathematics, and did not have very positive attitudes towards mathematics. A majority of the subjects had taken a remedial high-school level algebra course in college prior to taking this course. Furthermore, the post-test was administered at the end of the term, about four weeks after the subjects had studied probability principles. The subjects were also told that the test items would in no way count towards their grade in the course. They had only 50 minutes to answer about 20 questions and supply a reason for their answer in each case. The students' background, time constraints, time elapsed since studying probability and the fact that the test did not count should all be considered in interpreting the results. It is likely that each of these factors contributed to some of the failures on the items.

Some Representativeness Items

R1: The probability of having a baby boy is about 1/2. Which of the following sequences is more likely to occur for having six children?

(a) B G G B G B (b) B B B B G B (c) about the same chance for each

Give a reason for your answer.

TABLE 7.1

	Pretest responses to R1			Posttest responses to R1		
	BGGBGB	BBBBGB	Same	BGGBGB	BBBBGB	Same
Lecture	27	0	9	22	1	17
Experimental	23	2	9	12	2	24

(Entries in the tables represent the frequencies of the responses)

Students who chose B G G B G B indicated that this sequence fit more closely with a 50 : 50 expected ratio of boys to girls. B G G B G B appears to be a more 'representative' outcome than B B B B G B, although the outcomes are equally likely to occur. Reliance upon the representativeness heuristic was heavy on the pre-test, 50 out of the 80 subjects chose B G G B G B (Table 7.1). The posttest responses indicated that there was less reliance upon representativeness in the experimental activity-based groups.

R2: (same assumptions as R1) Which sequence is more likely to occur for having six children?

(a) B G G B G B (b) B B B G G G (c) about the same chance

Give a reason for your answer.

TABLE 7.2

	Pretest responses to R2				Posttest responses to R2			
	BGGBGB	BBBGGG	Same 3 B's	Same correct	BGGBGB	BBBGGG	Same 3 B's	Same correct
Lecture	15	0	11	8	9	1	13	16
Experimental	13	0	12	9	4	0	8	26

In their research Kahneman and Tversky [10] did not offer the option 'the same chance' on items similar to this one. They found that subjects heavily favored B G G B G B, since B B B G G G does not appear to be representative of the random process of having children. In the present study, many students picked 'the same chance', but gave as a reason "because each outcome has 3 boys and 3 girls." This reasoning is also indicative of reliance upon representative of the expected number of boys, 1/2 of the number of trials. Table 7.2 shows that there was a greater tendency for the lecture-based students to still be trapped by representativeness on this item after a formal course in probability.

R3: What is the probability that in 6 children, 3 will be girls? (same assumptions as R1)

Give a reason for your answer.

TABLE 7.3

	Pretest responses to R3			Posttest responses to R3		
	1/2	20/64	Other	1/2	20/64	Other
Lecture	23	0	10	19	0	21
Experimental	23	1	8	9	8	23

On the pretest, 46 of the 80 subjects indicated that the probability of 3 boys and 3 girls was 1/2 (or 50 : 50, 50%, etc.). These subjects, who were at the time naive about the binomial distribution, used the probability of one head on one toss as a representative estimate for the distribution of heads in multiple trials. The results of the posttest indicate that the experimental groups has much less tendency to employ the representative predictor 1/2 than the lecture groups did.

On each of the items R1, R2, and R3, the greater success of the subjects in the experimental groups on the posttest could (hopefully) be the result of actually performing activity 1 with the six coins. As mentioned earlier, mere exposure to probability concepts in a lecture format is not likely to be sufficient for overcoming the strong influence of the representativeness heuristic. The difficulty in overcoming representativeness is suggested by the number of failures on these three items, even in the experimental course.

R4: Which is more likely to occur?

(a) Pulling one red ball from a jar containing 10 red and 90 white balls or (b) Pulling four red balls in a orw from a jar containing 50 red and 50 white balls? (with replacement)

Give a reason for your answer.

TABLE 7.4

	Pretest Results on R4			Posttest Results on R4		
	(b) 4 red in a row	(a) correct reasoning	(a) guessed	(b) 4 red in a row	(a) correct reasoning	(a) guessed
Lecture	23	10	6	17	10	12
Experimental	22	10	6	8	26	6

On the pretest, the subjects tended to use the probability of getting one red ball on one pull from the 50 : 50 distribution as a representative predictor of the probability of four reds in a row. Many students wrote that the probability of getting 4 reds in a row was actually equal to 1/2. Table 7.4 shows that the experimental groups improved much more than the lecture groups on this item.

R5: The chance that a baby is born a boy is about 1/2. Over the course of an entire year, would there be more days when at least 60% of the babies born were boys in.

(a) a large hospital (b) a small hospital (c) makes no difference

Give a reason for your answer.

TABLE 7.5

	Pretest Responses on R5			Posttest Responses on R5		
	Small	Large	No Difference	Small	Large	No Difference
Lecture	6	9	25	14	3	22
Experimental	11	6	23	27	1	12

48 of the subjects chose 'makes no difference' on the pretest. This supports Kahneman and Tversky's claim that subjects tend to believe that both hospitals should be 'equally representative' of the population proportion of boys to girls.

Thus, for most of the subjects, sample size makes no difference. Reliance upon representativeness diminished substantially in the experimental groups on this item on the posttest. 27 of the 40 subjects in the experimental groups correctly chose the small hospital. The lecture-based groups did not exhibit this change on the posttest. In activities 7 and 8 the experimental groups had performed experiments to test the effects of sample size upon measures of central tendency and variability. The lecture-based groups had not carried out such experiments. Performing the experiments in these activities may have contributed to the greater success of the experimental groups on this item.

Some Availability Items

A1: Consider the grids below.

Grid A	Grid B	Are there:
XXXXXXXXX	XX	(a) More paths in grid A
XXXXXXXXX	XX	(b) More paths in grid B
XXXXXXXXX	XX	(c) about the same number of paths in each?
	XX	
	XX	
	XX	
	XX	
	XX	
	XX	

(Note: A 'path' was carefully defined for the subject as a sequence of line segments starting from the top row running down through each row to the bottom row, meeting one and only one symbol in each (horizontal) row of the array. Some subjects may have been misled on this problem by miscounting the number of rows in Grid B.)

TABLE 7.6

	Pretest Responses on A1			Posttest Responses on A1		
	Same	Grid A	Grid B	Same	Grid A	Grid B
Lecture	7	25	5	13	17	6
Experimental	8	28	3	22	12	1

According to Kahneman and Tversky [12] subjects favor grid A over grid B because there appear to be more paths 'available' in grid A. This was indeed the case on the pretest, as 53 out of the 80 subjects favored grid A. The reasons given for the choice of grid A included "there are more X's in grid A", and, "it is easier to draw a path in grid A than in grid B." There was still some tendency for subjects to favor grid A on the posttest, even after they had been exposed to counting principles. Reliance upon availability was less pronounced in the experimental groups on this item on the posttest than in the lecture groups. The extensive work in activity 4 in which the students in the experimental sections discovered their own counting principles may have contributed to the greater success of the experimental groups on this item.

A2: A man must select committees from a group of 10 people. Would there be:

(a) More distinct possible committees of 8
(b) More distinct possible committees of 2
(c) About the same number of committees of 8 as committees of 2?

Give a reason for your answer.

TABLE 7.7

	Pretest Responses on A2			Posttest Responses on A2		
	Same	Comm. of 2	Comm. of 8	Same	Comm. of 2	Comm. of 8
Lecture	9	25	4	11	16	12
Experimental	10	22	3	17	14	8

There was an overwhelming tendency on the pretest for the subjects to choose 'committees of 2.' These subjects used the availability heuristic because they felt that examples of committees of 2 were easier to construct than examples of committees of 8, and therefore the former must be more numerous than the latter. It is somewhat surprising that on the posttest more students did not get the item correct. Both the experimental activity-based and the lecture-based courses had covered such counting problems in detail. The subjects apparently did not see the complimentary nature of the problem, that a committee of 8 is equivalent to choosing a non-committee of 2. Perhaps retention was low because it had been four weeks since the students had worked any counting problems. The results on this item suggest that the availability heuristic is also a difficult crutch to overcome.

8. SUMMARY

The day-by-day observations made by the experimenter during one of the activity-based classes indicate that college students *can* learn to discover some elementary probability models and formulas for themselves while working on probability experiments in small groups. Furthermore, the effects of sample size upon measures of central tendency and variability can be learned by students working on activities such as those developed for the experimental course in this study. Making guesses for the probability of events and checking guesses with a hand-held calculator seems to help college students to be more cautious about probability estimates, and helps to make them aware about some of their own misconceptions about probability. Small-group problem solving, keeping a log of all class work and acitivities, and investigating the misuses of statistics all appeared to have a positive effect upon college students' attitudes towards mathematics, as indicated in the questionnaires filled out by the subjects in the experimental sections.

The present results of this study support the hypotheses of Kahneman and Tversky [10, 12] which claimed that combinatorially naive college students rely upon availability and representativeness to estimate the likelihood of events. Kahneman and Tversky were skeptical about the possibility of helping students to overcome their reliance upon availability and representativeness. The results on the posttest in this study suggest that the manner in which college students learn probability makes a difference in their ability to overcome misconceptions that arise from availability and representativeness. Mere exposure to probability concepts is not sufficient to overcome certain misconceptions of probability. Fischbein [9] notes that the "synthesis between the necessary and the possible – which is the basis of probabilistic thinking – does not in fact take place spontaneously . . . ." He claims that science education emphasizes only the deterministic aspect, and neglects the study of uncertainty. Thus peoples' intuition of probabilistic thinking is distorted by science education's emphasis on the necessary, and neglect of the possible. This experiment suggests that the course methodology and the teaching model used in an elementary probability course can help develop peoples' intuition for probabilistic thinking. A course in which students carry out experiments, work through activities to build their own probability models, and discover counting principles for themselves can help students to overcome their misconceptions about probability, and can help restore the synthesis between the necessary and the possible which is essential to probabilistic thinking.

Oregon State University
Corvallis

ACKNOWLEDGMENT

The author would like to thank the editor for his suggestions on revising this paper and for his correspondence concerning several technical aspects of the paper.

BIBLIOGRAPHY

[1] Tversky, A. and Kahneman, D., 'Belief in the Law of Small Numbers', *Psychological Bulletin* 76 (1971), 105–110.
[2] Shaughnessy, J. M., 'A Clinical Investigation of College Students' Reliance Upon the Heuristics of Availability and Representativeness in Estimating the Likelihood of Probabilistic Events', Unpublished Doctoral Dissertation, (1976).
[3] Cohen, J., 'Subjective Probability', *Scientific American* **197** (1957), 129–138.
[4] Cohen, J., *Chance, Skill, and Luck: The Psychology of Guessing and Gambling*, Baltimore, Penguin Books, 1960.
[5] Cohen, J. and Hansel, M., *Risk and Gambling*, New York, Philosophical Libraries Incorporated, 1956.
[6] Edwards, W., 'Conservatism in Human Information Processing', in B. Kleinmutz (ed.), *Formal Representations of Human Judgment*, New York, Wiley, 1968.
[7] Fischbein, E., Bărbat, I., and Mînzat, I., 'Intuitions primaires et intuitions secondaires dans l'initiation aux probabilités', *Educational Studies in Mathematics* 4 (1971), 264–280.
[8] Fischbein, E., *The Intuitive Sources of Probabilistic Thinking In Children*, Dordrect-Boston, D. Reidel, 1975.
[9] Fischbein, E., 'Probabilistic Thinking in Children and Adolescents', in R. Bechauf (ed.), *Forschung zum Prozess des Mathematiklernens*, Institut für Didaktik der Mathematik der Universität Bielefeld, 1976, 23–42.
[10] Kahneman, D. and Tversky, A., 'Subjective Probability: A Judgment of Representativeness', *Cognitive Psychology* **3** (1972), 430–454.
[11] Kahneman, D. and Tversky, A., 'On the Psychology of Prediction', *Psychological Review* **80** (1973), 237–251.
[12] Kahneman, D. and Tversky, A., 'Availability: A Heuristic for Judgeing Frequency and Probability', *Cognitive Psychology* **5** (1973), 207–232.
[13] Kahneman, D. and Tversky, A., 'Judgment Under Uncertainty: Heuristics and Biases' *Science* **185** (1974), 1124–1131.
[14] Carnap, R., 'What is Probability', *Scientific American* **189** (1953), 128–138.
[15] Mosteller, F., Kruskal, W. H., Link, R. F., Peiters, R. S., and Rising, G. R., *Statistics By Example*, Reading, Addison-Wesley, 1973.
[16] Mosteller, F., *Fifty Challenging Problems in Probability With Solutions*, Reading, Addison-Wesley, 1962.
[17] Huff, D., *How to Lie with Statistics*, London, W. W. Norton, 1954.
[18] Weiss, N. A., and Yoseloff, M. L., *Finite Mathematics*, New York, Worth, 1975.
[19] Pollak, H., 'On Some Problems of Teaching Applications of Mathematics', *Educational Studies in Mathematics* **1** (1968), 24–30.
[20] Thompson, M., 'Models, Problems and Applications of Mathematics', Unpublished Pre-conference paper for a Conference on Topical Resource Books for Mathematics Teachers, Eugene Oregon, 1974.

[21] Klamkin, M., 'On the Teaching of Mathematics so as to be Useful', *Educational Studies in Mathematics* 1 (1968), 126–160.
[22] Fitzgerald, W., 'The Role of Mathematics in a Comprehensive Problem Solving Curriculum in Secondary Schools', *School Science and Mathematics* (1975), 39–47.
[23] Freudenthal, H., 'Why to Teach Mathematics so as to be Useful', *Educational Studies in Mathematics* 1 (1968), 3–8.