4001

National Library of Canada | Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE | THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/NOM DE L'AUTEUR BARRY FREDERICK SQUIRE

TITLE OF THESIS/TITRE DE LA THÈSE PREFABRICATED CONCRETE IN GRADES ONE, TWO AND THREE

UNIVERSITY/UNIVERSITÉ UNIVERSITY OF ALBERTA

DEGREE FOR WHICH THESIS WAS PRESENTED/GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE MASTER OF EDUCATION

YEAR THIS DEGREE CONFERRED/ANNÉE D'OBTENTION DE CE GRADE 1975

NAME OF SUPERVISOR/NOM DU DIRECTEUR DE THÈSE DR W G CATHCART

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/DATÉ August 14, 1978 SIGNED/SIGNÉ B F Squire

PERMANENT ADDRESS/RÉSIDENCE FIXÉ 23 BOWMAN AVE,
ARMIDALE 2350
N.S.W AUSTRALIA

NL-91 (3-74)

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer; surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

THE UNIVERSITY OF ALBERTA

PROBABILITY CONCEPTS IN GRADES ONE, TWO, AND THREE

by

© BARRY FREDERICK SQUIRE

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF ELEMENTARY EDUCATION

EDMONTON, ALBERTA

FALL, 1978

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Probability Concepts in Grades One, Two, and Three submitted by Barry Frederick Squire in partial fulfilment of the requirements for the degree of Master of Education.

Supervisor

Date August 8, 1978

ABSTRACT

Research has shown that it is feasible to teach probability in upper elementary grades and that younger children appear to have some grasp of several basic ideas about probability prior to formal instruction.

The purposes of the present study were: (1) to determine the status of six basic probability concepts in grade one, two, and three children, (2) to investigate the level of quantification of probability present in these subjects, (3) to investigate differences in response due to the embodiment of probability settings, and (4) to examine the effect upon response due to the factors sex, grade, and IQ.

The six concepts studied were: events in a sample space, the most favorable event, the most favorable sample space for a given event, sample space equally favorable to a given set of events, impossible event, and certain event.

Seventy-two grade one, two, and three students from a suburban school were tested individually in an interview situation. Subjects made choice responses on concept items and predictions on quantification items.

Results showed that four of the concepts were understood by at least 75% of subjects in each grade and scores on all concept items were significantly greater than is attributable to chance. The scores on the quantification items were lower in all grades with an overall average of 42% correct on these items. For grades one and two the numbers of correct responses

were not significantly better than chance. There were no significant differences in scores due to different probability settings but performance decreased as the number of trials in the sample space increased.

Each concept question was presented in three embodiments, spinner, block, and box. No significant effect was found due to embodiment and there were no interactions between embodiment and sex or IQ.

On the total test score sex, grade, and IQ were all found to have significant effects. Grade and IQ were significant factors in the concept scores but no significant effects were found in the quantification scores. No interactions were found between any of the factors on the criterion measures.

Subjects were asked to state reasons for their responses in the probability test. Many correct rationalizations were given for concept responses but very few in relation to quantification predictions.

This study found that there was a substantial increase in understanding of probability concepts in children as they pass through grade three. At the same time there was little understanding of the numerical relationships between probability settings and frequencies of the outcomes, except when the number of outcomes is small.

Several implications are drawn for teachers of lower grades and for curriculum writers. It is suggested that provision be made for informal consolidation of existing concepts and ideas for grades one and two and a wider range of activities and

experiences from grade three onwards in which students are led to further concepts and into quantification of probability.

ACKNOWLEDGEMENTS

The writer is grateful for all the assistance and cooperation he received in conducting and reporting the study. In particular he wishes to express his sincere appreciation to the following:

Dr. W. G. Cathcart, his supervisor, for his help, encouragement, special interest, suggestions, and criticisms throughout the process of conducting the study and writing this report.

Dr. J. Kirkpatrick, a member of the thesis supervisory committee and earlier the writer's program advisor, for her continuing interest, and particularly her assistance in reading and commenting on the original draft of the report.

Dr. S. Hunka, an interdepartmental member of the thesis committee, for his interest in the study, for his advice in the analysis of the data, and for his consent to serve as a thesis committee member.

Dr. D. Sawada for his consent to serve as a thesis committee member and for his advice on various occasions throughout the study.

Gratitude is also expressed to the principal and teachers of the school in which the data were collected for their complete cooperation. Appreciation is also expressed to the Edmonton Public School Board for permission to conduct the study in the school.

A special word of thanks goes to the writer's wife, Jean, for her assistance, patience, and encouragement throughout the study and for typing the draft and final forms of the report.

B. F. S.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND STATEMENT OF THE PROBLEM

I. INTRODUCTION

One of the main tasks of curriculum design is the selection of appropriate content material. In recent years the criterion for judging topics to be appropriate for school age children has tended to be how well the study of such topics prepares those children for meeting life's situations. Most parents and educators have regarded a thorough grounding in the basic skills of language, expression, and arithmetic as providing this preparation yet it has been difficult to find agreement as to precisely what basic skills are necessary to cope with life.

Within the field of elementary school mathematics, there are many skills to be mastered and ideas to be understood, yet one topic has consistently been omitted from prescribed curricula even though suggested for inclusion by many writers in the past twenty years. This topic is the study of probability. The need for all children to have some experience in ideas of probability was well expressed by Cohen (1957) in a perceptive comment about the nature of schooling. Cohen said that

> Our system of education tends to give children the impression that every question has a single definite answer. This is unfortunate, because the problems they will encounter in later life will generally have an indefinite character. It seems important that during their years of schooling children should be trained to recognize degrees of uncertainty. (p. 137)

Restle (1961), Cohen (1964), and Estes (1976) reiterate how important a role subjective probability judgments play in our lives. We explain decisions made and conclusions reached on the basis of what likelihood we assign to events that are at best uncertain. Racha-Intra (1977) emphasized this function when he described the primary purpose of teaching probability as providing "a tool by which students comprehend the uncertainty model of the world" (p. 2). Yee (1966) called for the study of probability in elementary school on the grounds of the need to train children in decision-making skills.

The report of the Cambridge Conference on School Mathematics in 1963 recommended strongly that probability be a vital and appropriate part of the elementary school mathematics program. The National Advisory Committee on Mathematical Education (NACOME) (1975) similarly reported widespread support for the inclusion of probability as a necessary component of the elementary school curriculum. At the same time, the NACOME report cited a National Council of Teachers of Mathematics survey which found a deficiency in the training of teachers in probability and statistics resulting in a minimal treatment of probability topics by teachers in school. The instruction at school was found to be generally restricted to traditional graphing exercises and elementary descriptive statistics.

Although Fischbein (1976) observed "an increasing tendency to introduce the theory of probability in mathematical curricula" (p. 23), there has been a noticeable absence of this topic from

the official published curricula of most North American, British, and Australian elementary schools. At the same time there is no shortage of published programs and textual material on this topic designed for use in elementary schools. Examples are the School Mathematics Study Group series of four texts and commentaries, the Nuffield Mathematics Project Sequences on graphing and probability culminating in the "weaving guide" unit on probability and statistics, the Scottish Mathematics Project Series with relevant chapters for grades four through six, and sundry modern school texts which devote a chapter or two to activities involving chance and probability concepts.

The situation at present is that probability is included in some elementary school programs but not in others, the basis for inclusion often being the experience of the teacher. Three things need to happen before one can feel confident that all children will experience appropriate instruction in probabilistic ideas:

1. Starting points need to be determined by ascertaining what young children know about probability before receiving any formal instruction on the topic.

2. Curriculum materials then need to be prepared which begin at those starting points. Some of the already published programs and texts may be usable here.

3. Teachers and trainee teachers need inservice and pre-service courses to prepare them for teaching the material at appropriate levels.

It is the first of these tasks that this study was

designed to explore. The Cambridge Conference Report (1963) and Ausubel's (1968) theory of learning both had some influence on the design of the investigation.

The Cambridge Report called for as early a start as possible on the topic of probability; therefore grade one, two, and three children were chosen for this study. Most earlier studies involved pre-school and higher grade pupils with only one study investigating the first three grades.

It is assumed that meaningful learning is the desired product of classroom experiences. This, according to Ausubel (1968), occurs when two conditions are satisfied; the learner's anchoring ideas are ascertained, and the new material being presented is organized and modified accordingly. This agrees with Bruner's (1960) belief that subject matter can be modified and passed through stages of readiness just as children's thinking processes are. Ausubel goes further and says the subject matter must be modified and organized to match the child's readiness. The secret of teaching, he says, is to ascertain what the learner already knows and teach him accordingly. It seemed appropriate then to find the state of readiness of young children with respect to probability concepts prior to organizing learning experiences for them.

Reasons for the present study also came from the research literature. Many studies have been concerned with the level of understanding of chance at various ages and what conditions or environmental factors appeared to influence the judgments

made by children at these ages. The first conclusion to note is that "the global intuitions of relative frequency and probability are present even in pre-operational children" (Fischbein, 1976, p. 29). At this level, and for concrete operational children, many researchers have found that children have the ability to make only comparisons that are reducible to a binary operation. Not until the formal operational stage, according to Fischbein's interpretation of Piaget, can children make the synthesis between the possible and the deductive and understand quantification of probability as the relationship between the number of favorable and possible outcomes. This conclusion may have in fact contributed largely to the lack of research on probability at the elementary school level.

There have been those who claimed that preoperational children's comparison judgments have to be interpreted only as perceptual comparisons and not as probability judgments (Hoemann & Ross, 1971). Fischbein (1976) makes the point that probabilistic judgment is not completely reducible to quantitative estimation. There is a large element of understanding the meaning of the situation which implies an "intuition of chance" (p. 36). The existence of such an intuition appears to be indicated by Strohner and Nelson's (1974) finding that preoperational children are intuitively able to distinguish between probable and improbable events. In situations where there is a conflict between verbal meaning and factual probability the latter is stronger. For example, a 3-year-old child produces the situation "the girl feeds the baby" when asked to show "the baby

feeds the girl", as the former is seen to be more probable than what is requested.

Fischbein (1976) also holds the view that the intuitive background of probabilistic thinking desired in children needs to be built. If the spontaneous background of probabilistic thinking found in children is a mixture of correct and incorrect intuitions, as he suggests, then further research needs to be done to investigate the mixture at all ages up to adolescence so as to determine the foundation upon which curriculum building can occur.

Other research has found that the responses of young children tend to be egocentric, predictions are affected by an alternation tendency, reward is a significant factor at all ages, and mode of representation affects response. A number of investigations also found that pre-school children can "learn about probability" if the conditions tend to be largely nonverbal and reinforcing.

The present study sought to add to the current knowledge about early probabilistic thinking for the reasons outlined above. The study sought to embody several suggestions made by previous researchers as well as utilize features of past designs.

II. STATEMENT OF THE PROBLEM

The purposes of the present study were: (1) to determine the status of six basic probability concepts in grade one, two, and three children, (2) to investigate the level of quantification of probability achieved by these subjects, (3) to investigate

differences in response due to the embodiment of the probability setting, and (4) to examine the effect on performance on probability tasks due to the factors sex, grade, and IQ.

The six probability concepts in question were:

1. events in a sample space,
2. the most favorable event in a sample space,
3. the most favorable sample space for a given event,
4. sample space equally favorable to a given set of events,
5. impossible event, and
6. certain event.

These are defined in a later section of this chapter.

With respect to the second purpose there is some dispute among researchers as to when children begin to attend to the quantitative aspects of a probability setting. Fischbein, Pampu, and Manzat (1970) and Chapman (1975) found young children able to correctly compare ratios of the form a/b:c/b where the comparison could be reduced to a binary operation, a:c in this case. Fischbein et al. also found that 9-year-olds quite readily made correct probability judgments on the basis of relative frequencies of preferred events. On the other hand, Piaget (1975) maintained that proportional facility in probability situations only comes with the formal operations stage of development. It was hoped that this study would add to the evidence on young children's ability with quantification of probability.

The third purpose, the effect of embodiment on pupil response, was investigated partly to check Jones' (1975) findings and

partly to provide further guidance for the design of activities and other curriculum material.

Wilkinson and Nelson (1966) found that responses were affected by the familiarity of the situation to the subject. The devices and situations used in the present study were designed with this finding in mind. Every attempt was made to make all embodiments a new experience for the subjects.

The consideration of individual differences is vital for effective teaching. Sex, grade, and IQ are three easily discernible differences among pupils, thus the effect of these factors on response to probability questions was examined as the fourth purpose of this study. No attempt was made to test for the effect of the many other social and personal characteristics such as socio-economic status, cognitive style, conservation ability, vocabulary, and listening ability. Some of these may have contributed to error in the criterion measure and would thus impose a limitation on the study. Many of them would be worth examining in a further investigation.

III. MAJOR QUESTIONS AND HYPOTHESES

The four purposes stated above gave rise to the following questions and null hypotheses.

Purpose One

1. What proportion of subjects in each grade and in the total sample indicate an understanding of the six concepts investigated? These relative frequencies are taken to be

measures of the status of the concepts.

2. Null Hypothesis: The proportions derived in answering question one are not significantly different from chance proportions for each concept.

Purpose Two

3. What proportion of subjects in each grade and in the total sample indicate an understanding of the quantification items presented?

4. Null Hypothesis: The proportions derived in question three do not differ significantly from chance proportions.

Purpose Three

Null Hypotheses.

5. There are no significant main effects due to embodiment on performance on the probability test.

6. There are no significant main effects due to embodiment on performance on the probability test when sex, grade, and IQ are used as blocking variables in pairs.

7. There are no significant interactions between the embodiment and the factors sex, grade, and IQ.

Purpose Four

Null Hypotheses.

8. There is no significant main effect on probability test performance due to (a) sex, (b) grade, and (c) IQ.

9. There are no significant interactions between the independent variables sex, grade, and IQ on the criterion measures.

IV. DEFINITION OF TERMS

For the purposes of this study and report, the following terms are used as defined.

Probability Concepts

Probability concepts are taken to be the notions or ideas about random phenomena such as the outcome of rolling a die or whirling a spinner. The concern is more with "what will happen?" rather than "how often?".

Quantification of Probability

Quantification of the probability of an event is the assigning of a relative frequency to that event. This is a predicted measure of "how often" that event can be expected to occur. Due to the age of the subjects in the study proportions and fractions were avoided and expressions such as "four out of six" were used.

Embodiment of Probability

Situations were presented in three modes or embodiments: a spinner, a block, and a box containing counters of various colors. All had six "outcomes" to enable the construction of equivalent probability settings.

Random Devices

The devices referred to occasionally are the spinners, blocks, and boxes arranged as outlined in chapter 3 under Materials.

Probability Setting: a-b-c

Throughout the investigation, five different trinomial proportions were represented by the devices. These are referred to as settings. For convenience of labelling, the proportion a:b:c is rewritten a-b-c and used to identify which device is being used. For example, the 2-3-1 spinner means the spinner with two BLUE, three RED, and one YELLOW segment. This color order was invariant throughout the experiment.

Events in a Sample Space

A listing of all possible outcomes in a given situation constitutes the events in a sample space.

The Most Favorable Event in a Sample Space

This is the event with the greatest expected chance or relative frequency of occurring. For example, in using the 2-3-1 spinner RED would be the most favorable event.

The Most Favorable Sample Space for a Given Event

This is the sample space in which the given event has the greatest chance of occurring. For example, given the 2-2-2, 3-2-1, and 4-1-1 blocks the most favorable sample space for BLUE to occur is the 4-1-1 setting.

Equally Favorable Sample Space

This is the setting in which all events have the same chance of happening. In the previous example the 2-2-2 setting gives all events (colors) an equal chance of occurring on any one roll of the block.

Impossible Event

This is a phenomenon which is not in the sample space and hence cannot occur. For example, with the 3-3-0 spinner YELLOW would be an impossible event.

Certain Event

This is a phenomenon which is the only element in the sample space being considered and hence must occur every time. For example, with the 0-0-6 box YELLOW would be a certain event on each draw as the box contains only yellow counters.

Understanding

Understanding of a concept or a principle is indicated by the average per cent of correct responses on items relating to that concept or principle.

V. DELIMITATIONS AND LIMITATIONS

The sample for the study was selected from grade one, two, and three children at the one school in Edmonton made available by the Edmonton Public School Board. The subjects were selected on the basis of teacher perception of general reasoning ability within grades and sexes. The opportunity did not exist for a more random sample from a larger population.

The delimitations stated above impose some limitations on the generalizability of the results. The grade three classes tended to have very few low-IQ girls from which to select, and this would not have been representative of the wider population. Teacher perception is a subjective judgment, but in this case

a correlation of 0.778 with IQ scores was found across the sample when IQ scores were later available for grade one and two subjects.

No attempt has been made to control for socio-economic status, although an average rating was calculated for the overall sample. Rural students may also have different academic and social experiences from urban students and may respond differently to the test items and indicate a different level of understanding of probability concepts. It will be left to further research to determine if the differences mentioned above do exist.

VI. ASSUMPTIONS

One major assumption of the present study was that subjects' choice responses given in the test provided a true picture of the level of understanding of the related probability concepts. This was an assumption about both the validity of the instrument and about the child's understanding of the questions asked. A panel of experts judged the instrument as being a valid test of the basic concepts in question. The interview protocol was designed to ensure that subjects understood the questions asked.

A related assumption that is unavoidable in all forms of verbal testing is that the concepts and explanations expressed were the concepts and reasons believed. The researcher was successful in winning the confidence of the children so that most of them appeared to be relaxed enough to respond freely and openly. There is always that uncertainty, however, that accompanies one's judgment in such situations.

It was assumed that no subjects had received any formal

instruction in probability concepts. The school program contained no such activities up to the grade four level. It could not be assumed, on the other hand, that pupils had had no experience with notions of chance, as most acknowledged that they had played games of chance involving dice and/or spinners. This would be difficult to control for.

It was also assumed that no learning took place from one item on the test to the next. Students were not told whether they were correct or not, but were rather given a non-committal "O.K.", "Uh-uh", or a repeat of their answer with an inflexion. In addition, possible teaching effect was controlled for by randomizing the order of presentation of the items as described in detail in chapter 3. If learning did take place on an individual level it would then be uniformly distributed among the items. The interviews took only about twenty minutes each.

Finally, it was assumed that the IQ ratings were reliable measures of comparative cognitive abilities within the sample. The test writers advise that the ratings only be used on a comparative basis and within three months of the test being administered. Both these conditions were satisfied.

VII. SIGNIFICANCE OF THE STUDY

One of the points made in the introduction to this report was that educators' long-standing recommendations for early study of probability at school have not been implemented. Very few curricula include the study of probability as part of a core of study for all elementary school students.

Following the Cambridge Report's recommendation that the study of probability begin as early as possible in the elementary school, several research studies have been concerned with the design and teaching of instructional units in probability. Wilkinson and Nelson (1966) found that grade six children responded well to a specially designed unit but most subjects had strong preconceived intuitions about situations familiar to them from earlier experience. Many intuitions that were incorrect proved difficult to alter. Ojemann, Maxey, and Snider (1965, 1966) found that preliminary instruction of grade three children has a positive effect on probabilistic behavior. Fischbein, Pampu, and Manzat (1970) demonstrated that even 10-year-old children can rapidly understand the concepts of arrangements and permutations and solve problems involving simple combinatorial calculations. Shepler (1969) found it was possible to teach introductory probability and statistics to sixth-grade students to a high degree of mastery but he cautions against the use of activities and materials that call for too subtle an interpretation.

All the researchers indicate a belief that probability should be taught in elementary schools and call for further investigation into what should be taught, when experiences in probability should begin, in what sequence, and with what rigor. As indicated in the introduction, this study is concerned with finding anchorage points for the development of sequences of probability activities for young children. If meaningful learning experiences can then be organized into the curricula of the lower elementary grades, we may be able to prevent many

incorrect intuitions of chance developing and lay a more solid foundation for later studies of probability and statistics.

VIII. ORGANIZATION OF THE REST OF THE REPORT

Chapter 2 contains a review of the literature relevant to the development of notions of chance in children. Chapter 3 presents a description of the probability test and its design, the IQ test, research procedures, and analysis of the probability instrument. Chapter 4 presents the results of the data analysis and chapter 5 gives a summary and discussion of the main findings.

CHAPTER 2

REVIEW OF RELATED LITERATURE

Much of the related research has been concerned with the feasibility of teaching concepts of probability in the upper elementary grades or later. This may well have been influenced by Piaget's statement that the quantification of probability demands the onset of formal operational thought. Whatever the reason, we know very little about the actual readiness of young children for this topic.

The research that has been done can be divided into three main groups:

1. Piaget and Inhelder's study of the child's understanding of chance (1975), and studies directly related to their resulting theory of development of the idea of chance,

2. studies concerned with the status of probability concepts and significant factors affecting their development, and

3. studies concerned primarily with the preparation and trial of instructional units in probability and their implementation with, and effects upon, elementary school children.

There is unavoidable overlap between these categories as some studies which began with Piaget's theory also examined factors affecting the development of chance concepts. Likewise, others could have been reported in either group 2 or group 3. As near as is practical, the studies are reported in chronological order within each section.

I. PIAGET AND RELATED STUDIES

The Theory

Piaget's (1960) genetic epistemological theory of development is well known among researchers for the way it presents intellectual development as the organization of intellectual operations into structured systems, progressively producing cognition of greater genetic maturity. Logical and arithmetical operations are seen as internalized actions organized into systems which are basically characterized by rigorous composition and reversibility. Such reversible composition makes possible deduction and lawful predictions. However, chance transformations and chance events are not rigorously reversible and cannot be rigorously composed in deductive systems.

Piaget (1975) applied his general theory to the development of the concept of chance in children and proposed the following:

1. The preoperational child is influenced more by contiguity in space and time than by causality, and he is unable to distinguish possibility from necessity. Fischbein (1976) elaborates on Piaget's meaning: "Being unable to understand necessity, he is unable to understand-by-opposition the unnecessary, the fortuitous" (p. 27). In the words of Flavell (1963), "for the preoperational child, nothing is deductively certain and nothing is genuinely fortuitous . . . ; his thought is forever at midstation between these poles" (p. 342).

2. During the concrete operational stage (approximately age 7 through age 10) the child begins to separate the necessary

from that which is simply possible; he becomes able to distinguish chance events from the strictly deductive ones. This is not enough to produce probabilistic thinking in its full sense because the child only understands the possibility of operational systems related to chance events as he discovers them in an incomplete and empirical fashion.

3. Not until the formal operational stage (about age 12) can the probability concept begin to be completely mastered. Ability with combinatoric operations and proportions at this stage allows for the synthesis between the possible and the deductive which is the source of probabilistic thinking.

The Experiments

Except for a brief article in 1950, Piaget's only work on chance concepts in children is the book he and Inhelder wrote in 1951 which was not translated into English until 1975. The conclusions given above arose from a dozen or so experiments reported in the book. There were three main experiments in Part I and three more of relevance in Part II which are summarized below.

Part I. Chance in Physical Reality

1. Random mixture and irreversibility.

To test for notions of random mixture and irreversibility a rectangular tray was set up as a see-saw in which 2 groups of balls, 8 red and 8 white, separated by a divider, were arranged along its width. At each see-saw movement, the balls would

roll to the opposite end then return to the original end when the box is tipped back. The balls invariably collided with each other and returned into a different arrangement. Prior to each tipping, the subjects were asked to predict the arrangement of the balls when they return to the starting end. Questions were used such as: Will the red ones stay on one side and white ones on the other? Or will they mix up? In what proportions?

After several successive predictions and tippings, subjects were asked to predict the result of a large number of moves of the tray. Of particular interest was any prediction of a progressively random mixture, or of a final reordering to their original sides, the two extremes which Piaget encountered.

2. Centred and uniform distributions.

(a) To test for notions about centred distributions, five funnelled boxes were shown to the subjects one at a time. The first four were funnelled in the centre of the top, the fifth at the right side of the top. Boxes #1 and #5 had 2 bins at the bottom, #2 had 3 bins, #3 had 4 bins, and #4 had 18 bins below a network of nails as in a regular quincunx.

As each box was used, several marbles were dropped in the funnel one at a time with the subject predicting then explaining the outcome. Finally, about sixty were let go with the subject first giving a prediction then an explanation.

(b) For uniform distributions, small glass beads which did not roll too easily were dropped from a kind of trellis

sieve onto a sheet of grid paper and the subjects were asked about the chance of uniform distribution as a function of the greater or smaller number of beads dropped. (p. 27)

3. Constant relationship in conflict with fortuitous uniform distribution.

To find out how the mind of the child succeeds in dissociating what is due to chance from what is due to non-chance, a more "flamboyant" (Flavell, 1963) experiment was used. The apparatus was essentially a roulette wheel with an iron bar as a pointer. There were 16 equal sectors of 8 colors, opposite sectors being of like color, and the wheel spun "honestly" until a set of matchboxes was placed on its colors. The matchboxes looked identical but all contained wax in which were embedded pieces of metal of various kinds to make four sets of four boxes of different mass. Two boxes in one set contained magnets so that the spinner could be made to stop on a chosen color.

The initial inquiry was about the honest wheel, that is, about where the pointer would be likely to stop on a given spin, about the distribution of stops over a large number of spins, and so on. Once the child had noted the fortuitous distribution of stopping places, the matchboxes were introduced to give an unexpected constant element designed to throw off the predictions of the child. This assumed he saw the dispersion of the stopping points as uncertain. His reactions to and explanations for this new situation were recorded.

Results of Part I Experiments

The analysis of each experiment contains extensive examples from the subjects' protocols followed by the experimenters' explanations and hypotheses. These examples of children's reasoning make this book one of the easier and more interesting of Piaget's to read. In each of the three studies, the preoperational children made the more interesting responses. They tended to impute a hidden lawfulness to the randomization process in all three experiments (e.g., "blue this time as it was red last time"), and held quasi-magical views of causality on occasion (e.g., "it would go to green if you concentrated hard enough"). Flavell (1963) made a delightful footnote to gambling-oriented readers on this point referring to preoperational tendencies! (p. 344)

Responses of older subjects indicated a gradual ontogenetic development of notions of irreversibility with respect to the order of the marbles in the see-saw tray. Likewise, the ability to predict distributional form was slow in developing. This is in keeping with Piaget's view that a grasp of "the law of large numbers" depends on an understanding of proportions, something not acquired in force until the formal operational period.

With regard to the wheel experiment, younger subjects inferred more predictability than was justified and were not surprised by the results of the dishonest spins. They considered them

> not beyond the pale of the hodge-podge of quasi-magical causal relations already thought to be at work in the genuinely random turns. (Flavell 1963, p. 344)

Older children, on the other hand, sensed a trick and went on to discover the cause.

Part II. Random Drawings: The Experiments and Their Results

1. Chance and "miracle".

The first experiment in this section is similar in one aspect to the third experiment in part I in that it allowed the experimenter to "cheat chance", to produce "the miracle" at will. Two sets of about 15 counters were used; each one in the first set bearing a circle on one side and a cross on the other, each in the second having crosses on both sides. Showing the subject the first set, the experimenter asked for a prediction for a single throw, then for the distribution of crosses and circles if all were thrown at once. Then the experimenter surreptitiously substituted the second false set, threw them all at once, and gauged the child's reactions.

The younger children again merely registered mild surprise whereas older ones suspected a trick and quickly turned a few counters over to confirm their suspicions. The startling thing is that even after the trick was revealed, the younger subjects were inclined to think that the same result could readily be reproduced using the true counters.

This experiment was repeated in a slightly different form using two sacks of marbles. The first had red and blue marbles in it, the second had only blue marbles. Again the trick was employed, reactions analysed, and the trick explained.

At the end of each test, without the subject knowing

whether the mixed or homogeneous set was being used, the experimenter threw one by one the trick counters, or drew out the marbles one by one from the trick sack (blue). The aim was to catch the moment the subject recognizes for sure that it is an experiment with the homogeneous elements, and to explore his reasoning. According to Piaget, "this last experiment often gives the surest index of the judgment of probabilities of which the child is capable" (1975, p. 97).

In both experiments the reasons of the younger subjects appeared to be egocentric and phenomenological and Piaget concluded that these children understood nothing about the notion of random mixture.

2. Random drawing of pairs.

The first of the experiments having a bearing on quantification involved four unequal sets of different colored counters which were mixed thoroughly in a bag (e.g., 15 yellow, 10 red, 7 green, and 3 blue). Identical sets of each color were left on the table as a memory aid. The child made successive drawings of pairs of the counters from the bag and was asked to predict the most probable pair before each drawing. No replacements were made and each pair drawn was placed on the table in front of the subject as he drew them. It was possible for him to know what remained in the bag, but no explanations of this were given.

The reported results again matched Piaget's earlier conclusions. Younger children predicted according to a variety of bases, among them color preference and the imputed lawfullness

of "taking turns". Concrete-operational children tended to base their forecasts on relative frequencies but to forget that each drawing changed the frequencies, and so failed to keep their estimates up to date as they continued to draw each pair. The oldest children tended to keep a running tally of the changing distribution and quantified the probability as a function of the counters left in the sack.

3. Quantification of probabilities.

In the final study summarized here, counters were used, some with a cross on the back, the others with no cross. The experimenter made up two collections of counters and showed them to the subject. All collections had a small number of counters, for example, 2 with crosses, 2 without, and 1 with and 2 without. When the subject had noted the makeup of each set, they were separately mixed up and placed blank face up on the table. The task was to judge whether there was a greater chance of drawing a "cross counter" from one set than from the other. A number of such problems were posed, some very simple (e.g., comparing a 2-2 set with a 0-4 set) and others more difficult (e.g., a 1-2 set and a 2-3 set).

The same three developmental stages appeared in the responses, according to Piaget. In the four to seven year group, there was the absence of any comparisons based on the proportions in play. Occasionally there was an intuitive comparison when striking disproportions were perceived. For those in the second stage, seven to eleven years, there was a beginning of quantification but with a consistent error: the prediction was solely on the

absolute number of counters with crosses in each set rather than in terms of the ratio of these to total counters in each set. The subjects compared favorable with unfavorable cases but did not construct a relationship between the favorable and the possible. This relationship appeared to be established in all of the eleven-years-and-over subjects: they tended to solve each problem by a calculation with fractions.

Other experiments were reported by Piaget and Inhelder where they investigated even further the quantification of probabilities and combinatoric operations, but the six outlined above are sufficient background for this review which is concerned more with Piaget's stage I and II subjects.

Criticisms and Related Studies

Very few details were given by Piaget about the size and nature of his samples, and this is one of the grounds on which he is criticized. It was noted by Flavell (1962) that Piaget's probability experiments also called for a high level of verbalization of the child's understanding of various technical aspects of mathematical probability. All we are told about the samples is that the age range was from 3 to 12 years and samples of 7 to 14 subjects were used. No attempt was reported to control for effects of age, sex, or other relevant variables, and no provision was made for analysing the results statistically.

The main area of disagreement with Piaget has been with his conclusion that children up to the age of about 7 show no evidence of being able to think probabilistically. Contrary

evidence began to come from the work of Messick and Solley (1957), Stevenson and Zigler (1958), Stevenson and Weir (1959), and Siegel and Andrews (1962) who examined reinforcement conditions and found young children tended to adopt maximizing strategies in probabilistic situations.

Yost, Siegel and Andrews (1962) listed several shortcomings of Piaget's technique, designed a model to overcome these, and compared it with a Piaget-type procedure. Their desire to minimize the verbal content of the procedure led them to a decision-making technique involving a choice between two transparent boxes of counters instead of a prediction with one box and a duplicate set of tokens displayed in a fixed, initial manner. They also sought to control for color preference and provided reinforcement other than knowledge of the outcome. The scores under the two conditions were different enough for Yost et al. (1962) to conclude that

> Probability judgments made by 4- and 5-year-old children have been observed in two situations. In the situation in which controls are introduced, amount of reinforcement is increased, and an opportunity for non-verbal decision-making is presented, children tend to make correct responses significantly more frequently (p. 780).

Goldberg (1966) replicated this experiment, with some modifications, endorsed the above conclusions, and emphasized the importance of task conditions in observing the level of performance of four- and five-year-olds on probabilistic judgment tasks.

The Yost and Goldberg studies involved only 19 and 32 subjects respectively and both researchers called for further

investigation. Davies (1965) designed and administered non-verbal and verbal tests on probability concepts to 112 subjects, 8 male and 8 female at each of ages three through nine years. Her results supported Piaget's interpretation of the acquisition of this concept as a developmental phenomenon but there appeared to be evidence that preoperational children frequently behave according to event probabilities even before they can adequately verbalize the probability concept by which they are responding. No significant sex difference was found in the development of this concept and Davies concluded with a suggestion that age of appearance of non-verbal and verbal demonstrations of probability concepts be related in future studies to other variables such as MA or IQ, socio-economic status and educational level of the home, and patterns of child-rearing.

Offenbach (1964, 1965) also found evidence for probabilistic thinking in kindergarten children and this contention seems to be refuted only by Hoemann and Ross (1971). They argued that preschool children's probability judgments are really only perceptual comparisons of two arrays of objects and not judgments about likely or unlikely outcomes at all.

Fischbein (1976) disagreed with Hoemann and Ross and contended that understanding the meaning of the situation in which they placed their subjects implies necessarily the intuition of chance and that this intuition and the intuitive estimation of probabilities appear early in childhood. This would seem to be the state of this argument at the present time,

but many more aspects are considered in the next two sections of this review.

II. THE STATUS OF THE CONCEPTS AND SIGNIFICANT FACTORS

Some research has studied the type and level of chance ideas which are acquired naturally, that is, without formal instruction. This is the first concern of this section of the review.

Doherty (1966) examined the status of four concepts of probability in a sample of 54 grade four, five, and six children. She concluded that the subjects had acquired naturally, from everyday experiences, considerable familiarity and ability with the four concepts: (1) the idea of a sample space, (2) the probability of a simple event in a sample space, (3) the probability of the union of non-overlapping events in the sample space, and (4) the idea of the difference between mutually independent events and mutually exclusive events. No significant difference was found in level of difficulty between concepts nor were sex or chronological age significant factors for this sample. Significant differences were found between ability levels, mental age levels, and mathematical and average achievement levels.

Leffin (1969) surveyed 528 children randomly selected from the grade four through seven population of the Wausau Public School System, Wisconsin. The sample was categorized on the basis of sex and three IQ ranges. Three concepts were examined: (1) points in a finite space, (2) probability of a simple event in

a finite sample space, and (3) the quantification of probability. (The labelling of this last item as a concept is questionable.) Three tests were administered to groups of subjects, one test for each concept. The overall mean performances were significantly different among IQ groups, sex, and grades. The most significant outcome reported was that the children demonstrated that they had acquired considerable knowledge about the three concepts in question, and that such knowledge must have developed as a result of their background, experience, and intuition. Both Doherty and Leffin called for an inclusion of probability topics in the elementary curriculum for intermediate grades, and for the development of methods and materials appropriate for the task.

Jones (1975) appears to be the only researcher who looks specifically at grades one, two, and three with respect to performance on concepts of probability. The five concepts were: (1) outcomes of a sample space, (2) most favorable event in a sample space, (3) most favorable sample space for a given event (same number of outcomes in each space), (4) sample space equally favorable to a given set of events, and (5) the most favorable sample space for a given event (same number of favorable outcomes in each space). A sample of 162 subjects was chosen on the basis of the grades and three IQ levels. Three embodiments were used: spinners with equal sectors comprised the first embodiment (unit), spinners with unequal sectors comprised the second (gross), and containers of discrete objects comprised the third (set). Five tests were administered

individually by the investigator in an interview situation. Corresponding test items for the three embodiments were isomorphic in a probability sense. Also half the items were contiguous and half the items had non-contiguous outcomes. Items from both sets were matched in a probability sense.

It was found that grade two and three children performed significantly better than grade one children on concepts one, four, and five, but there were no significant differences between grades two and three on any of the tests. Several differences were reported regarding the embodiment and contiguity factors. This led the investigator to conclude that some topics in probability should be introduced into the primary curriculum but the concepts which are included should be presented in as many different settings as possible.

The second concern of this section of the review is with studies which have primarily focused on factors which affect children's response in probabilistic situations.

Offenbach (1964) systematically studied the effect of reward and punishment on the learning behavior of 30 kindergarten and 30 grade four children. Using marbles and ten-cent toys as the prizes for correct guessing of the next card in a specially made deck, he found that the reward-punishment groups of subjects chose the more frequent event more often than the control group. The level of reward-punishment did not appear significant, in agreement with the findings of Brackhill, Kappy, and Starr (1962). The absence of probability matching behavior in the control group was at variance with

results reported by Siegel and Andrews (1962). Offenbach suggested the inconsistencies could be attributable to methodological differences, simultaneous presentation of stimulus in pairs by Siegel versus successive presentation by Offenbach. In a subsequent study Offenbach (1965) found little effect due to the method of stimulus presentation except that the simultaneous procedure made it easier for both age groups (K and 4) to respond on the basis of previous outcomes. In both studies Offenbach found a tendency for the fourth graders to try to find rules governing the occurrence of the events while the kindergarten children responded to the immediate situation in isolation. The older children appeared to be more aware of a possible sequential nature of the task. These and other intratask behaviors reveal age differences consistent with Piaget's stages of logical thinking but take issue with Piaget's belief that probabilistic thinking doesn't occur before age seven.

Mullenex (1968) group-tested a class in each of third, fourth, fifth, and sixth grades to determine the level of understanding of four probability concepts and to study the relationship of this understanding to the variables of sex, age, general ability, and basic skill in school subjects. None of the variables appeared to be relevant predictors of the criterion measure, but judgment was reserved in regard to sex, basic reading skill, arithmetic skill, and problem solving skill as measured by the Iowa Tests of Basic Skills. The investigator recommended individual testing and instruction

to more precisely determine the relationship between these variables.

Carlson (1969) used several Piagetian-type tests of varying difficulties to examine 160 eight- to eleven-year-old children for development of probabilistic thinking. Apart from supporting Piaget's conception of the development of probability concepts in children at this level, Carlson noted that age and socio-economic status were the most significant factors, level of general intelligence was less important, and no differences in performance were attributable to the subject's sex.

Fischbein, Pampu and Manzat (1970) found that age and instructional conditions were highly significant factors in children's responses in a ratio-comparison experiment related to concepts of chance. Sixty children from each of grades K, three, and six* were seen individually and presented with 18 problems. Twenty subjects at each grade level were assigned to one of these instructional conditions each of which began with a guessing game concerned with black and white marbles. The interesting conclusion drawn by the investigators is that a little amount of instruction enabled the nine-year-old subjects to correctly estimate chance by comparing ratios whereas prior to the brief instruction the nine-year-olds' spontaneous responses differed little from those of the five-year-olds. This finding led the investigators to question Paiget's hypothesis about the proportionality concept only

* American schools equivalent would be grades 1, 4, and 7.

coming at the formal operational stage and to argue that probability topics should be started in primary school.

Hoemann and Ross (1971) conducted four experiments on probability judgment tasks with children ranging in age from preschool to early adolescence. As mentionsd in part I, their point of view was that the preoperational child made supposed probability choices only on the basis of magnitude discrimination with no probability inference being involved. They acknowledged that there was a contrary point of view but argued that their results supported Piaget and Inhelder's account of the development of probability concepts in children.

One particular result from Hoemann and Ross's study of relevance to the present study was that

> preschool children of CA 4½ performed at the 75% level on the two-array task that allowed a direct comparison, while this level was not reached until CA 6 in a single-array task. (p. 235)

The single array task called for a prediction which Piaget had suggested was more difficult than a comparison choice, although his experiments required mostly prediction responses. Yost, Siegel, and Andrews (1962) and Goldberg (1966) had utilized choice between arrays instead of prediction with one array in their decision-making models and found improved response as a result.

Hoemann and Ross found it unnecessary to employ the elaborate controls for color preference and odds displays that Yost et al. and Goldberg had used. Their use of black and white is an indication of this but they also state the fact

explicitly. While acknowledging that preschool children do indeed sometimes prefer color to quantity, they affirm that no probability inference is involved in either case.

Finally in this section we consider a study on the role of verbalization in probability learning. Stevenson and Weir (1963) found that subjects responding in pairs did not perform differently from a single subject. Subjects who were forced to verbalize the basis of their responses did not differ in their choice behavior from those who did not verbalize explicitly. The sample in their experiment was comprised of 78 twelve-, fifteen-, and eighteen-year-olds and it is not certain that these findings can be assumed applicable at a younger age.

III. CLASSROOM STUDIES OF MATERIALS AND INSTRUCTIONAL METHODS

Many studies have investigated how the ability to think in probability terms can be developed most effectively. Ojemann, Maxey and Snider (1965, 1966) devised a program of guided experiences designed to help the child learn elementary aspects of probability. A series of five 30-minute lessons were administered on consecutive days to a third grade class of 20 pupils with a corresponding class of 21 as the control group. One of the main concerns in designing materials and experiences for the lessons was that responses be encouraged which relate to the information available rather than only to previous experiences of interaction with the environment. Four tests were used to assess the effect of the instruction. Taken

together the results indicated that the experimental subjects were acquiring considerable ability to relate their "predictions" to the information available.

> They showed significantly greater ability to relate their predictions to the probable and they tended to wait before making a prediction when only a small amount of information was available and more would be supplied. (p. 326)

The investigation seemed to indicate that grade three children can benefit from instruction in concepts of risk, chance maximization, and prediction.

At about the same time, and also in Iowa, Wilkinson and Nelson (1966) conducted a three week trial-teaching experiment with a sixth-grade class in the area of probability. The study sought answers to questions about content suitability and organization of activities. The class was taught for 45 minutes each day and this sequence appeared to be the first formal experience with probability encountered by any of the subjects. The most basic consideration was that experiences and ideas should be meaningful to the students. The lessons began at low levels of sophistication and activities were carried as far as possible while class interest was maintained.

The general technique was to initiate discussion with carefully stated questions and to encourage students to develop ideas through actual experimentation. Exploratory questions such as "Why?", "What does 'usually' mean?", "Are you sure?", or "How can you find out?" were employed and instructor responses were mainly neutral. Experience with eighteen concepts and five skills was incorporated into the

sequence of lessons which employed three probability situations: personalistic, familiar, and unfamiliar. The first involved data from the student's own environment such as birthdays, telephone numbers, and personal statistics. The second situation used coins, dice, and cards and the third dealt with objects such as thumb tacks, bent paper clips, paper cups, cardboard cylinders, and drawing beads from a container.

Approximately equal time was spent on each situation and questions were asked about likelihood, fairness, number of possible outcomes and combinations, and certainty. No concerted effort was made for precision in defining concepts; rather they were mostly dealt with on an intuitive level only.

In many of the experiences subjects realized the uncertainty of the situation and the need for a consensus of methods as indicated by a comment: "We've just got to make some agreements and then we can get somewhere" (p. 102). In dealing with familiar situations, the investigators were surprised at the number of inaccurate beliefs and intuitions held by the subjects concerning coins in particular.

> When testing quarters to see if engraving differences cause tails to occur more often than heads, students were willing to make generalizations from ten flips if the results agreed with their prejudice. They said that six tails and four heads in ten trials would prove their point , . . More than six tails . . . was especially satisfying to them. Five tails . . . was treated as something you have to expect once in a while, but which doesn't disprove anything . . . Fewer than four tails in ten trials . . . led some to say that something had been dishonest (pp 102-103).

Less prejudice was found with dice and none with cards except that a few students were aware of the existence of trick

decks. No rationale was given by students for their objection to the use of a large die and a small die together but it seemed "unfair" to them. No such prejudice seemed to be held with regard to the unfamiliar situations. The students were encouraged in hypothesis-making and -testing activities which all agreed were interesting and worthwhile.

> The important aspect of this experiment is the progression from a beginning estimate of a probability without necessarily being "exact" at any time (p. 105).

This would seem to be the essence of practical probability experiences: many questions in everyday life have answers of an indefinite character and a good guess is often the best answer possible.

Wilkinson and Nelson concluded with six recommendations that have implications for those involved in designing probability units for elementary school children. In brief summary these are:

1. Don't let intuition lead you too far.
2. Avoid pre-prejudiced situations.
3. Keep vocabulary useful and simple.
4. Spread the teaching out. Two- or three-day units spaced throughout the year would be more suitable than a three-week unit.
5. Don't overstructure the experiences.
6. Use experiences meaningful to students, experiences planned especially for them, not watered-down versions of high school or college-level probability.

Another significant study of teaching probability to sixth-

grade children was reported by Shepler (1969). He designed a four-week unit which was taught to a class of 25 students of average ability by a trained elementary school teacher. The instructional goal was to demonstrate "mastery learning" of the behavioral objectives of the unit.

Pre- and posttests were administered, the first to measure preknowledge of 14 objectives, the second to help both in re-analyzing the unit regarding modifications to be made and in testing the feasibility of the study. The test consisted of 72 items based on one- and two-dimensional finite sample spaces generated by models using coins, dice, spinners, and boxes of objects.

The criterion for instructional success was that 90% of the students should score 90% or better on each of the measured objectives. This was satisfied in the posttest in the case of 11 of the 14 objectives. There was a dramatic change in the performances on pre- and posttests with the mean scores increasing from 38% to 93%, and the variances decreasing from 74 to 11. The instruction was judged highly successful and the author attributed the large gain in raw score to the developmental analysis used and the mastery learning techniques employed.

The author proposed the following sequence for developing research-based curriculum materials:

> Start with a content outline and establish behavioral objectives. Task analyse these objectives and write an instructional treatment to meet them. Proceed to the important step of actually trying these materials with children, while recognizing the possibility of iteration through preceding steps (pp. 202-203).

Shepler followed his own advice in analyzing the items which measured the three missed objectives. He concluded that poor wording in some items and lack of emphasis of objectives in others were plausible reasons for not achieving the objectives.

Lovell (1971), in commenting on the above study, stated "there is no doubt . . . that, given first-class teaching, selected sixth-grade pupils can be introduced to notions of probability" (p. 135). Despite its undoubted success, Shepler himself admitted the study had only narrow implications and was not generalizable. Based on his experience in training the teacher for the study he suggested that the typical elementary school teacher could adequately teach lessons on probability using a one-dimensional sample space. He cautioned, however, that more research and development needs to be done to determine more feasible ways of presenting problems in a two-dimensional sample space. As mentioned earlier in this report, reservations were also placed on the use of graphing situations calling for subtle interpretations.

A follow-up study of retention of probability concepts was conducted by Romberg and Shepler (1973) to examine the effects of the mastery learning technique utilized by Shepler in the study just reported. The same 72-item test was administered exactly four weeks after the posttest with no instruction or practice being given in that period. Posttest and retention test scores had a correlation of 0.78. The authors claimed a high retention rate. On the posttest 21 out of 25 (84%) had

achieved a 90% score. Seventeen of these 21 were still above a 90% level of performance and the other four were still above the 80% level.

The results also indicated that if the objective was originally mastered it was retained. If not, there was some loss. The eleven objectives successfully mastered in the instruction period had retention ratios in excess of 0.80; the other three objectives had ratios of 0.74, 0.54, and 0.43.

The authors acknowledged a weakness in such a study. The same test being used three times in a period of eight weeks could have caused test-retest interaction which was not controlled for. Also four weeks may not have been long enough to be practically significant. If further retention studies using parallel tests over longer periods were to corroborate the findings of their study, the investigators would then recommend further use of mastery-learning principles with this age group.

Two researchers did venture a little lower into the elementary grades in investigating the teaching of probability concepts. McLeod (1971) investigated the feasibility of teaching an eight- to ten-day unit to second- and fourth-grade children. He conducted two consecutive parallel studies. Experience gained in Study A was used to make modifications in instructional treatments and in the outcome measures in Study B. At both grade levels in a single school in each study, three experimental treatments were assigned at random to whole classes: laboratory participation (LP), teacher demonstration (TD), and no instruction

(M) between pretests and posttests. In study B control classes (C) were added from outside the basic school. Only the posttests were administered to these groups under C.

All the activities carried out in LP or observed in TD involved repeated chance events using the drawing of red and blue marbles from bags. The grade four activities were a little more extensive than those in grade two, otherwise the treatments were the same and common worksheets were used. Parallel 44-item tests were administered as pretests, posttests, and retention tests five weeks later.

Evidence from the pretests indicated that most second-grade as well as most fourth-grade children were able to apply the concepts of likely, more likely, equally likely, less likely, and unlikely before instruction began. Groups LP and TD were significantly superior to group C at both grade levels in Study B on the early posttest measures. No clear treatment effect was found for groups M, LP, and TD at either grade level on either posttest or retention measures. Learning apparently occurred under all three treatments. No clear effect was found for high or low groups classified on reading ability using Stanford Achievement Test scores and no effect was found due to sex.

The treatments LP and TD apparently improved the subjects' performance but so did M, no instruction apart from the pretest. This may have been due to a Hawthorn effect except that M was not significantly more effective than C, posttest only. Where knowledge is being measured rather than rate of production,

such a result tends to indicate that the knowledge is already present at the beginning. The pretest may have acted as an organizer of schema and consequent posttest performance was sufficiently improved so as to indicate no significant difference in the three treatment measures.

Rather than having strong implications about instructional procedures McLeod's study indicated the presence of a considerable range of understanding about chance situations in grade two and four children. Gipson (1971), at the same time, was examining just two concepts of probability and sought to give an accurate account of eight children's responses in learning the two concepts, finite sample space and probability of a simple event.

A procedure very similar to that outlined by Shepler (1969), reported earlier in this chapter was used by Gipson to develop an instructional sequence. Two pilot studies were used, the first to identify materials and appropriate concepts. In the second pilot study four children were taught on an individual basis and as a result three lessons were sequenced for presentation to third- and sixth-grade children. The lessons were taught to six children and audiotaped, and two more children were videotaped while being taught.

Pretest and posttest results and analysis of the children's protocols indicated the instructional sequence to be successful. The researcher reported that the interview-type procedure gave a deeper insight into how children think about probability concepts. The children were often able to explain how they

arrived at their answers relating to the two concepts. The conclusion was made that third grade is therefore an appropriate grade to introduce selected probability concepts.

Analysis of the protocols indicated that the most difficult performance objectives for the study were those related to specifying the estimated probability of the results for experiments and comparing equally likely outcomes using different objects. Gipson called for similar investigations to be made of children in grades four, five, and six.

It appears to the present investigator that Gipson's conclusion that third grade is the appropriate starting place for probability topics could be hasty. It would seem necessary to examine the situation with grades one and two more thoroughly before such a decision could properly be made. As indicated by studies reviewed in part II of this chapter, there is ample evidence that young children do have some understanding of some probability concepts. The nature and level of this understanding needs to be determined as a first step in designing curriculum material and instructional techniques suitable for the early grades.

Writers such as Harvey (1972) and Engel (1966, 1970) have provided detailed outlines of topics within probability that should be taught at upper elementary and junior high school levels. Engel's approach is a set theoretical one and rapidly becomes too advanced for most elementary students. Harvey, the other hand, begins with descriptive statistics and generates behavioral objectives for over 130 tasks. These also

seem to be pitched at upper elementary grades and are very statistics-oriented, but some of the earlier tasks appear suited to junior grades.

IV. SUMMARY

Piaget's experiments led him to propose three stages in the development of probabilistic thinking in children. In the first stage, four to seven years, children have little or no idea of random mixtures and fortuitous events. They tend to impute a hidden lawfulness to random processes and to explain outcomes in terms of egocentric and quasi-magical causal relations. In the second stage, seven to eleven years, children begin to recognize randomization and irreversibility in probabilistic situations and can understand most probability concepts. Piaget maintains that not until the third stage, beginning at age eleven or twelve, can the child really understand the process of random mixture and deal with the quantification of probability.

The main criticism of Piaget's theory relates to his first-stage proposal. Several studies have shown that young children, even of preschool age, can learn basic ideas about probability if the instructional conditions are reinforcing and the experiences are meaningful to the pupils. Other research has found that school children of all ages appear to develop concepts of probability prior to receiving any formal instruction in the topic. The intuitions about probability that children have are often incorrect and need to be determined and understood before

a meaningful program can be designed.

Classroom studies of material and instructional methods have proven the feasibility of teaching probability topics in elementary grades. At the same time these studies have shown the value in adopting a behavioral-objectives and mastery-learning approach to developing instructional units.

There is general agreement with Piaget that quantification of probability seems unlikely until age eleven or twelve although Fischbein reported that nine-year-olds had been successfully taught combinatoric skills which enabled them to compare ratios in probability situations.

Most studies showed that age and instructional conditions were significant factors in children's performance on probability items. Socioeconomic status and IQ were often found to be significant though no strong generalization could be made from all studies collectively.

Only Jones (1975) seems to have investigated the level of probability concepts at the grade one, two, and three levels. This present study was in part a replication of his study in that the effect of embodiments was investigated with these same grades and a game situation was used as motivation to maximize strategies. Choice between arrays was the required response for one part of the investigation and prediction of the estimated probability was required in another part. While the test required largely non-verbal response, found by Yost and others to be preferable with young subjects, verbalization of the basis for response was encouraged throughout the interviews.

Every attempt was made to fit the task-conditions to the subject's level of development in order to maximize their responses. Details of the design are given in the next chapter.

CHAPTER 3

INSTRUMENTATION AND RESEARCH PROCEDURES

I. INSTRUMENTATION

To test the hypotheses stated in Chapter 1 it was necessary to construct and administer an appropriate probability test and to administer the Canadian Cognitive Abilities Test. These instruments are described below.

Probability Test

Purposes. The purposes of the probability test were:

1. to obtain a measure of each subject's understanding of the six probability concepts in question,

2. to explore the extent to which subjects were able to quantify the probabilities in the settings presented to them, and

3. to determine the kinds of reasons that subjects gave for their responses.

Each concept was presented in three embodiments using spinners, blocks, and boxes. Throughout the test a total of five different probability settings were employed to elicit responses about the concepts being examined. Three of the settings were used in examining subjects' ability to quantify a probability representation.

Materials and apparatus. Three sets of devices were made:

1. five spinners each with equal sectors outlined in black

and colored with plastic adhesive tape,

2. five 2 cm wood blocks covered with the same plastic adhesive tape, and

3. five plastic boxes each containing six plastic counters, which were the same shades of color as the adhesive tape.

The proportions of blue, red, and yellow for each of the fifteen devices in the test are given in table 1.

TABLE 1

PROPORTIONS OF BLUE, RED, YELLOW IN THE DEVICES USED IN THE PROBABILITY TEST

Device	Proportions (B:R:Y)				
Spinner	2:2:2	3:2:1	4:1:1	3:3:0	0:6:0
Block	2:2:2	2:3:1	1:4:1	3:0:3	6:0:0
Box	2:2:2	1:2:3	1:1:4	0:3:3	0:0:6

A white laminated race-game board was made (see Appendix A) and a collection of six markers was used, one of each color blue, red, yellow, white, green, and brown. The researcher was careful to arrange the materials so that counters, markers, and plastic tape were the same shade of blue, red, and yellow.

For use in the introductory activity, a half green half white spinner and a half red half yellow block were also made. Except for these two, the devices were assembled so that no one color appeared to be used more often than another and color, devices, and correct response in the test were randomly associated.

Responses. Three types of response were sought from each subject corresponding to the three purposes of the test.

1. Choice response, to measure understanding of the six concepts. For example, "This spinner" (pointing to it) or "This marker" (picking it up or touching it). Twenty one such responses comprised the concept subtest.

2. Predictive response, to measure subjects' quantitative understanding of probability. For example, "Four times out of six goes I'll get yellow" although "Four" was the usual abbreviated version of such a response. Eighteen such responses sought from each subject comprised the quantitative subtest.

3. Rationalization response, in answer to the question "Why did you choose that one?" after each type (1) response, or "Why that number of times?" after each type (2) response. Three possibilities existed at this point:

(a) A rational, correct explanation was given based on the probability settings present, for example "it has more blue than red or yellow" or "there are four red sides on this block so we will be likely to get red four times out of six".

(b) A non-rational explanation was given involving influences such as favoritism of color, position, or quasi-magical properties attributed to a color or to a device.

(c) No response was given or the child said "I don't know" or simply "No reason".

A subject was never pressed for an answer beyond one repetition of the question and every effort was made to ensure that the subject did not feel threatened but in fact enjoyed the

interview. The researcher was encouraged to believe he succeeded in this by the number who expressed a wish to return and "play again" and by the number of children not in the sample who approached him at the school to request inclusion.

Description of the Probability Test

Introductory activity: the game. As most of the questions in the test were presented in the setting of a simple race-game the interview began with an introduction to each other and to the game. The researcher placed the game-board, the green/white spinner, the green marker, and the white marker in front of the subject who was asked if he liked to play games. All students answered in the affirmative. Most had seen a spinner before but only the type with numerals on the face. The researcher explained and demonstrated how the two markers were used, that one advanced one square on the board when the spinner stopped on its color. The first to reach "finish" was the winner. The subject was invited to choose one of the markers and the game was played to a conclusion with either subject or investigator operating the spinner.

The spinner and markers were removed and the red/yellow block and matching markers were placed on the board. After a brief examination of the block and its use in the game (e.g. red the uppermost face means the red marker moves up one square) the subject and investigator played the game again to a conclusion. By that stage all subjects said they knew how to play the game and the test was then begun.

The test organization and design. The test contained fifteen tems which were related to the concepts, quantification trials, and mbodiments as shown in Table 2.

TABLE 2

ITEMS AND EMBODIMENTS BY WHICH CONCEPTS AND QUANTIFICATION WERE TESTED

Concepts	Items	Embodiment*
Sample space events	1(a),(b); 2(a),(b); 3(a),(b)	2-2-2
Most favorable event	4(a), 5(a), 6(a)	3-2-1
Most favorable sample space	7(a), 8(a), 9(a)	4-1-1
Equally favorable sample space	10(a), 11(a), 12(a)	2-2-2
Impossible event	13(a), 14(a), 15(a)	3-3-0
Certain event	13(b), 14(b), 15(b)	6-0-0
Quantification		
Six trials	Part (b) of 4 through 12	3-2-1 4-1-1 2-2-2
Twelve trials	Part (c) of 4 through 12	3-2-1 4-1-1 2-2-2

* each entry represents the three settings which are isomorphic to the given one (e.g. 3-2-1 means 3-2-1, 2-3-1, and 1-2-3 were used).

Items one, two, and three were presented, in random order, followed by items four through fifteen, again in random order. In the first three items more effort was made to "draw out" the desired response than in later items. It was felt that the concept of sample space was basic to all the others. Without an appreciation of what the possible events in a given situation were, a subject would find later items unnecessarily difficult if not meaningless. Subjects were not asked for rationalizations for their responses on these first three items.

From Table 2 it can be seen that a total of 21 questions was asked relating to concepts and 18 relating to quantification predictions. The items are now described in detail.

Description of the Test Items

Item 1.

(a) The researcher placed the six markers near the game board. The 2-2-2 spinner was shown to the subject who was asked "If this spinner were used to play the game which of these markers would be used?" Usually the subject selected the three correct markers and placed them in the starting squares on the board. If only one marker was selected, the question "Is that all?" was asked until the subject gave a firm "yes".

(b) With just the 2-2-2 spinner before the subject, the researcher asked "If I were to spin this spinner what could I get?" With few exceptions subjects responded correctly in one of two forms. For example, "A red, or a blue, or a yellow" or "You could get any of the colors". The probing question "Anything else?" was

allowed once with each subject if needed.

Item 2.

(a) The board and six markers were presented as described in Item 1(a) but this time the 2-2-2 block was used and the question adjusted to read "block" instead of "spinner".

(b) With just the 2-2-2 block before the subject the researcher asked. "If I were to roll this block what could I get?" The same procedure for probing as in item 1 was employed as needed.

Item 3.

(a) The game board and six markers were presented as in items 1(a) and 2(a). The 2-2-2 box was presented and its use explained to the subject. When the researcher felt sure the subject understood how the box could be used to play the game he asked the question, "If this box were used to play the game which of these markers would be used?".

(b) With just the 2-2-2 box before the subject the researcher asked "If I were to reach into this box and draw out a counter without looking what could I get?". The same probing procedure as described in item 1 was employed as needed.

In each of the following items the game board was before the subject with a blue, a red and a yellow marker in the start squares. Each question relating to the game was in the context of these three "players".

Item 4.

The 3-2-1 spinner was presented and the subject was asked "If using this spinner which marker would you choose to

win the game?". When a response was given the subject was then asked "Why that one?".

(b) The subject was then asked "In six spins how many times would you expect to get blue?". If a response was given the subject was encouraged to state a reason for his response by the question "Wh[illegible]he number)?".

([illegible] subject was then asked "In twelve spins how many times would you expect to get blue?". Again, if he responded, his reason was elicited by the question "Why (the number)?".

Item 5.

This item was identical to item 4 except that the embodiment was the 2-3-1 block and the wording of the questions was altered accordingly:

(a) "If using this block which marker would you choose to win the game?"

(b) "In six rolls how many times would you expect to get red?"

(c) "In twelve rolls how many times would you expect to get red?"

Item 6.

This item was identical to item 4 except the embodiment was the 1-2-3 box and yellow was the color most favored. The questions were worded as follows:

(a) "If using this box which marker would you choose to win the game?"

(b) "In six draws how many times would you expect to get yellow?"

(c) "In twelve draws how many times would you expect to get yellow?"

Item 7.

(a) The 2-2-2, 3-2-1, and 4-1-1 spinners were placed before the subject who was asked "Which spinner gives blue the best chance of winning?". When a response had been given, the question was asked "Why that one?".

(b) Referring only to the 4-1-1 spinner the researcher asked "In six spins how many times would you expect to get blue?". The 2-2-2 and 3-2-1 spinners were removed from view prior to putting this question in order to encourage the subject to attend only to the 4-1-1 setting. The subject's reason for any response was sought by asking "Why (the number)?".

(c) The subject was then asked with regard to the 4-1-1 spinner only "In twelve spins how many times would you expect to get blue?". Again the follow up probe "Why (the number)?" was used.

Item 8.

This was a replication of item 7 in the block embodiment.

(a) The 2-2-2, 2-3-1, and 1-4-1 blocks were presented and the subject was asked "Which block gives red the best chance of winning?" and "Why that one?".

(b) Only the 1-4-1 block was left before the subject who was asked "In six rolls how many times would you expect to

get red?". Upon giving a response the subject was asked "Why (the number)?".

(c) Referring only to the 1-4-1 block the researcher asked "In twelve rolls of this block how many times would you expect to get red?" and "Why (the number) times?".

Item 9.

This was the box embodiment of item 7.

(a) The 2-2-2, 1-2-3, and 1-1-4 boxes were presented and the subject was asked "Which box gives yellow the best chance of winning?".

(b) Referring only to drawing from the 1-1-4 box, the researcher asked "In six goes how many times would you expect to get yellow?".

(c) Again referring to box 1-1-4 the question was asked "In twelve goes how many times would you expect to get yellow?". Each time a response was given the subject's reason for that response was elicited by the question "Why that one?" or "Why that number?".

Item 10.

(a) The 2-2-2, 3-2-1, and 4-1-1 spinners were presented to the subject who was asked "Which spinner gives each player (indicating the three markers in place on the game board) the same chance of winning?". Occasionally a subject appeared not to understand the question whereupon the researcher rephrased the question to "Which spinner makes the game fair for each of the three players?". The subject's reason for his response was

again sought by the question "Why this one?".

(b) The 3-2-1 and 4-1-1 spinners were removed from the subject's view and, referring only to the 2-2-2 spinner, the subject was asked "In six spins how many times would you expect to get red?" and "Why that number?".

(c) Again referring only to the 2-2-2 spinner the subject was asked "In twelve spins how many times would you expect to get red?" and "Why that number?".

Item 11.

The procedure was as described in item 10 except that the 2-2-2, 2-3-1, and 1-4-1 blocks were used in (a), yellow was the chosen color in (b) and (c), and appropriate wording was used in the questions (e.g. "rolls" instead of "spins"). The subject was always given the opportunity to state his reason for his response as described in previous items.

Item 12.

The procedure was as described in item 10 except that the 2-2-2, 1-2-3, and 1-1-4 boxes were used in (a), blue was the chosen color in (b) and (c), and appropriate changes were made to the terms in the questions (e.g. "draws" or "goes" instead of "spins"). Again the probe questions "Why that one?" or "Why (the number)?" were always asked.

Item 13.

(a) The 2-2-2, 3-2-1, 4-1-1, 3-3-0, and 0-6-0 spinners were placed before the subject who was instructed to look

carefully at all of them. The researcher said "We have these three players (pointing to the markers) in the game. Which one of these spinners makes it so that just one of the players can never win?". This was repeated carefully with the alternative "will always lose" being given at the end. When a spinner was selected the subject was asked "And which marker will always lose if we use this spinner?". Following a nomination of marker(s) the researcher asked "Why?".

(b) Restoring the spinner selected in (a) to the array of five spinners the researcher said, "Look carefully at the spinners and tell me which spinner makes it so that one of these players will always win?". This was repeated if necessary and again the subject was required to indicate a spinner and a marker and to give a reason for the choice.

Item 14.

The 2-2-2, 2-3-1, 1-4-1, 3-0-3, and 6-0-0 blocks were placed before the subject. The questions (a) and (b), with probes, were asked as described in item 13 with the appropriate change of wording to "blocks" instead of "spinners".

Item 15.

The 2-2-2, 1-2-3, 1-1-4, 0-3-3, and 0-0-6 boxes were placed before the subject. Questions (a) and (b) and probes as described in item 13 were asked using the word "boxes" instead of "spinners".

Pilot Study of the Probability Test

A draft form of the test was piloted on March 23, 1978 in a single Edmonton school using three subjects from each of grades one, two, and three. The test was administered individually to each subject in a manner similar to that described in the preceding paragraphs. A seven inch reel-to-reel tape recorder was used to record the interviews. The first draft of the test contained only items which investigated the six concepts in question but not quantification.

The purpose of the pilot study was to trial the items, the materials, the interview procedure, and the recording technique. In addition it helped to satisfy the researcher that the instrument was appropriate to the grade levels involved and that the items were valid measures of the elementary probability concepts. Feedback from a panel of six independent mathematics educators to whom the findings of the pilot study were reported supported the claim for validity of the items.

Two major changes were made as a result of the pilot study. The tape recorder provided marginal assistance in the collecting of data and was therefore not used in the main study. The responses given by the subjects were easily categorized, coded, and recorded by the researcher on an answer sheet and the tape added little additional information. Secondly it was decided to inquire further into the subjects' probabilistic understanding by including the quantitative parts of the test items which were not included in the pilot form of the test (parts (b) and (c) of items 4 through 12). These were then trialled separately on the

researcher's own three children, ages nine, seven, and five years, mainly to streamline the actual wording of the questions and the presentation of the materials.

The pilot study proved to be of great value also in providing the investigator with practice in interviewing and recording with young children.

Canadian Cognitive Abilities Test (CCAT)

The second instrument used in the study was for the purpose of gaining a measure of the subject's general reasoning ability. As grade three subjects in the school system had recently been tested with the Canadian Cognitive Abilities Test it was decided to administer the companion tests in the same series to the grade one and two subjects.

"The Cognitive Abilities Test is part of an integrated test series designed to assess the cognitive development of abilities from kindergarten through grade nine" (Thorndike, 1968, p. 4). Primary 1 and Primary 2, designed for grades one and two respectively, are group tests using pictorial materials and oral instructions. There are four short subtests in each, oral vocabulary, relational concepts, multi-mental ("one that doesn't belong"), and quantitative concepts.

Norms were established for CCAT Primary in 1966 using 50 schools across Canada supplying approximately 2 000 pupils in each of grades K to 4. Construct validity was established using a factor analysis which showed that all batteries in the series give measures of a general reasoning factor. Reliability

coefficients of from 0.769 to 0.887 are given for the tests.

A table is supplied to allow easy conversion of raw scores and chronological age into deviation IQ's with a mean of 100 and standard deviation of 16. These are standard scores directly comparable from age to age.

II RESEARCH PROCEDURES

Selection of Sample

The Edmonton Public School Board made one school available for data collection. It was situated in a rapidly developing residential neighborhood regarded as representative of Edmonton's middle to low socio-economic status areas. The principal of the school was contacted and asked to make 24 pupils available to the researcher from each of grades one, two, and three. A further request was made for the sample to have equal numbers of boys and girls and equal numbers of pupils of high and low general reasoning ability.

As no standardized IQ scores were available for grade one and two pupils the principal suggested that the teachers' perception be the means of judging the general ability levels of all of the subjects. In this way the sample of 72 subjects was selected from a population of 176 grade one, two, and three pupils. The principal had the list of names of subjects ready for the researcher on his arrival at the school for data collection.

Data Collection

The data were collected in the last two weeks of April, 1978. The researcher, assisted by his wife, spent eight days at the school interviewing subjects individually and administering the Cognitive Abilities Test in several sittings to the grade one and two subjects.

The probability test was administered on an individual basis in a small room made available by the principal. The procedure was as follows: familiarize the subject with the game by playing it as outlined previously, present the items 1, 2, and 3 in a random order, and present items 4 through 15 in a random order.

To facilitate this procedure question cards were made for each of the fifteen items. Each card indicated the item number, listed the apparatus to be used and showed each part of the question as it was to be stated. Labelling of the apparatus with the appropriate tri-numeral description of the probability setting enabled the researcher to quickly change the materials between items.

After each interview the researcher or assistant ordered the cards according to the next row in a table of random numbers generated especially for the purpose. In this way an attempt was made to control for a teaching effect which might otherwise influence scores in items occurring towards the end of the test.

The average time for an interview was approximately twenty minutes. Twelve subjects were tested each day, usually seven

in the morning and five after lunch.

Each subject's responses were recorded by the researcher on individual profiles. These responses were later coded, along with all the identification information, onto a data form using a one or zero according to whether the response was correct or not. The rationalizations for responses given throughout the interviews were noted and later were tallied into a frequency table for analysis. The range of replies was sufficiently narrow for the researcher to classify them into three groups as described in the first section of this chapter.

The remaining data, age and IQ were secured from the cumulative record cards in the case of grade three subjects. For grade one and two subjects, as mentioned above, the IQ had to be obtained from direct testing by the researcher. As the school would have had little use for IQs for grade one and two pupils at that late stage in the school year the principal preferred that only the sample subjects be tested.

The Primary 1 Form 1 test in the Canadian Cognitive Abilities Test series was administered to the 24 grade one subjects by the researcher and his assistant in three thirty minute sittings over two days. The Primary 2 Form 1 test in the same series was similarly administered to the 24 grade two subjects. The IQs were derived according to the test manual and were added to the subject's data card.

Finally the occupation of each subject's father was noted from the cumulative record cards for the purpose of gaining an average occupational class rating of the sample (Blishen, Jones,

Naegele, & Porter, 1968). The Blishen index was based on income and education characteristics of incumbents of 320 occupations in the 1961 Canadian census and is utilized in this study simply as a check on the judgment of both the researcher and the school principal as to the socio-economic status of the sample. An average index of 43 on a scale ranging from 25 to 77 confirms the assessment of middle to low SES which was stated earlier.

Analysis of Data

The data were analysed with the assistance of the computer facilities in the Division of Educational Research Services at the University of Alberta. The encoded data were entered into a file and several standard statistical programs were utilized.

Test statistics were calculated initially for the total sample and these are reported in the analysis of the instrument below. Further analysis of the test responses included item frequencies, responses according to embodiment, and means and variances within IQ, grade, and sex groups. One- and three-way analyses of variance were performed to test for effect of embodiment and effect due to sex, grade, and IQ.

These analyses are reported and discussed in the next chapter along with an analysis of subjects' rationalizations.

III. ANALYSIS OF THE PROBABILITY INSTRUMENT

This section reports a post-administration analysis of the probability test using the data collected in the present study.

Table 3 contains the means and standard deviations for the

total test scores and two subtest scores for the whole sample and for each grade.

TABLE 3

MEAN AND STANDARD DEVIATION ON PROBABILITY TEST AND SUBTESTS USING TOTAL SAMPLE AND GRADES

	Total score		Concept		Quantification	
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
Total sample	22.81	5.91	17.21	3.31	5.63	4.10
Grade 1	19.88	6.09	15.13	3.44	4.83	4.37
Grade 2	21.79	4.66	17.08	3.15	4.71	3.42
Grade 3	26.75	4.57	19.42	1.35	7.33	3.91

The three scores are out of 39, 21, and 18 respectively and the means indicate the relative difficulty of the quantitative items.

An analysis of variance indicated that grade was a significant factor in all three criterion measures. The Scheffe method of multiple comparisons of means was used to compare the performance of grade levels two at a time (Ferguson, 1971, p. 270-271). The probabilities derived from the Scheffe test are given in Table 4. Significant differences at the 0.01 level occurred between the total score means for grades one and three and grades two and three. The grade one and grade three means were also significantly different at the 0.01 level on the concept subtest.

TABLE 4

PROBABILITIES FOR SCHEFFE MULTIPLE COMPARISONS OF GRADE MEANS ON TOTAL AND SUBTEST SCORES

Grade means compared	Test		
	Total	Concept	Quantification
1 vs 2	0.4553	0.0675	0.9942
1 vs 3	0.0001	0.0000	0.1039
2 vs 3	0.0071	0.0231	0.0831

The distribution of scores on the test is shown in Figure 1. For the total sample the distribution of scores was found to be normal by the chi-square goodness of fit test. The skewness was -0.028 and the kurtosis was -0.268. The chi-square of 5.744 with 6 degrees of freedom had a probability of 0.453. Neither of the subtests was indicated to be normally distributed, the chi-square probabilities being less than 0.001 and 0.006 for concept and quantification subtests respectively. These measures are summarized in Table 5.

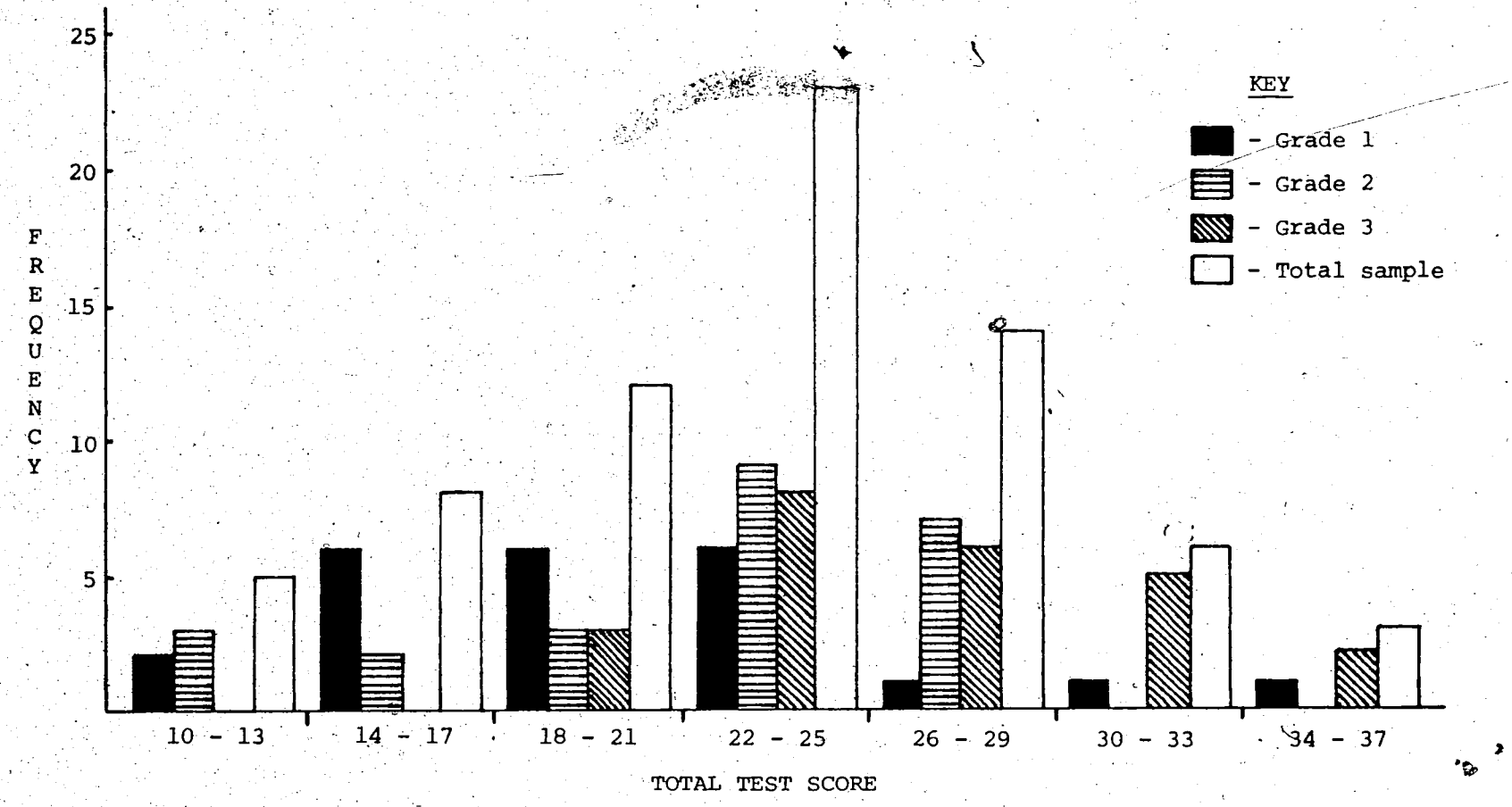
TABLE 5

SKEWNESS, KURTOSIS, AND CHI-SQUARE FOR TEST AND SUBTESTS ON TOTAL SAMPLE

Test	Skewness	Kurtosis	χ^2	df	Prob. of χ^2
Total	-0.028	-0.268	5.744	6	0.453
Concept	-1.069	0.858	19.699	4	0.001
Quantification	0.722	0.182	10.146	2	0.006

FIGURE 1

DISTRIBUTION OF SCORES ON THE PROBABILITY TEST BY GRADES AND TOTAL SAMPLE



As a measure of item difficulty the percent of responses correct was calculated for each item. These percentages are presented in Table 6 where the items are grouped in the two subtests.

The concept items were correctly answered at least 73% of the time except for those relating to concept five, the impossible event. By contrast the percentages for the quantification items were less than 50% except for item 9(b). More discussion of particular groupings occurs in chapter 4.

TABLE 6

PERCENTAGES CORRECT ON PROBABILITY TEST ITEMS FOR TOTAL SAMPLE

Concept items				Quantification items			
Item	%	Item	%	Item	%	Item	%
1 a	100	9 a	96	4 b	41	9 b	53
b	99	10 a	74	c	31	c	17
2 a	98	11 a	73	5 b	42	10 b	41
b	98	12 a	77	c	32	c	12
3 a	96	13 a	57	6 b	45	11 b	41
b	99	b	81	c	25	c	16
4 a	92	14 a	46	7 b	35	12 b	46
5 a	80	b	87	c	13	c	19
6 a	81	15 a	46	8 b	34		
7 a	91	b	85	c	13		
8 a	82						

Reliability

A test-retest reliability check was made on the final version of the probability test. Eighteen of the subjects (25% of the sample) were randomly chosen for retesting which was done in one day in the third week of May, 1978, three weeks after the original testing. All items were administered in the same interview situation as employed initially. Pearson product-moment coefficients of correlation were calculated on test and retest total, concept, and quantification scores. The correlation coefficients and test and retest means and standard deviations are given in Table 7.

TABLE 7

TEST-RETEST RELIABILITY OF PROBABILITY TEST

Criterion score	Test		Retest		Correlation coefficient
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	
Total	23.6	5.27	24.4	4.78	0.837
Concept	17.1	2.88	18.5	2.27	0.796
Quantification	6.6	3.67	5.9	3.02	0.753

The differences between the test and retest parameters were judged to be small enough and the correlation large enough for the test to be considered reliable for experimental use.

CHAPTER 4

RESULTS OF THE INVESTIGATION

This chapter reports the results of testing the hypotheses and of other analyses made on the data. The results are reported separately for each of the four major purposes stated in Chapter 1. The questions and hypotheses listed under each purpose are discussed in turn.

I. STATUS OF THE PROBABILITY CONCEPTS

Question One Restated: What proportion of subjects in each grade and in the total sample indicate an understanding of the six concepts investigated?

At the end of the previous chapter an indication of item difficulty was provided in Table 6 as the percentage correct on each item in the probability test for the whole sample of 72 subjects. To arrive at a status index for each probability concept embodied in the test items average relative frequencies were computed for the set of items relating to each of the six concepts. For concept one, items one, two, and three (two parts in each) constituted the set of related items. Each of the other five concepts was tested by three items, one for each embodiment. The relative frequencies for each concept are given as percentages in Table 8 for the whole sample and for each grade.

The percentages reported in Table 8 indicate that the first concept, sample space, was understood by almost all subjects,

whereas only half of the responses in the whole sample were correct to items about concept five, impossible event. The other concepts ranked between these extremes in the following order: most favorable sample space, certain event, most favorable event, and equally favorable sample space.

TABLE 8

MEAN PER CENT CORRECT RESPONSES FOR PROBABILITY CONCEPTS IN EACH GRADE AND THE WHOLE SAMPLE

Concept	Grade (N=24)						Total Sample (N=72)	
	1		2		3			
	n	%	n	%	n	%	n	%
Sample space	23	96	24	100	24	100	71	99
Most favorable event	18	75	19	79	23	96	60	83
Most favorable sample space	18	75	23	96	23	96	64	89
Equally favorable sample space	14	58	17	71	22	92	53	74
Impossible event	9*	38	11	46	16	67	36	50
Certain event	18	75	20	83	23	96	61	85

* Not significantly better than chance (0.01)

The increase in percentages on all concepts through grades one to three indicates that these concepts develop with age.

Analysis within grades indicates that four of the concepts

were understood by at least 75% of the subjects irrespective of grade. These were concepts one, two, three, and six. If 75% were taken as a minimum criterion for acceptable level of probability concepts before formal instruction should begin, then this study indicates that four of the six concepts investigated meet that criterion in each grade and in the total sample.

Hypothesis Two Restated: The proportions derived in answering question one are not significantly different from chance proportions for each concept.

In order to determine the number of correct responses beyond the number obtainable by chance, the cumulative binomial distribution was used. The items relating to concepts two, three, and four had three possible outcomes and those relating to concepts five and six gave five alternatives. If a subject were to respond at random, the probability (p) of a correct response based only on a random choice would be 1/3 and 1/5 respectively. With p=1/3 and N=72, the binomial probability of any item obtaining more than 33 correct responses is less than 0.01. With p=1/5 and N=72, the critical number is 23. Table 8 reveals that all the concepts received more than the critical number of correct responses from the whole sample.

The binomial probabilities were also found for the number in each grade in the sample. For p=1/3 and N=24, the probability of any item obtaining more than 13 correct responses by chance is less than 0.01. For p=1/5 and N=24, the critical number is 10

responses. Table 8 shows that only one concept has less than the significant number of correct responses in any of the grades. This is concept five, impossible event, at the grade one level. With this one exception, the hypothesis is not accepted. The conclusion is that all six concepts received correct responses in the whole sample significantly more often than is attributable to chance.

II. LEVEL OF QUANTIFICATION OF PROBABILITY

Question Three Restated: What proportions of subjects in each grade and in the total sample indicate an understanding of the quantification items presented?

The data in Table 6 were used to compute the mean per cent of correct responses to the three questions on each of the six quantification situations. Table 9 presents the mean number and per cent of correct responses for each grade and for the total sample. As expected the items on quantification were much more difficult than those on the concepts. This is indicated by the lower percentages throughout Table 9, the highest rate of success being 58% achieved by grade three subjects in predicting the expected frequency of an event in six trials using the 2-2-2 settings.

Performance improved with the age of subjects except for a decline in grade two on two of the settings. The average percentages of correct responses indicate little difference between grades one and two with averages of 25% and 24% respectively,

but grade three subjects performed significantly better with an average of 40% over all situations. (A more complete analysis of differences between grades is deferred until hypothesis eight is discussed.)

TABLE 9

MEAN PER CENT CORRECT RESPONSES ON QUANTIFICATION ITEMS IN EACH GRADE AND THE WHOLE SAMPLE

Probability setting	Grade (N=24)						Whole Sample (N=72)	
	1		2		3			
	n	%	n	%	n	%	n	%
2-2-2								
6 trials	8*	33	8*	33	14	58	30	42
12 trials	3*	13	1*	4	6*	25	11*	15
3-2-1								
6 trials	8*	33	10	42	12	50	30	42
12 trials	4*	17	6*	25	11	46	21	29
4-1-1								
6 trials	9*	38	8*	33	11	46	29	40
12 trials	4*	17	2*	8	4*	17	10*	14

* Not significantly better than chance (0.01)

Hypothesis Four Restated: The proportions derived in question three do not differ significantly from chance proportions.

In answering the questions concerning quantification, the subjects were asked to predict how many times in six or twelve

trials a certain event would occur. If a subject were to respond in a completely random manner, the chance of his response being the correct one would be 1/6 or 1/12 respectively.

On this basis, the cumulative binomial distribution was used to determine the number of correct responses beyond the number obtainable by chance. With p=1/6 and N=72, the binomial probability of any item obtaining more than 20 correct responses is less than 0.01. With p=1/12 and N=72, the critical number is 12. Table 9 shows that, for the whole sample, responses on all three settings were significantly different from chance responses for the six-trial items but significant only on the 3-2-1 setting for the twelve-trial items.

Similarly, the critical points were found for N=24 and p=1/6 and 1/12, so that within each grade the number of correct responses obtainable by chance alone could be compared to the performance of subjects in that grade. For the six-trial items the binomial probability of any item obtaining more than nine correct responses is less than 0.01, while the critical number for the twelve-trial items is six. Table 9 shows that 13 of the 18 mean responses within the three grades are not significantly better than chance at the 0.01 level of confidence.

For grade one, hypothesis four should be accepted for all settings and trial sizes. For grade two, the hypothesis should be accepted except for the six-trial 3-2-1 setting. For grade three the hypothesis is not accepted except for the twelve-trial questions with the 2-2-2 and 4-1-1 settings.

In view of the above analysis, it would be unwise to judge

this hypothesis solely on the total sample response as it is biased by the significantly higher correct response rate of the grade three students. This analysis supports the conclusions stated regarding question three that subjects in grades one and two performed poorly on the quantification items while over 50% of the grade three subjects indicated understanding in the six-trial situations.

An indication of the distribution of the actual responses to the quantification items is given in Tables 17 and 18 in Appendix B.

III. THE EFFECT OF EMBODIMENT

Hypothesis Five Restated: There are no significant main effects due to embodiment, on performance on the probability test.

The hypothesis was tested using a one-way analysis of variance with repeated measures. Each concept and quantification question was essentially asked three times, once with each embodiment of isomorphic probability settings. The embodiments can thus be regarded as repeated measures of the same criterion.

The analysis of variance was carried out for each grade and for the total sample. The results of the comparison of mean scores on each embodiment, within grades and within the whole sample, are given in Table 10.

The conservative probability of F was calculated by making allowance for unequal covariances among the correlated measures (Winer, 1971, pp. 281-282).

At the 0.01 level of confidence there were no significant differences between means for the three embodiments at any grade level. The hypothesis is therefore accepted for all three grades.

TABLE 10

COMPARISON OF EMBODIMENT MEANS FOR EACH GRADE AND THE TOTAL SAMPLE

Subjects	Embodiment means			F	d.f.	Conservative probability of F
	Spinner	Block	Box			
Total sample	7.61	7.35	7.81	2.23	1, 71	0.140
Grade 1	6.46	6.13	7.29	6.57	1, 23	0.017
Grade 2	6.96	7.29	7.42	0.73	1, 23	0.401
Grade 3	9.42	8.63	8.71	2.76	1, 23	0.110

Hypothesis Six Restated: There are no significant main effects due to embodiment on the probability test performance when sex, grade, and IQ are used as blocking variables in pairs.

The hypothesis was tested using a three-way analysis of variance with the embodiment factor as a repeated measure. Three analyses were carried out using, two at a time, the factors sex, grade, and IQ for the blocking variables. A summary of the analyses of variance within subjects relating to embodiment is given in Table 11.

None of the F-ratios for embodiment (E) are significant at

the 0.01 level of confidence. The hypothesis, that there are no significant main effects due to embodiment, is retained.

TABLE 11

SUMMARY OF THREE ANALYSES OF VARIANCE WITHIN SUBJECTS DUE TO EMBODIMENT AND SEX, GRADE, AND IQ

	Source of variance	df	MS	F	Prob. of F
(a)	Embodiment (E)	2	3.81	2.40	0.095
	E x Sex (A)	2	1.25	0.79	0.456
	E x Grade (B)	4	5.37	3.38	0.011
	E x A x B	4	2.16	1.36	0.250
	Error	132	1.59		
(b)	Embodiment (E)	2	3.81	2.26	0.108
	E x Sex (A)	2	1.25	0.74	0.477
	E x IQ (C)	2	4.85	2.87	0.060
	E x A x C	2	0.37	0.22	0.806
	Error	136	1.69		
(c)	Embodiment (E)	2	3.81	2.48	0.088
	E x Grade (B)	4	5.37	3.50	0.010
	E x IQ (C)	2	4.85	3.15	0.046
	E x B x C	4	2.08	1.35	0.254
	Error	132	1.54		

Hypothesis Seven Restated: There are no significant interactions between the embodiment and the factors sex, grade, and IQ.

The same three-way analysis of variance used to test hypothesis six also tested for interactions between the variables. The data in Table 11 indicate that the only interaction significant at the 0.01 level is between grade level and embodiment when grade and IQ are the blocking variables. When considered with sex, grade just fails to show a significant interaction with embodiment at the 0.01 level of confidence.

The hypothesis of no interaction with embodiment is accepted for sex and IQ but judgment is reserved in the case of grade.

IV. THE EFFECTS OF SEX, GRADE, AND IQ

Hypothesis Eight Restated: There is no significant effect on performance on the probability test due to (a) sex, (b) grade, and (c) IQ.

This hypothesis was tested by a three-way analysis of variance for the three criterion variables, total score, concept subtest score, and quantitative subtest score. The results are reported separately for each criterion in Table 12.

On the total test score, all three factors are shown to have an effect which is significant at the 0.01 level of confidence. On the concept subscore, the effects of grade and IQ are shown to be significant at the 0.01 level, but sex only at the 0.05 level. On the quantitative subscore all factors appear to have a significant effect only at the 0.05 level of confidence.

TABLE 12

SUMMARY OF ANALYSIS OF VARIANCE

FOR RESPONSES TO THE TEST AND SUBTESTS

Source	df	Total		Concept		Quantification	
		F	Prob.	F	Prob.	F	Prob.
Sex (A)	1	9.02	0.004	4.55	0.037	5.40	0.024
Grade (B)	2	13.67	0.000	16.63	0.000	3.29	0.044
IQ (C)	1	13.13	0.001	11.69	0.001	4.84	0.032
A x B	2	1.10	0.340	0.23	0.795	1.56	0.219
B x C	2	1.96	0.149	1.23	0.300	1.49	0.233
A x C	1	0.00	0.985	2.48	0.120	1.42	0.237
A x B x C	2	0.17	0.845	0.99	0.378	0.25	0.780
Error	60						

On the basis of the analysis reported above, the hypothesis is rejected for each of the factors on the total score. It did make a difference to the total test score whether the subject was a boy or a girl, whether in grade one, two, or three, and whether high or low in general reasoning ability.

These differences in performance are evident in Table 13 which gives the mean on each criterion measure for each grouping of subjects, and in Table 14, which gives the means for sex and IQ groupings nested in grades. Figures 2, 3, and 4 are graphical representations of the means in Table 14 for the total score on the probability test and the two subscores. On each criterion

TABLE 13

CRITERION MEANS BY SEX, GRADE, AND IQ

Criterion measure	Factors							Grand mean	S.D.
	Sex		Grade			IQ			
	B	G	1	2	3	Low	High		
Total (/39)	21.4	24.2	19.9	21.8	26.8	20.8	24.8	22.8	5.9
Concept (/21)	16.6	17.8	15.1	17.1	19.4	16.3	18.2	17.2	3.3
Quant. (/18)	4.8	6.5	4.8	4.7	7.3	4.6	6.7	5.6	4.1

TABLE 14

CRITERION MEANS BY SEX AND IQ NESTED WITHIN GRADES

Criterion measure \ Grade	1				2				3			
Sex	B		G		B		G		B		G	
IQ	L	H	L	H	L	H	L	H	L	H	L	H
Total (/39)	14.5	21.4	18.1	24.2	18.7	23.8	20.6	24.8	23.9	24.0	28.4	30.4
Concept (/21)	12.5	15.5	14.9	17.0	14.0	19.3	17.1	18.0	18.3	19.4	19.8	20.3
Quant. (/18)	2.0	5.9	3.3	7.6	4.7	4.5	3.4	6.8	5.6	4.6	8.6	10.1

measure girls outscored boys, grade three scores were higher than those of grade one or grade two subjects, and the high-IQ subjects scored higher than the low-IQ subjects.

Hypothesis Nine Restated: There are no significant interactions between the independent variables sex, grade, and IQ and the criterion measures in the test.

The analysis of variance summarized in Table 12 included tests for interaction effects. No significant interactions were found between the variables. The lowest probability assigned to any of the interactions is 0.120 for the Sex x IQ interaction. The hypothesis is therefore accepted. Sex, grade, and IQ did not interact with one another to produce any differential effects on probability test achievement.

FIGURE 2

TOTAL SCORES BY IQ-SEX GROUPS FOR EACH GRADE

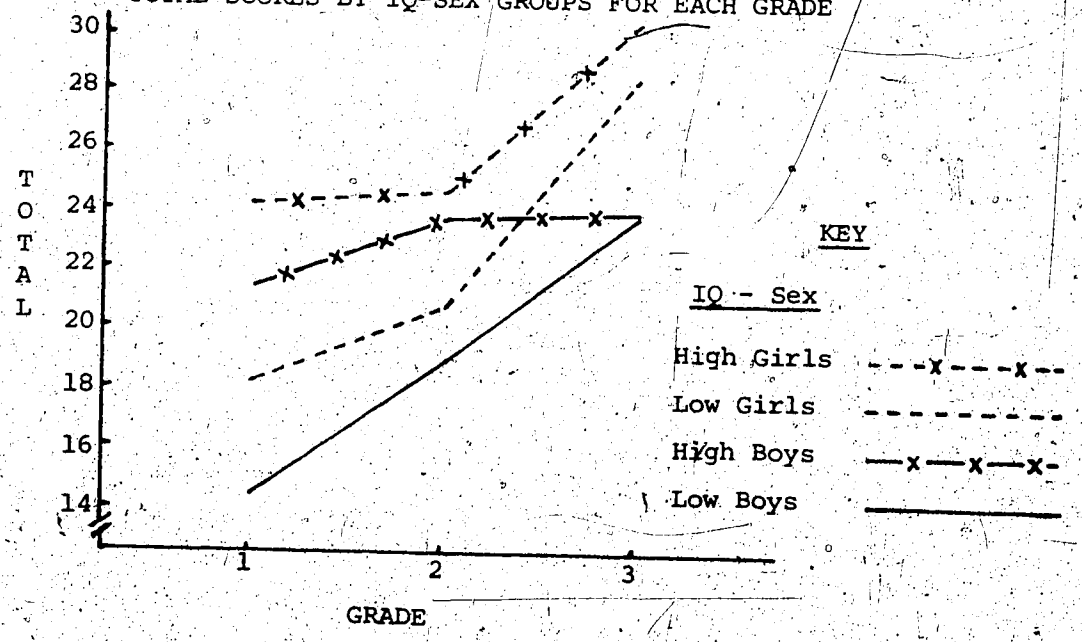


FIGURE 3

CONCEPT SCORES BY IQ-SEX GROUPS FOR EACH GRADE

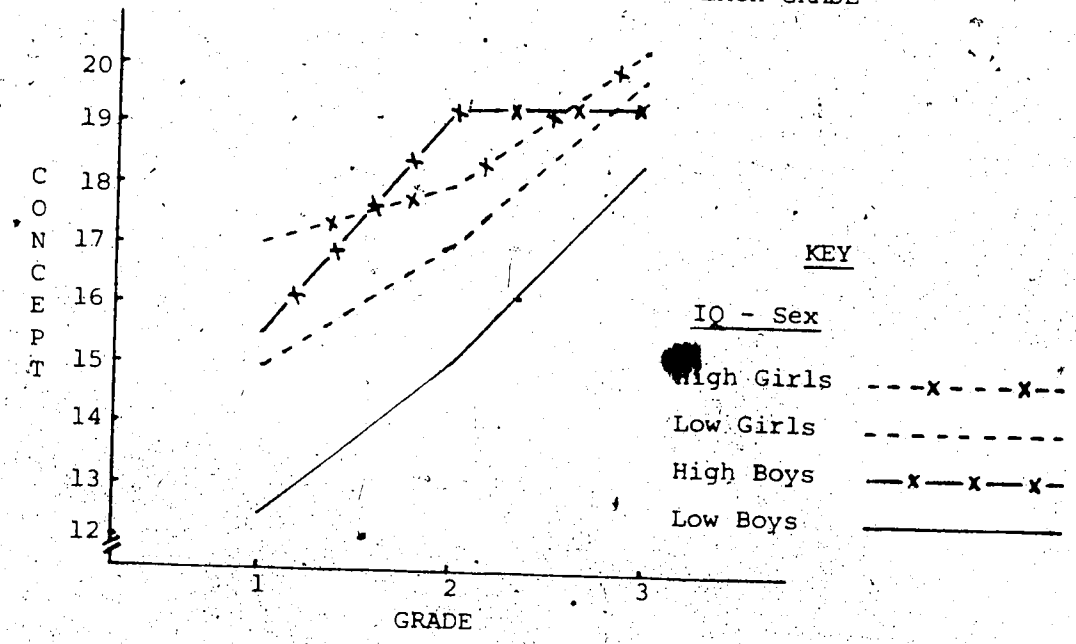
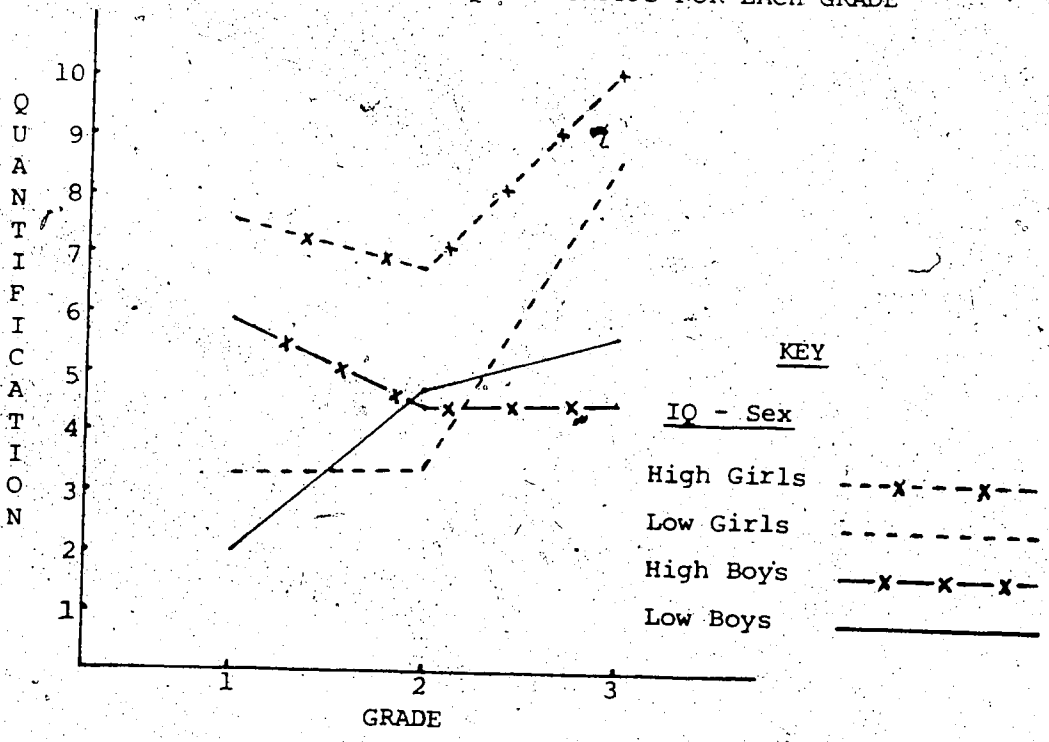


FIGURE 4

QUANTIFICATION SCORES BY IQ-SEX GROUPS FOR EACH GRADE



V. RATIONALIZATION USED BY SUBJECTS

In addition to the foregoing questions and hypotheses, a subsequent question, arising from the third purpose of the probability test, was asked:

What rationalizations are given by subjects for their responses to probability questions? Are there any observable trends related to grade level?

As indicated in chapter 3, subjects were found to rationalize their responses in one of three ways: (a) by correct reasoning, (b) by reference to color preference, position, or quasi-magical properties, or (c) by indicating they had no reason or didn't know.

These categories applied only to responses related to concepts. When asked for reasons for their answers to the quantification questions, only two subjects attempted to explain their answers. One was a second-grade girl who attended closely to the relative frequencies of the favorable color. The other was a third-grade boy who was reluctant to make predictions about events which he perceived to occur with random and uncertain frequency. For example, when asked to predict the number of times that the 2-3-1 block would turn up red out of six rolls he replied "You couldn't tell really, as it could be red every time in one lot of throws and not at all in another lot". When pressed for a commitment he said "Three times" and gave as his reason "Three faces are red". From these and other comments, the experimenter judged this subject to be well in advance of his peers in

understanding probability concepts.

The main analysis relating to the subsequent question, then, concerns rationalizations for concept responses. Table 15 contains frequencies of the types of rationalizations given for responses to items about all concepts except the first one, events in a sample space. As outlined in chapter 3, subjects were not required to explain the basis for their responses to items one, two, and three of the test.

In a number of cases a subject gave a correct rationalization even though his actual response may have been incorrect. This was particularly so with items relating to the fifth concept and accounts for an apparent discrepancy between the percentages given for this concept in Tables 6 and 15.

A number of trends are apparent. For four of the five concepts an acceptable rationalization was given more than 65% of the time with the whole sample. Except for the fifth concept, grade three subjects rationalized correctly more than 95% of the time. Color preference, position, and quasi-magical reasons seemed to be evident only in grade one and two subjects (with one exception) and then only regarding concept two, to any marked extent. Concept five proved to be the most difficult for all students a[illegible]alize their responses.

Examples of Rationalization

The following are some examples of rationalizations used by the subjects. These are grouped according to the concepts for which they were given and are referred to as types (a) or (b).

TABLE 15

FREQUENCIES OF RATIONALIZATION RESPONSES FOR CONCEPTS BY GRADES AND FOR THE TOTAL SAMPLE

Concept	Type of reason given [1]	Grade (N=24)			Total sample (N=72)	
		1	2	3	n	%
2. Most favorable event	a	12	13	23	48	67
	b	10	8	1	19	26
	c	2	3	0	5	7
3. Most favorable sample space	a	18	23	24	65	90
	b	2	1	0	3	4
	c	4	0	0	4	6
4. Equally favorable sample space	a	14	15	24	53	74
	b	1	1	0	2	3
	c	9	8	0	17	23
5. Impossible event	a	10	12	18	40	56
	b	0	0	0	0	0
	c	14	12	6	32	44
6. Certain event	a	19	20	23	62	86
	b	1	0	0	1	1
	c	4	4	1	9	13

1. a - correct; b - incorrect; c - none or didn't know

Most favorable event.

(a) Subjects most often responded that there was more of the selected color, or they counted segments, faces, or counters and concluded that there were numerically more of them of that color than other colors.

(b) Those responses categorized as type (b) included the following reasons, each of which indicates what the subject appeared to be attending to: "I like red; It's my favorite color", "Blue will win because it is my second favorite color", and "It's my (or the) best color".

Other reasons which were not related to a color preference included "It's the first one" (from a subject who chose the device nearest to him), "It's faster" (from one subject who attributed this trait to the color red), and "It's lighter" (from another subject who thus explained why the yellow side of the block would occur most often).

Most favorable sample space.

(a) Subjects tended to either give the correct rationalization for this concept or give no reason at all. Examples were: "Each color has two", "There are two of each", and "The colors are all even (or equal)". One subject incorrectly selected the 2-3-1 device and reasoned "It has all the colors". He appeared not to see the importance of the relative frequencies of the possible events.

Impossible event.

(a) An example of correct reasoning was "There's no yellow, it (the yellow marker) wouldn't get to move at all,"

(b) Examples of false reasoning were "I don't like yellow, it always loses" and "Blue couldn't win (selecting the 1-2-3 device) as there's only one blue". In one case a subject selected the 4-1-1 device and explained "One color can win and two can lose".

Certain event.

(a) Most subjects who gave reasons for their responses to items related to this concept were correct in their rationalizations, and also in their responses. A typical response was "It's all red, so only red would go. It would always win".

(b) The one erroneous reason given was related to color preference which caused the particular subject to select the 1-4-1 device and explain "Red would always win because it's my favorite color".

Concluding Statement

Although there is often considerable interest in "erroneous" rationalizations used by children, and indeed much can be learned from them, the researcher was impressed by the quality and the frequency of subjects' correct rationalizations on most of the concept items in the probability test. On the average, 75% of rationalizations were correct in the whole sample, while over 93% of the grade three responses were correct.

VI. SUMMARY

Chapter IV contains the results of answering three questions and testing seven hypotheses associated with the four major purposes of the present study.

The first purpose of the study was to determine what percentage of students appeared to have an understanding of the six basic probability concepts. It was found that over 74% of subjects in the whole sample showed an understanding of five of the six concepts. Over 92% of the grade three subjects showed such understanding and there was a general improvement in performance evident as age increased. The one concept that received fewest correct responses in all grades was number five, impossible event. The next poorest response was on concept number four, equally favorable sample space, which was correctly identified by 74% of all the subjects, but only by 58% of the grade one children.

When the rate of response was compared to what might be expected from purely random responses, it was found that the responses on all six concepts were significantly better than chance at the 0.01 level of confidence.

The second purpose of the study dealt only with quantification of probability. Scores were much lower than on the concept items; an average of only 42% of subjects responded correctly on the quantification items. Grade three subjects performed significantly better than grade one or two subjects. No difference was found between grade one and two scores on either subtest.

There was no significant difference between scores on three different probability settings (2-2-2, 3-2-1, and 4-1-1) but as the number of trials in the sample space increased the number of correct predictions decreased.

The third purpose of the study was to examine the effect of embodiment on pupils' responses to probability questions. No significant main effects were found due to embodiment at any of the grade levels. No interactions were evident between embodiment and sex or IQ but judgment was reserved in the case of interaction between embodiment and grade.

The fourth purpose of the study was to investigate the effects of sex, grade, and IQ on the criterion scores. Significant effects were found, at the 0.01 level of confidence, for all three factors. Girls were found to score higher than boys, grade three subjects scored higher than those in grades one and two, and the high-IQ group scored higher than the low-IQ group. There were no interactions between sex, grade, and IQ on probability test achievement.

The fifth discussion in chapter 4 related to the rationalizations used by subjects to explain their responses. Only 3% of subjects offered reasons for quantification responses while over 85% of subjects responded on four of the six concepts, and 77% and 56% on the other two. Of the rationalizations given, an average of 75% were correct across the whole sample, while 93% of the grade three rationalizations were correct.

In the next and final chapter, the study is summarized in terms of its purposes, the instrumentation and procedures used,

and the analysis of the results. Implications for teaching, curriculum design, and further research are made in conclusion.

CHAPTER 5

SUMMARY, DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS

I. SUMMARY OF THE INVESTIGATION

The present study developed from a need for more information about young children's understanding of basic probability notions. It is widely accepted that the elementary school curriculum should include topics on probability in order to give children early experience in dealing with degrees of uncertainty and to prepare them for later studies in statistics. According to Ausubel (1968), meaningful learning experiences can only be constructed on the basis of the learner's existing knowledge and understanding. The first task for a curriculum writer or a teacher is to ascertain the level of readiness a learner has for a topic.

The main purpose of the present study was to investigate young children's readiness for instruction in probability by determining how well six basic concepts were understood by children in grades one, two, and three. A sample of 72 pupils was chosen, 24 in each grade, and a test was administered to each one in an interview with the researcher. A game was employed in the test items as a motivating agent and to encourage subjects to maximize their responses. There were 21 questions dealing with the six concepts. In addition 18 quantification questions were added and subjects were asked to give

rationalizations for each response. The materials and apparatus used in the study were specially constructed so as to represent each of five probability settings in three embodiments, spinner, bl[redacted] and box. This allowed for testing for the effect of embodiment on subjects' responses.

A test-retest reliability check on the probability test gave Pearson product-moment correlation coefficients of 0.837, 0.796, and 0.753 for total and subtest scores; all were significantly different from zero at the 0.01 level.

Analysis of the results was organized into five sections corresponding to the four purposes of the study and a report on the rationalizations used by subjects. Three questions were asked and seven hypotheses were tested. A brief summary of the main findings is now given under five headings.

Concepts

Four of the six concepts tested were understood by at least 75% of the subjects in each grade. They were: sample space events, most favorable event, most favor e sample space, and certain event. The other two concepts, equally favorable sample space and impossible event, were understood by 74% and 50% respectively of the total sample. The scores on all concept items were significantly greater than was attributable to chance.

Quantification

The quantification of probability proved to be less understood than the concepts. The average correct response rate for all subjects on the 18 quantification items was 42%.

There was no significant difference in scores due to different probability settings but performance decreased as the number of trials in the sample space increased from six to twelve.

Embodiment

No significant effect was found due to embodiment at any grade level, and no interactions were evident between embodiment and sex or IQ.

Significant Factors

Sex, grade, and IQ were all found to have significant effects on the total test score. Girls scored higher than boys, grade three subjects scored higher than those in grades one and two, and the high-IQ group scored higher than the low-IQ group. Only grade and IQ had significant effects on the concept score and no significant effects were found on the quantification score. No interactions were found between sex, grade, and IQ on any of the criterion measures.

Rationalization

When subjects were asked for their answers to the test questions, one of three types of responses was given: (a) a correct explanation based on the proportions within the probability settings in question, (b) an erroneous explanation involving the subject's color preference, the position of the devices, or quasi-magical properties seen in the situation, or (c) no answer, or an indication of having no reason or of not knowing.

Considering all of the concept items, an average of 80% of subjects rationalized with type (a) or type (b) responses. Of these, 75% were correct for the whole sample and 93% were correct for grade three. Samples of rationalizations given for the responses for each concept were included with the report in chapter 4. Only two subjects gave rationalizations for quantification responses. Most gave no reason for their predictions on these items.

II. DISCUSSION OF THE FINDINGS

Concepts

The first concept, events in a sample space, presented little problem for any of the children. The researcher found that most of the subjects were almost puzzled that such an "obvious" question be asked as, for example, "What color could I get if I spin this spinner?" Once the meaning of the question was grasped (and that sometimes proved to be the main problem) almost all subjects readily supplied the correct answer. The children in the study were able to recognize all the possibilities in a situation when they understood that this was what was required of them.

The second concept, the most favorable event, was understood by 83% of all the subjects and by 98% of the third graders. Most of the children simply selected from the three colors present the color that was most plentiful. This meant counting the number of sectors on the spinner, the faces on the block, or the counters

in the box that were of each color. Eleven of the twelve subjects who responded incorrectly were in grades one and two and selected their favorite color or the closest one to them, or chose on the basis of imputed advantage of one color over the others.

The third concept, most favorable sample space, was understood by almost all the second and third graders and 75% of first graders. In the items relating to this concept subjects were presented with three probability settings, for example, 2-2-2, 3-2-1, and 4-1-1. It appeared to be relatively easy for most subjects to compare these settings and to choose the one which maximized the chance of a particular color being the outcome, 4-1-1 and BLUE in the example. It made little difference whether the settings were embodied by spinners, blocks, or boxes. The main rationalization given by subjects concerned the relative amount of the favored color. With each embodiment subjects were able to count discrete units of each color, make comparisons of the amount of color, and make their choice on that basis. In each case their reasoning was correct, there being an equal number of units (six) in each setting and device. Subjects in grade three often verbalized their reasoning in clearer terms than the researcher had expected.

The fourth concept, equally favorable sample space, was the second most difficult for the subjects with an overall correct response rate of 74%. Third-grade subjects (92% correct) had little difficulty selecting the correct 2-2-2 setting, but many of the grade one and two children baulked at the three 3-way

comparisons that confronted them. For some of these children the exact meaning of "fair", "same chance", or "equal chance" may not have been clear when there were three possible outcomes. This was indicated by several subjects selecting two of the devices which had the same amount of one particular color. These responses serve to emphasize the need for extreme care when using verbal instructions or questions with 6- and 7-year olds. The same problem was not encountered with the third-grade subjects.

The fifth concept, impossible event, was by far the most difficult for all subjects. The grade one response was no better than a chance response and only 67% of grade three subjects responded correctly. The most common incorrect response was the selection of the 3-2-1 or the 4-1-1 setting (or equivalent) and identification of the one-unit color as the impossible outcome. To ensure that the subjects giving these responses were not just confusing "impossible" with "unlikely", the investigator always repeated the question with emphasis on "never win" and always lose". For more than half of the grade one and two subjects it made little difference; the two words, impossible and unlikely, appeared to suggest the same idea. A third of the grade three pupils also saw no differences.

Several subjects selected the 6-0-0 setting in their response and nominated as impossible outcomes the colors with no representation in the setting. The investigator accepted such a response as indicative of an understanding of the concept in

question. This judgment was generally confirmed by later asking the subject what would happen in the game if the 3-3-0 device were used. Except for one case (no response), the subjects responded in terms of one event never occurring or one color (or marker) never winning.

The final concept, certain event, was correctly understood by 85% of all subjects, and by 96% of the third graders. The responses were usually given quickly and subjects often reacted as though the questions relating to this concept were "obvious", as with the first concept. Few of the subjects had difficulty understanding the meaning of the questions. The incorrect responses generally came from some of the same subjects who were wrong on the concept five questions, and for similar reasons. For example, the 3-2-1 or 4-1-1 setting was selected and BLUE identified as the certain outcome. The ideas of certainty and likelihood were apparently being equated by these students; this tendency was not shown by any of the third grade subjects.

Quantification

Subjects' responses on the quantification questions indicated a low level of understanding of the relationship between the proportions in the settings and the chances of the various outcomes. The numbers of correct responses made by grades one and two were not significantly better than chance. The significantly higher scores by grade three subjects, even though still only 50% were correct, suggest that grade three is the earliest that such quantification ideas should be introduced.

There was little difference in the number of correct responses due to the probability setting, but doubling the number of trials (from six to twelve) resulted in more than a 50% drop in correct responses.

Embodiment

The probability test was arranged so that the subjects answered thirteen questions in each of the three embodiments, spinner, block, and box. No significant differences at the 0.01 level were found between the mean scores for these embodiments for any of the grades. This was decided using the conservative probability of F in a univariate analysis with embodiment as a repeated measure. Using a three-way analysis with sex, grade, and IQ as blocking variables in pairs, no significant interactions were found between embodiment and sex or IQ. Judgment was reserved in the case of interaction with grade as it was significant when blocked with IQ, but not significant when blocked with sex. (Both probabilities of F were close to 0.01.) This meant that subjects' responses to embodiment tended to vary according to their grade level. This variation is shown in Table 16 by the rank of embodiment responses within grades. The grade one subjects responded correctly to the box embodiment 17% more often than to the block, and 13% more often than to the spinner. The grade three subjects, on the other hand, responded correctly to the spinner embodiment 9% more often than to the block, and 8% more often than to the box embodiment. The grade two means were in the order box, block, and spinner within a 6% range.

TABLE 16

RANK OF EMBODIMENT RESPONSES WITHIN GRADES

Embodiment	Grade 1	Grade 2	Grade 3
Spinner	2	3	1
Block	3	2	3
Box	1	1	2

A possible reason for the grade one preference for the box embodiment is the easier task of counting discrete objects rather than proportions on a disc or faces on a cube. Many of the early number activities at school involve counters; this would predispose younger children to this mode of counting. By third grade, subjects had gained greater ability to count and reason about non-discrete items that were fixed in their spatial relationship.

Significant Factors

The findings of the present study agree with those of earlier studies that grade (or age) and IQ are significant factors in responses to probability questions. There was a highly significant difference ($p<0.001$) between the grade three performance and grades one and two on the concept and total scores, but little difference between grades one and two on any of the criterion scores. This tends to indicate a substantial increase in understanding of probability concepts in children as they pass through grade three, about age 8 years. According

to Piaget's theory, this jump in understanding is the beginning of stage II development, proposed by Piaget to begin at around seven years.

In the present investigation, sex was a significant factor in the total test score but not in either of the subscores. Girls scored higher than boys on all criterion measures and within most IQ and grade groupings.

On the quantification subtest, none of the factors was significant at the 0.01 level. The subjects in all three grades, of both sexes, and in both IQ groups found the quantification questions uniformly difficult, although grade three scores were higher than grades one and two, as already mentioned.

Rationalizations

The main impression formed by the investigator as a result of the subjects' rationalizations is that most of the children in the sample correctly understood at least four of the basic concepts and could express quite adequately the basis of their responses to the probability concept questions. The absence of rationalizations for quantification responses indicated that many subjects had little understanding of the numerical relationships between the probability settings and the frequencies of the outcomes. Many responses appeared to be guesses, although third-grade subjects scored well on the six-trial questions and may have been at the threshold of understanding quantification in sample spaces with a small number of events.

III. SOME IMPLICATIONS OF THE FINDINGS

For Teachers

From the discussion in the previous section a number of classroom implications arise. The first relates to the main finding that at least four of the probability concepts were understood by the majority of the grade one, two, and three subjects. The writer would encourage teachers to provide experiences in probability in these grades (especially grade three), experiences which allow children to become involved in the real world activities of decision making. This is not too ambitious for young children, for their vocabulary, and performance in this study, indicate they already have begun to appreciate the uncertainty in many situations which they face. Having begun to accumulate ideas about chance events at a young age, children can easily form incorrect intuitions which are difficult to alter, as Wilkinson and Nelson (1966) found. As teachers, we need to include probability experiences in the lower grades that will produce correct intuitions and so build a better base for the more formal study of probability and statistics in later grades.

Further suggestions can be made from the writer's observations in this study:

1. Beginning activities should involve only sample spaces with small numbers of outcomes which are easily identified by the students as possible occurrences.

2. Comparison and choice responses should be used initially

in preference to prediction responses. The probability settings being compared should have the same number of units to allow comparisons on the basis of absolute number. Otherwise ratios with unequal denominators are involved; these are not handled with any substantial skill until later grades.

3. Care should be taken to minimize misunderstanding due to word meaning. The problem of communicating ideas accurately is ever present at all grade levels. As far as possible, non-verbal methods and materials that communicate ideas and command action or decision with little direction needed from the teacher should be used with young children.

For the Curriculum

The present study indicates that grade one, two, and three children have sufficient intuitive understanding of probability concepts to form a basis for the development of special instructional units appropriate to their grade level. Early activities in grade one and two need to provide informal consolidation of existing concepts and ideas. From grade three onwards, units should provide a wider range of experiences and activities in which students are led to further concepts and into quantification of probability.

IV. RECOMMENDATIONS FOR FURTHER RESEARCH

No attempt was made in the present study to test the feasibility of teaching probability topics to grades one, two, and three. The present study was designed mainly to survey the

field for understanding of six basic concepts. There are two main recommendations for further research: a replication of the study, with improvements, and the development and trial of appropriate instructional units in probability for primary grades.

One improvement in a replication study would be to use a larger sample chosen from a wider variety of schools. Changes could be made in the instrument to include the testing of a further concept, unlikely event, for comparison with the response to impossible event. Special attention would need to be given to the wording and presentation of these items. Another suggestion would be to use grade two, three, and four students as little difference was found between grades one and two. There may then be little difference between grades three and four as third-graders in the present study scored near the maximum on the concept subtest. Such a replication may, on the other hand, give further insight into quantification understanding around that age or grade level.

Many worthwhile studies have been done regarding instructional units and methods appropriate to the senior grades in the elementary school. The implications stated in section III of this chapter lead to a recommendation for comparable research and development at the junior grade level. Wilkinson and Nelson (1966) gave six practical suggestions to those who would design probability units for elementary school children. Shepler (1969) proposed a sequence for developing research-based curriculum materials using behavioral objectives and task analysis. Using

a combination of these guidelines (outlined in chapter 2 of this report), and taking the four concepts sample space, most favorable event, most favorable sample space, and certain event as a basis, a unit on probability should be designed and tested with junior grades.

Teachers-in-training need to be instructed in recent curriculum changes. Teacher-preparation programs and in-service courses need to include updated components related to the teaching of probability in elementary grades if any significant implementation of the recommended changes in curriculum is to occur.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ausubel, D. P. Educational Psychology: A Cognitive View. New York: Holt, Rinehart, & Winston, 1968.

Blishen, B. R., Jones, F. E., Naegele, K. D., & Porter, J. Canadian Society: Sociological Perspectives. Totonto: MacMillan, 1968.

Brackbill, Y., Kappy, M. S., & Starr, R. H. Magnitude of Reward and Probability Learning. Journal of Experimental Psychology, 1962, 63, 32-35.

Bruner, J. S. "On the Learning of Mathematics - A Process Orientation". In D. B. Aichele & R. E. Keys (Eds.), Readings in Secondary School Mathematics. Boston: Prindle, Weber, & Schmidt, 1974.

Cambridge Conference on School Mathematics. Goals for School Mathematics. N. Y.: Houghton Mifflin, 1963.

Carlson, J. S. Children's Probability Judgments as Related to Age, Intelligence, Socio-Economic Level and Sex. Human Development, 1969, 12(3), 192-203.

Cohen, J. Subjective Probability. Scientific American, Nov. 1957, 197, 128-138.

Davies, C. M. Development of the Probability Concept in Children. Child Development, 1965, 36, 779-788.

Doherty, Joan. Level of Four Concepts of Probability Possessed By Children of Fourth, Fifth, and Sixth Grade before Formal Instruction (Doctoral Dissertation, University of Missouri, 1965). Dissertation Abstracts, Dec. 1966, 27, 1703A. (University Microfilms No. 65-14, 465).

Engel, Arthur. Mathematical Research and Instruction in Probability Theory, The Mathematics Teacher, Dec. 1966, 771-782.

Engel, Arthur. Teaching Probability in the Intermediate Grades. International Journal of Mathematical Education in Science and Technology, 2, July/September, 1971, 243-294.

Estes, W. K. Research and Theory on the Learning of Probabilities. Journal of the American Statistical Association, 1972, 67, 81-102.

Estes, W. K. The Cognitive Side of Probability Learning, Psychological Review, 1976, 83(1), 37-64. (a)

Estes, W. K. Some Functions of Memory in Probability Learning and Choice Behavior. In G. H. Bower (ed.), The Psychology of Learning and Motivation (Vol. 10). New York: Academic Press, 1976. (b)

Ferguson, G. A. Statistical Analysis in Psychology and Education, 3rd Edition. N. Y.: McGraw-Hill, Inc., 1971.

Fischbein, E. The Intuitive Sources of Probabilistic Thinking in Children. Dordrecht, Holland: D. Reidel Pub. Co., 1975.

Fischbein, E. Probabilistic Thinking in Children and Adolescents. In Forschung zum Prozeb des Mathematiklernens (Addresses Given at the Third International Congress on Mathematics Education at Karlsruke). Universitat Bielsfeld, 1976, 23-42.

Fischbein, E., Pampu, I., & Manzat, I. Comparison of Ratios and the Chance Concept in Children. Child Development, June 1970, 41, 377-389.

Flavell, J. H. The Developmental Psychology of Jean Piaget. New York: D. Van Nostrand Co. Inc., 1963.

Gipson, J. Teaching Probability in the Elementary School: An Exploratory Study (Doctoral Dissertation, University of Illinois, 1971). Dissertation Abstracts International, Feb. 1972, 32, 4325A-4326A. (University Microfilms No. 72-6933, 374).

Glass, G. V., & Stanley, J. C. Statistical Methods in Education and Psychology. Englewood Cliffs, N. J.: Prentice-Hall, 1970.

Goldberg, S. Probability Judgements by Pre School Children: Task Conditions and Performance. Child Development, March 1966, 37, 157-167.

Harvey, J. G. An Immodest Proposal for Teaching Statistics and Probability. School Science and Mathematics, March 1972, 129-144.

Hoemann, H. W. & Ross, B. M. Children's Understanding of Probability Concepts. Child Development, 1971, 42, 221-236.

Jones, G. A. The performances of First, Second, and Third Grade Children on Five Concepts of Probability and the Effects of Grade, IQ, and Embodiments on Their Performance (Doctoral Dissertation, University of Indiana, 1974). Dissertation Abstracts International, Jan. 1975, 35, 4272A-4273A. (University Microfilms No. 75-1710, 387).

Leffin, W. W. A Study of Three Concepts of Probability Possessed by Children in the Fourth, Fifth, Sixth, and Seventh Grades (Doctoral Dissertation, University of Wisconsin, 1968). Dissertation Abstracts, June 1969, 29, 4188A-4189A. University Microfilms No. 68-17, 913).

Lovell, K. proportionality and Probability. In NCTM, Piagetian Cognitive-Development Research and Mathematical Education. Washington, D. C.: NCTM, 1971.

McLeod, G. K. An Experiment in the Teaching of Selected Concepts Of probability to Elementary School Children (Doctoral Dissertation, Stanford University, 1971). Dissertation Abstracts International, Sept. 1971, 32, 1359A. (University Microfilms No. 71-23, 535, 231).

Messick, S. J., & Solley, C. M. Probability Learning in Children: Some Exploratory Studies. Journal of Genetic Psychology, 1957, 90, 23-32.

Mullenex, J. L. A Study of the Understanding of Probability Concepts by Selected Elementary School Children (Doctoral Dissertation, University of Virginia, 1968). Dissertation Abstracts, May 1969, 29, 3920A. (University Microfilms No. 69-4010, 89).

National Advisory Committee on Mathematical Education. A Technical Report: Overview and Analysis of School Mathematics Grade K - 12. Washington, D. C.: Conference Board of the Mathematical Sciences, 1975.

Offenbach, Stuart I. Studies of Children's Probability Learning Behavior: I. Effect of Reward and Punishment at Two Age Levels. Child Development, 1964, 35, 709-715.

Offenbach, Stuart I. Studies of Children's Probability Learning Behavior: II. Effect of Method and Event Frequency at Two Age Levels. Child Development, 1965, 36, 951-962.

Ojemann, R. H., Maxey, J. E., & Snider, B. C. The Effect of a Program of Guided Learning Experiences in Developing Probability Concepts at the Third-Grade Level. Journal of Experimental Education, Summer 1965, 33(4), 321-330.

Ojemann, R. H., Maxey, J. E., & Snider, B. C. Further Study of Guided Learning Experiences in Developing Probability Concepts in Grade Three. Perceptual and Motor Skills, 1966, 23, 97-98.

Piaget, Jean. The Psychology of Intelligence. Patterson, N. J.: Littlefield, Adams & Co., 1960.

Piaget, J. & Inhelder, B. The Origin of the Idea of Chance in Children. New York: W. W. Norton & Co., Inc., 1975.

Racha-Intra, S. Basic Inferential Statistics in Grade Nine. Unpublished Doctoral Dissertation, University of Alberta, Edmonton, 1977.

Restle, F. Psychology of Judgment and Choice: A Theoretical Essay. New York: Wiley, 1961.

Romberg, T. A. & Shepler, J. Retention of Probability Concepts: A Pilot Study into the Effects of Mastery Learning with Sixth-Grade Students. Journal of Research into Mathematics Education, Jan. 1973, 4(1), 27-32.

S.M.S.G. Probability for Primary Grades, Teacher's Commentary. Pasadena, Ca.: A. C. Vroman, 1966.

Shepler, J. L. Parts of a Systems Approach to the Development of a Unit in Probability and Statistics for the Elementary School. Journal for Research in Mathematics Education, Nov. 1970, 1(5), 197-205.

Siegel, S. & Andrews, J. M. Magnitude of Reinforcement and Choice Behavior in Children. Journal of Experimental Psychology, 1962, 63, 337-341.

Stevenson, H. W. & Weir, M. W. Variables Affecting Children's Performance in a Probability Learning Task. Journal of Experimental Psychology, 1959, 57, 403-412.

Stevenson, H. W. & Weir, M. W. The Role of Age and Verbalization in Probability Learning. American Journal of Psychology, 1963, 76, 299-305.

Stevenson, H. W., & Zigler, E. F. Probability Learning in Children. Journal of Experimental Psychology, 1956, 56, 185-192.

Strohner, H. & Nelson, K. E. The Young Child's Development of Sentence Comprehension, Influence of Event Probability, Nonverbal Context, Syntactic Form and Strategy. Child Development, 1974, 45, 567-576.

Thorndike, R. L., Hagen, E., Lorge, I., & Wright, E. N. Canadian Cognitive Abilities Test, P2/F1; Examiner's Manual. Toronto: Thomas Nelson, 1970.

U. S. Army, Ordnance Corps. Tables of the Cumulative Binomial Probabilities. Washington: 1952.

Wilkinson, J. D. & Nelson, O. Probability and Statistics - Trial Teaching in Sixth Grade. Arithmetic Teacher, Feb. 1966, 13, 100-106.

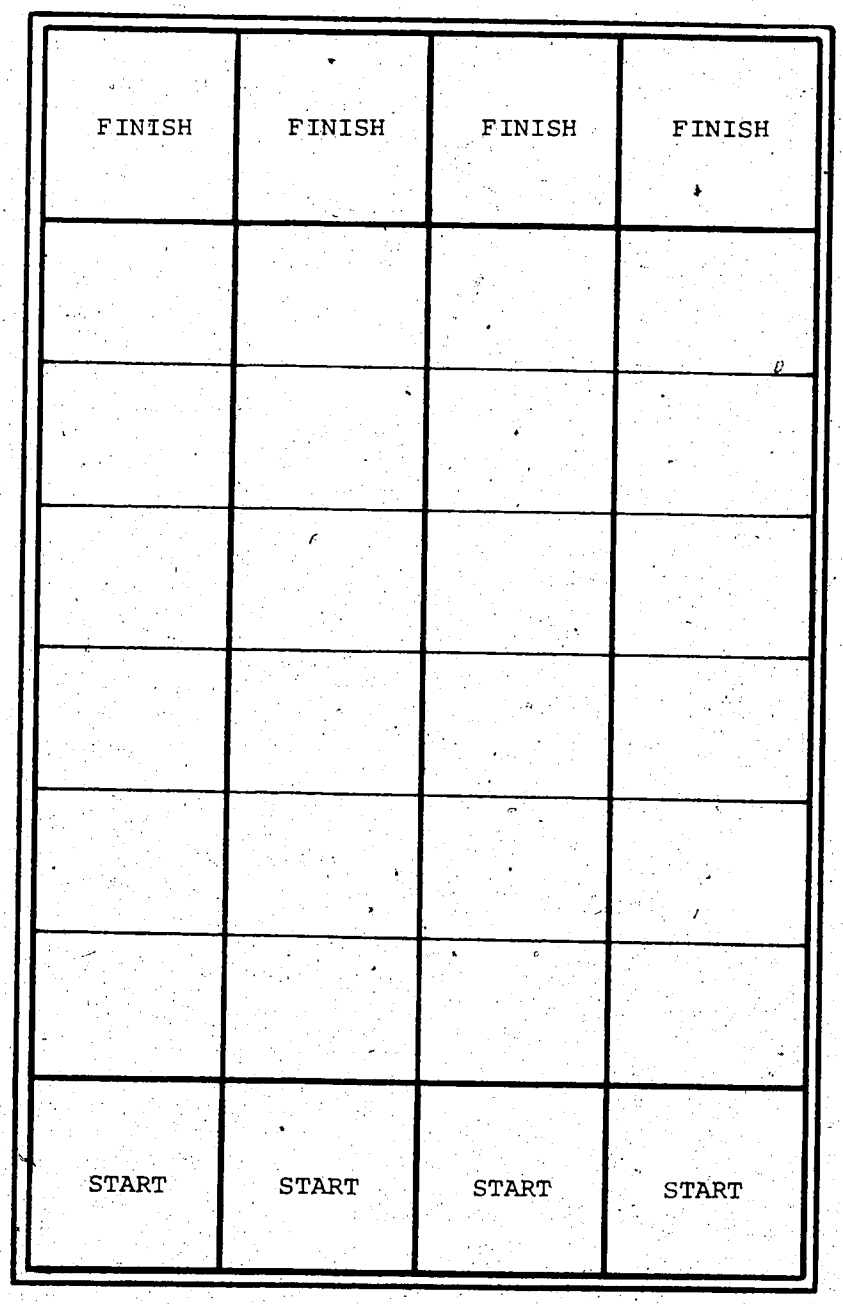
Winer, B. J. Statistical Principles in Experimental Design. N. Y.: McGraw-Hill, 1971.

Yee, Albert H. Mathematics, Probability and Decision-Making. The Arithmetic Teacher, May 1966, 13, 385-387.

Yost, P. A., Siegel, A. E. & Andrews, J. McM. Non Verbal Probability Judgments by Young Children. Child Development, 1962, 33, 769-780.

APPENDIX A

GAME BOARD USED IN PROBABILITY TEST


FINISH
FINISH
FINISH
FINISH
START
START
START
START

APPENDIX B

TABLES OF FREQUENCIES AND RELATIVE FREQUENCIES OF QUANTIFICATION RESPONSES

TABLE 17

FREQUENCIES AND RELATIVE FREQUENCIES OF RESPONSES ON THE SIX-TRIAL QUANTIFICATION QUESTIONS

Subject's response	Setting					
	2-2-2		3-2-1		4-1-1	
	f	rf	f	rf	f	rf
1	9	0.044	8	0.039	3	0.014
2	94*	0.461	29	0.142	9	0.043
3	40	0.196	93*	0.456	44	0.210
4	26	0.127	25	0.123	90*	0.429
5	13	0.064	19	0.093	34	0.162
6	22	0.108	30	0.147	30	0.143

* correct responses (also modal responses in these cases).

TABLE 18

FREQUENCIES AND RELATIVE FREQUENCIES OF RESPONSES ON THE TWELVE-TRIAL QUANTIFICATION QUESTIONS

Subject's response	Setting					
	2-2-2		3-2-1		4-1-1	
	f	rf	f	rf	f	rf
1	6	0.028	6	0.029	3	0.014
2	34	0.156	7	0.034	5	0.024
3	14	0.064	40	0.196	8	0.038
4	31*	0.142	8	0.039	38	0.181
5	18	0.083	13	0.064	8	0.038
6	59#	0.271	60*#	0.294	55#	0.262
7	7	0.032	12	0.059	14	0.067
8	13	0.060	14	0.069	29*	0.138
9	5	0.023	7	0.034	7	0.033
10	9	0.044	11	0.054	12	0.057
11	5	0.023	2	0.010	9	0.043
12	17	0.078	24	0.118	22	0.105

* correct responses.

\# modal responses.

40529

National Library of Canada / Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE / THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* DENNIS M. STANG

TITLE OF THESIS/*TITRE DE LA THÈSE* THE ECONOMICS OF REGULATION OF THE PAY-TV INDUSTRY

UNIVERSITY/*UNIVERSITÉ* University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/*GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* Master of Arts

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1978

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* W. C. Riddell

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* Aug 16 78 SIGNED/*SIGNÉ* Dennis Stang

PERMANENT ADDRESS/*RÉSIDENCE FIXÉ* 12120-86 St.
Edmonton, Alberta
T5B 3K7

NL-91 (3-74)

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

THE UNIVERSITY OF ALBERTA

THE ECONOMICS OF REGULATION OF THE PAY-TV INDUSTRY

by

DENNIS MICHAEL STANG

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF ARTS

DEPARTMENT OF ECONOMICS

EDMONTON, ALBERTA

FALL, 1978

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled THE ECONOMICS OF REGULATION OF THE PAY-TV INDUSTRY submitted by DENNIS MICHAEL STANG in partial fulfilment of the requirements for the degree of Master of ARTS

.....

Supervisor

.....

.....

Date August 4, 1978

ABSTRACT

This thesis examines the effects of different institutional arrangements and supply structures in the television industry on the use and allocation of scarce resources. In particular, economic analysis is applied to some of the regulatory issues involving pay television. Pay-TV is a new industry in Canada, and much confusion currently exists concerning its potential economic performance and the type of governmental regulation it should be subjected to.

Particular attention is paid to the issue of optimal variety in the television programming market. This includes an analysis of the ability of pay-TV to offer consumers a programming mix that more closely corresponds to their preferences than the programming presently offered by the advertiser-supported television system. Current television program uniformity and the reasons behind the lack of diversity is also evaluated.

This thesis also concerns itself with defining a set of conditions in an industry which might warrant government regulation from an economic viewpoint. Because regulation has become so prolific in our economy, it is very important to develop a general theoretical framework in which to evaluate its rationale. Justification for regulation will be based on the ability of the regulatory process to improve resource efficiency in an industry. This framework for regulatory justification is then applied to the pay-TV industry in Canada to determine what type of regulation the pay-TV industry should be subjected to.

It is discovered that although government regulation cannot be expected to significantly improve the economic performance of the pay-TV industry, regulation is politically inevitable because of the potential effects of pay-TV on the current broadcasting system, consumer equity, and some of the cultural aspects of our society.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Regulation can be defined as "the codification of modes of behavior for individual agents directed at maintenance of the social system". In other words, regulation places both incentives and restrictions on decision-making units aimed at inducing these units to perform in a desirable manner. Regulation is thus aimed at controlling or directing behavior.

The role of economic theory, on the other hand, is to identify the important determinants of economic behavior, so that hypotheses can be formulated concerning their relationships and impact on the various aspects of performance. Thus, although regulation of industry is a social phenomena, economics has much to offer in the formation and evaluation of regulatory policies.

More specifically, economic theory can be utilized in analyzing the effects of regulation on the use and allocation of scarce resources, on the institutional structure of the economy, and on the individual agent in the economy. Economic theory also provides principles and behavioral rules that, if followed, will produce optimum results (using widely accepted criteria of optimality).

The purpose of this thesis is to apply economic analysis to some of the issues involved in governmental regulation of the pay-TV industry in Canada. Because little economic literature has been generated on pay-TV, much of this thesis involves an identification of the major issues and a delineation of the relevant variables involved - a necessary step before a rigorous economic analysis of pay-TV is possible. Particular

attention will be paid to the issue of optimal variety in the television programming market, as this appears to be a crucial issue in the debate over pay-TV.

Because this analysis is economic in nature (it is essentially impervious to many of the political, legal, and administrative considerations which normally form an integral part of any regulatory policy), it in itself is not intended to lead to the operationally ideal regulatory policy. Rather, it serves to isolate the areas in which economic analysis can be useful in determining the optimal regulatory policy stance.

The prescription of a purely economic ideal solution designed to achieve the desirable economic performance goals will also allow for a measurement of the costs of deviating from this ideal to achieve the non-economic performance goals. In other words, it can serve as a benchmark in assessing the economic costs of achieving any political objectives when it is compared to the actual regulatory policy adopted.

Some may question the appropriateness of applying the type of economic analysis presented in this thesis to the television industry. The economic theory presented is largely based in the context of the efficiency of the market system, and some may argue that the concept of television broadcasting in Canada, as established by various pieces of legislation and judicial decisions, is one of public service designed to meet political objectives, and not as an outcome of the market place. However, a television programming market certainly does exist, with both supply and demand components, and political objectives can simply be viewed as including those goals which the market fails to attain thus requiring government manipulation of the market to allow for their achievement. The television industry, and in particular pay-TV, lends itself quite readily to economic analysis in a market context.

The first question which should be answered in analyzing any type of regulation is: should the industry be regulated in the first place? The justifications for government regulation, if any, will then determine the form which regulation should take. The author suspects that sometimes government officials do not sufficiently consider the justifications for government intervention in a market, but begin by first asking: how should the industry be regulated?

Within this paper I will attempt to develop a framework for the economic justifications of government regulation. Although justifications for regulation are politically determined, there are certain conditions which can justify government intervention in a market solely on economic grounds. The pervasiveness and importance of regulation in our economy justifies a rigorous economic analysis of its rationale.

The pay-TV industry was chosen for this thesis because at the commencement of writing regulation of pay-TV was a timely issue in Canada. The unfortunate aspect of the pay-TV industry from a research point of view is that its limited experience as an industry forces research to sometimes rely on speculation and inference, a procedure which may lead to error in analysis. In addition, the lack of sufficient empirical data makes a rigorous analysis of the industry and prediction of potential economic performance very difficult.

There can be little doubt that pay-TV is potentially very important to our society. Although pay-TV has only been implemented in a relatively few cases to date, there is a very good chance that someday pay-TV, in the absence of restrictive regulation, will offer subscribers in most of our large metropolitan markets a diversity of services. In the words of

Canada's Minister of Communications, Jeanne Sauve, "the establishment of pay television service on a large scale is inevitable."[1]

Pay-TV can be expected to one day supplement the current advertiser-supported television system, and any type of television media is important because of its prominent effects on the political, social, and economic aspects of our society. The media plays an important role in our society in disseminating ideas and information. Almost every Canadian home has a television set,[2] and the average person views several hours of television daily.[3] Television has a pervasive impact on these people.

The Pay-TV Industry

In this first section the pay-TV industry will be defined in terms of its operations, history, potential growth and desirable characteristics. Before a regulatory policy can be developed, a knowledge of the technical and institutional environment of the industry is necessary. The political reaction to pay-TV in Canada will also be discussed.

---

[1] Canada, Dept. of Communications, "Pay Television", Notes for a Speech: By the Honourable Jeanne Sauve, (Ottawa: 1976), p. 2.

[2] Approximately 97 per cent of Canadian Homes have television sets as reported in the 1975 - 1976 Annual Report, Dept. of Communications (Ottawa: Supply and Services Canada, 1976).

[3] Daily television viewing per household in Canada averages about 6 hours 11 minutes. See E.S. Hallman, Broadcasting in Canada, Case Studies on Broadcasting Systems (London: Routledge and Kegan Paul, 1977), p. 31.

Definition of Pay-TV

Per-program or per-channel charge

Pay television, as its name suggests, is a system in which viewers pay directly for either each individual television program shown or for access to additional channels. Thus, pay-TV shifts part of the financing of television from advertisers to consumers and "introduces the box office concept into television."[4] Pay-TV provides additional or specialized programming to viewers who are willing to pay for it.

It appears that initially a per-channel charge will be necessary in Canada in order to generate sufficient revenues to make a pay-TV system financially viable. However, a per-program charge may eventually evolve to allow the delivery of programs that cater to the tastes of small viewer groups.[5]

How pay-TV works

Pay-TV is normally delivered (or "piggy-backed" as they say in the industry) over cable, although it can be implemented in an over-the-air system or even transmitted via satellite (the satellite is the ultimate in delivery systems; reception by a household in such a case could be through a disc-shaped sphere attached to the roof and connected to the television set). Pay-TV is most likely to operate through cable for a number of reasons.

---

[4] Roger G. Noll, Merton J. Peck, and John J. McGowan, Economic Aspects of Television Regulation (Washington, D.C.: The Brookings Institution, 1973), p. 129.

[5] Recently, both Bell Canada and Western Coded Television Ltd. have separately developed a pay-TV system which would allow individual program payments by consumers instead of flat monthly charges. See "Royal Commission may delay pay-TV decision," Edmonton Journal, 18 June, 1977.

The term cable actually embraces three distinct forms of subscriber-supported communications service. Two of the categories form the cable industry as it exists today. The original CATV (community antenna television) brings improved television service to areas that experience poor reception. The second form not only strengthens existing signals, but also imports additional signals into an area or even originates programs of its own, thus providing a greater variety in programming. This service may involve a periodic billing to the customer, or a per-program or per-channel payment (pay-TV). As such, these two forms are primarily an extension of over-the-air broadcasting.

The third form of cable has the potential to become a comprehensive communications medium in its own right, offering a wealth of entertainment and information services. As many as a forty or eighty-channel system could not only provide a greater diversity of broadcasting service with improved signal quality, but it could also offer a wide array of non-broadcast entertainment and information services, including two-way communications.

Cable-delivery has several advantages for pay television, expecially when compared to the most likely alternative which is over-the-air pay-TV delivery. Most importantly, cable delivery takes advantage of sales and service economies that will make possible rapid introduction of pay-TV at the lowest possible cost.[6]

Pay-TV firms can capitalize on the investment already made in cable systems. Cable companies can expand the field capabilities necessary for

---

[6]Canadian Radio - Television and Telecommunications Commission, COMMENTS: Pay Television (Ottawa: 1976), comment #28 by PIN Pay Television Network, p. II-6 to II-10.

pay-TV with only marginal cost increases, whereas a pay-TV network operating a separate field organization would have to incur major fixed costs. Cable companies already have an in-field organization that can perform the essential functions of sales, installation, servicing, and ongoing billing and customer service.

With a substantial subscriber list, cable-TV is also in an advantageous position to attract pay-TV customers. And consumers will also find it more convenient to deal with only a single television-related field organization — cable customers are accustomed to cable companies making service calls and thus the addition of a new field may only increase consumer confusion and create additional costs and thus possibly decrease consumer confidence in pay-TV services.

An over-the-air system would be burdened with shortages in off-air frequency and the sizeable cost of purchasing a UHF television station. The capacity of the coaxial cable, on the other hand, has no inherent limits, and cable offers the potential of widening the selection of channels far beyond what the present television system can offer. Moreover, the home equipment required for unscrambling pay-TV signals and for subscriber billing is simpler and cheaper with cable than with over-the-air broadcasting.[7]

Cable delivery is also desirable if pay-TV is ever to be developed into a system where payment is made on a per-program basis. Such a system requires the ability of the viewer to interact with the exhibitor (a two-way system), and this will only be feasible on a cable-based system.

---

[7] Richard Adler and Walter Baer, The Electronic Box Office: Humanities and Arts on the Cable, Praeger Special Studies in U.S. Economic, Social, and Political Issues (New York: Praeger Publishers, 1974), p. 20.

The coaxial cable has an information retrieval capacity that makes this possible.

Perhaps the biggest drawback of a cable-delivery system is its lack of universal availability. However, a reasonably high level of cable penetration currently exists in Canada (almost three million Canadian homes received cable by 1976 over 351 systems[8]), and in the near future almost all major communities in Canada can be expected to have cable services. According to data provided by PTN, cable should pass 70% of all Canadian households by the beginning of 1978.[9] Although an over-the-air system could deliver pay-TV programming to considerably more subscribers, this potential advantage is effectively eliminated by higher hardware costs that would have to be passed on in the form of higher subscriber charges.

From an economic viewpoint, the market if left to itself would certainly have sufficient motivation to find the best delivery system for pay-TV programming (best in terms of direct costs and benefits). However, equity considerations may favor a form of delivery which might not naturally result from the functioning of the market.

The necessary technology for a pay-TV system already exists, as there are a variety of devices and methods that can be used in the delivery of a pay-TV signal, whether it be cable or over-the-air delivery. Generally speaking, there are two common ways in which pay-TV systems operate. One is by sending scrambled signals to all subscribers (commonly called subscription television). A subscriber's TV set would be equipped with a

---

8 Canada, Dept. of Communications, 1975 - 1976 Annual Report.

9 CRTC, COMMENTS: Pay Television, Comment #28 by PTN Pay Television Network, p. II-8.

converter that decodes the scrambled images and sounds for the viewer. A different code is used for each channel or program, and subscribers pay for a particular program or channel allowing their converter to decode the scrambled signal. Some pay-TV systems involve periodic billings, some require the deposit of coins into a device associated with the decoder, and some offer tickets that can be used to obtain the specific programming desired.

Another way in which pay-TV programming can be delivered is through the use of a "trap" device that is installed right into the coaxial cable. A trap device or "tap" will be used if cable companies adopt a pay-by-program approach. An "addressable" tap can automatically direct programs to individual households. (However, the necessary hardware for this system is considerably more expensive than that used in a converter system.)

It may take many years and millions of dollars to develop, market and install the hardware for a pay-TV system. A typical pay-TV company will face large depreciation charges for hardware in their first few years of business, thus requiring a large market penetration for feasible operation. Actually, it is difficult to safely predict at this time exactly what type of technology will eventually be used by pay-TV systems in Canada, as the electronics industry is continually developing better and less expensive equipment for the delivery of pay television programming.

What pay-TV can offer

Initially, pay-TV can be expected to provide greater options in viewing entertainment. Judging from past experience, major film productions and sports events are likely to be the types of shows offered at first. In an analysis of consumer demand for pay-TV, the Stanford Research

Institute estimated the relative demands among program categories (Table 1.).

TABLE 1.

ESTIMATED PERCENTAGE OF PROGRAMS TO BE OFFERED BY PAY-TV, BY CATEGORY

Program Category	Percentage of Programs Per Year
Major theatrical films	25
Major sports	52
Major performing arts	5
Educational specials	5
News specials	2
Popular entertainment concerts	11
Total	100%

Source: Stanford Research Institute, Analysis of Consumer Demand For Pay Television (Final Report); prepared for: Office of Telecommunications Policy (May, 1975), p. 72.

These estimates were based on the assumption that pay-TV will offer primarily "box office" programs, that is, programs that appeal to fairly large audiences who would be willing to pay high prices for live attendance at these events.

To date the vast majority of pay-TV programming has centered around movies, but over time programming can be expected to become more specialized, appealing to individualized markets such as, for example, opera lovers or people with an interest in science documentaries. These markets may not be large enough in numbers to warrant specialized programming on the current advertiser-supported television system, but pay-TV will be able to offer such programming if these well-defined markets have a demand sufficient to generate enough revenues from viewers to pay for the programming.

(Reasons for why pay-TV would be able to expand viewer choice will be discussed further on in this chapter.) Also, over time, technological innovations, larger subscriber lists and the achievement of scale economies are likely to allow pay-TV to offer more of these "marginal" programs.

Industry Structure

The pay-TV industry can be separated into two major groups. One group is the network entity, which acquires, packages and distributes the programming. The second major group is the exhibitors, who are responsible for "delivering" the program packages. In Canada the exhibitors are composed primarily of the cable companies. In many cases both of these functions are filled within one company, usually an existing cable company.

The functions of the network entity are many, although, as stated above, it is essentially responsible for acquiring, packaging and distributing programming. The acquiring of programs includes the actual financing and production of some programming, although the network usually acts as a middleman between program producers and the exhibitors. The packaging of programming includes the scheduling of programs for local broadcast stations, and thus involves an evaluation of programs and consumer attitudes and response. The distribution of programming also includes the marketing function. In addition, the network may participate in the management processes of both producers and exhibitors.

History of the Industry

Pay-TV has received much attention and publicity over the years, and has been technically feasible since the early 1950's.

But a combination of factors — government regulations, difficulty in obtaining programming, an industry reluctance to invest heavily in research and development, consumer disinterest, opposition by broadcasters and others, and poor economic conditions — have succeeded in keeping the industry from blossoming as initially envisioned. The situations in which pay-TV systems have been implemented, both on cable and over-the-air stations, have been relatively small in size and primarily experimental in nature; and, financially, these attempts have been by and large unsuccessful in establishing pay-TV as a viable system.

Pay-TV was attempted as early as 1951 in Chicago, and in 1953 a pay-TV system was implemented in Palm Springs, California; both experienced little success because of inadequate audience support. In the late 1950's another unsuccessful experiment was tried in Oklahoma, an experiment which attracted much attention at the time to pay-TV and its potential. The programming in the Oklahoma experiment was confined solely to motion pictures.

In 1960 one of the largest pay-TV systems to date was introduced in Etobicoke, a suburb of Toronto, providing subscribers with three channels and a choice of types of programming. The system was a pioneer in leased cables. Operations of the Etobicoke system closed in 1965 because of poor sales. During its period of operation it was discovered that current movies and sports were the most popular programs, and the use of pay-TV generally appealed to "selective" tastes, resulting in no readily discernible adverse effects on commercial television viewing (although revenues per subscriber were considerably lower than had been anticipated).[10]

---

[10] W. Qual and L. Martin, Broadcasting Management (New York: Hastings House, 1968), p. 410

Perhaps the most famous pay-TV system was implemented in 1962 in Hartford, Connecticut as a planned seven-year experiment. Its purpose was to offer high quality programs which were not available on free-TV. It was suggested that the financial failure of the experiment can be attributed to the fact that less cultural programming and more general entertainment was offered than had been expected (movies accounted for approximately 86.5% of the programs offered). A wealth of data flowed from the Hartford experiment.

In recent years pay-TV systems have been springing up all over the U.S. and Canada. As of June, 1976, one source has estimated that 253 cable systems in the U.S. provided 766,000 subscribers extra programming for which an additional fee was charged on top of the regular cable charges.[11] One of the primary causes of this increase appears to be the improved programming.

In Canada, three of the largest cable companies (National Canadian Cablesystems Ltd., and Rogers Cable TV Ltd.) recently formed the PTN Pay Television Network. This consortium of cablecasters was organized because of a realization of the potentialities of pay-TV, and their goal is to establish control over the production and distribution of pay-TV. Their stated objective is to "bring together cable companies, broadcasters, independent producers and other members of the program production community to discuss and plan the most effective approach to introducing pay television in Canada."[12]

---

11 CRTC, COMMENTS: Pay Television, "A Report on Pay Television in the U.S.A.", Comment #76 by Paul Hagon and Associates, p.22.

12 CRTC, COMMENTS: Pay Television, Comment #28 by PTN Pay Television Network, p. 2.

Stage In Development.

In determining regulatory policies, the stage in an industry's life-cycle should be taken into consideration. As an industry changes and matures, the optimal regulatory policy for it will also change. William Shepherd has developed a "utility life-cycle" that is useful in analyzing an industry's stage in development.[13] Shepherd defines a regulated utility as a system which provides services to a spectrum of users whose levels and elasticities of demand usually vary significantly. The life-cycle involves four stages through which most utilities pass:

Stage 1: usually a brief but decisive stage for the system, often leading to control by patents.

Examples:		
	Railways	(1820 - 35)
	Electricity	(1870 - 85)
	Airlines	(1920 - 25)
	Cable-TV	(1950 - 55)

Stage 2: a system of creation and growth; cross-subsidies among users and a separation of creamy and skim markets becomes embedded in the price structure; the service seeks regulated status for permanence, legitimacy and market control, and the regulatory act as a promoter of the service.

Examples:		
	Railways	(1820 - 80)
	Electricity	(1885 - 1960)
	Airlines	(1925 - 66)
	Cable-TV	(1955 -)

Stage 3: the system becomes complete in terms of technology and market saturation; a defensive system is adopted; competing new technologies arise, beyond the utility's control; physical layout and pricing structure do not fit evolving city patterns as well as before; users in lucrative markets challenge the prices they face; the utility more and more finds itself trying to obstruct new technology or find ways to fit it into its private optimum structure.

Examples:		
	Electricity	(1960 -)
	Airlines	(1965 -)

---

[13] W.G. Shepherd and T.G. Gies, eds., Regulation in Further Perspective (Cambridge, Mass.: Ballinger Publishing Co., 1974), p. 6 - 7. The characteristics of the various stages have been constructed through observation of a number of regulated industries.

Stage 4: the systemic monopoly yields to these pressu of competition and technology, and reverts to conventional competitive processes.

Examples: Railways (passenger) (1935 -)

According to Shepherd, regulation is introduced in most industries in stage 2, after profit expectations and rate levels and structure have been established. Thus, so as not to upset industry balance, regulation "harmonizes" with the interests of the utility and its larger industrial customers, and its primary effect is to legitimize and to smooth out the interest-group conflicts. As Shepherd points out, the rate structure is never thoroughly assessed and changed, and thus regulation introduced at this stage is often not directly aimed at protecting the best interests of the general population.

For this reason regulation of pay-TV will be most beneficial to society if it is implemented now in the early stages of the industry's development:

> ...if action is not taken now to develop a nationally coordinated service, with high-quality Canadian content, we run the risk of ending up with many independent systems that would fail to achieve the objectives now set forth for pay television. But if we can quickly work towards a single national network, pay television has the potential of achieving many valuable results.[14]

It took about ten years of cable development (1960 - 70) before the Canadian government recognized the need for, and acted to formulate a national policy for Canadian cable operations, and the result has been costly to the Canadian economy and the cable industry. Many of the problems and headaches being experienced now could have been avoided by the introduction of a national policy in the early stages of the industry's development.

---

[14] CRTC, COMMENTS: Pay Television, Comment #28 by PTN Pay Television Network, p. 2.

According to Shepherd's classification, the pay-TV industry in Canada is still in stage 1, but it appears to be on the verge of a prominent growth which will take it into further stages of development. Referring to the detailed analysis of consumer demand for pay-TV by the Stanford Research Institute, Table 2. gives the upper-bound growth projection for pay-TV subscribership and revenues in the U.S. which was estimated.

TABLE 2.

PROJECTED UPPER-BOUND ON PAY-TV GROWTH

	End of Year				
	1974	1977	1979	1982	1985
Subscribers (millions of households)					
Pay-cable	0.1	1.6	5.2	12.1	14.7
Subscription-TV	0.0	0.5	1.8	3.3	1.5
Total	0.1	2.1	7.0	15.4	16.2
Revenues ($ millions)	8.	249	821	1,708	1,903

Source: Stanford Research Institute, Analysis of Consumer Demand For Pay Television (Final Report), prepared for: Office of Telecommunications Policy (May, 1975), p. 101.

For 1977, the upper-bound on subscribership is 2.1 million households for a joint pay-cable cable-STV industry, with revenues of about $249 million. By 1985, U.S. subscribership is projected to grow to about 16 million households with revenues of approximately $1.9 billion.

Some of the potential of pay-TV is reflected in the growing predominance of cable systems in our society. The prospects for pay-TV have brightened over the last several years with the rapid growth of cable distribution systems for homes, apartments, and hotels. In Canada almost

three million households received cable by 1976.[15]

On the other side of the picture, the projected growth and prominence of pay-TV may not necessarily be fully realized even in the absence of restrictive regulation. While it is easy to be optimistic about the industry, one must be careful to realistically assess its past performance and future potential. After all, the concept of pay-TV has been around for over two decades, but yet it has not flourished to any great extent as an industry to date.

Despite all the reasons provided by industry enthusiasts for why the industry has not flourished, in the end analysis the biggest reason has been lack of consumer acceptance. Consumers, in areas where pay-TV has been offered, have not been breaking down the doors of pay-TV companies to get their services; if they were it is likely that the industry would have somehow solved all the other problems facing them. (If a demand or market exists, the business world usually will find some way to meet that demand.)

Some possible explanations for the rather anemic consumer acceptance of pay-TV include: 1) although pay-TV has the ability to provide programming more closely correlated with consumer preferences, it has not done so to any great degree as programming has been rather narrow in scope; 2) consumers are reluctant to pay for a type of product they perceive to be available for free; 3) viewers are content with existing television programming, so that payment for additional programming is not

---

[15] Canada Dept. of Communications, 1975 - 1976 Annual Report.

at active to them; 4) most pay-TV programming has been offered on a per-channel basis.

There may be other explanations for the consumer disinterest in pay-TV, but an analysis of the current demand or market for television services is not included in this paper. All that will be said is that the eventual success or failure of pay-TV will depend on the extent to which consumers are willing to pay for programs not available from the current advertiser-supported television system.

In conclusion, pay-TV is likely to eventually become a regular service of some magnitude in Canada, and thus it is an issue which should be considered now so the appropriate regulatory policy can be developed before the industry becomes established in basic structure.

The Political Reaction to Pay-TV

The political debate over pay-TV in past years has been quite extensive although very inconclusive in large part. In separating out the various issues at hand one has to be aware of the self-interests of the various groups concerned in the debate. Arrayed against the efforts of pay-TV systems to become established have been the businesses or entities who are threatened by pay-TV — primarily over-the-air broadcasters and theatre owners.

Over-the-air broadcasters are those most concerned about pay-TV and the threat it offers in luring away part of their viewing audience. More will be said on this matter in Chapter III.

Similarly, theatre owners are concerned with the potential loss of patrons to pay-TV systems. Pay-TV can not only compete with theatre owners by offering a lower price to view movies (because larger audiences can be reached), but pay-TV can also be more physically convenient.

Pay-TV is attractive to that group of people who "don't care for the time and money it takes to get a babysitter, get dressed up, drive to a theatre, pay to park and get into the theatre."[16] A pay-TV system provides a safe and convenient means of viewing a new movie or other major entertainment event.

Also, pay-TV further threatens movie theatres by competing with them for a limited number of quality motion pictures. Thus, unless the emergence of pay-TV results in a significant increase in the amount of movies produced, prices paid for the rights to exhibit movies are likely to be bid up.

On the other hand, it has been suggested by some that movie theatres will not suffer as much as is sometimes claimed. The potential audience for pay-TV may be very different from that of the present movie-going audience, and new and different types of movies will be produced and shown on pay-TV.

Also, as the Motion Picture Theatre Associations of Canada themselves have stated in their submission to the CRTC, the motion picture distributors are currently not looking to use pay-TV to replace theatres as the primary medium for their films, but rather they are looking at pay-TV as a supplementary source of income before selling a film to free-TV.[17] It appears that the policy of the major film suppliers concerning pay-TV is to release a film approximately eight to twelve months after the initial theatre release. Whether film suppliers will in fact be able to ignore any bids by pay-TV firms for new movies will depend on their

---

[16] A. Rimberg, "Pay-TV May Hold Key to Cable-TV's Future in Vast Urban Market," Wall Street Journal, (17 May 1973).

[17] CRTC, COMMENTS: Pay Television, Comment #26 by Motion Picture Theatre Associations of Canada, p. 1.

ability to use any market power they have to induce non-competitive behavior — some sort of collusion amongst suppliers along with entry barriers to the industry would be needed to make such a policy effective.

For these reasons and others, pay television has encountered tremendous opposition in North America, particularly by broadcasters and theatre owners who have been continuously pressing for legislation to severely restrict or even entirely prohibit pay-TV.[18] It is only natural that these groups are interested in their own self-preservation and growth.

Because of this strong opposition, pay-TV, if it ever does emerge, will almost certainly spring up alongside the present advertiser-supported television system. Political interests will be sure to maintain the present existing networks. The government usually sees its role

---

18 It should be noted that the various owners of media groups are not wholly against pay-TV. Possibly because of the perceived potential threat of pay-TV and cable systems in general, the various media have invested heavily in cable operations. Industry statistics for 1974 reveal the interesting fact that media groups held partial ownership in approximately three-fourths of all U.S. cable companies: broadcast related ownership in 33% of cable systems, program producers in 25%, newspapers in 15%, publishing companies in 7% and theatre owners in 4% of cable operations. See: "A Short Course in Cable, 1974", Broadcasting (Apr. 22, 1974), 23; and "A Short Course in Cable, 1975", Broadcasting (Apr. 14, 1975), 56.

as to come to some sort of a political compromise and balance out the conflicting interests of the major groups involved:

> ...we must take care that pay television is introduced in an orderly and controlled fashion. We must ensure that it brings maximum benefit on a national plane to all the groups interested in broadcasting — private and public broadcasters, cable operators, program producers, actors, writers, and most of all, the Canadian public.[19]

It is true that any regulation policy involving pay-TV should be developed within the framework of the entire broadcasting industry and not in isolation. Pay-TV exists in the wider context of the communications system, and is not independent of the existing structures, and thus integration of pay-TV into a national framework of broadcasting policies and regulation is needed to maintain direction and balance within the broadcasting system.

However, the public has a large interest in pay-TV also because of its effect on the quantity, quality, and expense of television programming, and this interest should have considerable weight in determining regulatory policies. Even through pay-TV may be "bad" for some interest groups, it may be good for society as a whole in the long-run, and thus the industry should be analyzed in all respects in an attempt to determine how, in the best interests of the whole of society, the industry should be regulated if indeed it should be regulated at all.

Unlike the U.S. situation where hearings by the Federal Communications Commission (FCC) on pay-TV began as early as 1955, the Canadian political reaction to pay-TV has been concentrated into the last few

---

19 Canada, Dept. of Communications, "Pay Television", Notes For a Speech: By the Honourable Jeanne Sauve, p. 2

years, as it is only recently that pay-TV has threatened to become prominent and established as a major industry in Canada.

Realization of the inevitability of pay-TV and the concern by government officials over its impact on over-the-air broadcasting, general social patterns, and the television and film industry, led the Canadian Radio - Television and Telecommunications Commission (CRTC) in 1976 to invite interested groups to submit briefs commenting on government regulation of pay-TV and the introduction of a pay-TV system in Canada. (Under the Broadcasting Act the CRTC has been given the right to make all federal broadcasting policies about non-technical matters, although the debate over pay-TV has involved jurisdictional problems between the federal and provincial governments.)

A total of 140 submissions were received by the Commission, with the majority being strongly opposed to pay-TV. Among those opposed were the Canadian Conference on Arts and the two major broadcasting networks in Canada. The cablecasters felt pay-TV should be allowed to operate as a network whose programming is distributed on cable, but under regulation of profits.

The view which emerged from the government in 1976 appears to be that they feel pay-TV should be run on a "non-profit" basis.[20] The rationale for this, however, has not been explicitly given. There is no clear economic reason why the pay-TV industry should be regulated on a "non-profit" basis.

---

20 "Cable Industry Reacts Favorably to Comments on Pay Television," Globe and Mail, 13 November 1976, p. B2.

Also, it is not clear whether the government intends the non-profit criterion to be applied also to the cable companies involved in delivering pay-TV program packages as well as the group or network entity responsible for acquiring and packaging the programming. (Apparently the cable industry has, quite naturally, interpreted the government position as only applying to the network entities.[21]) It is also unclear as to what form the government feels regulation should take so as to ensure that excess profits are not earned.

If the rate of return for the industry is regulated, for political reasons, consideration should be given for the stage in development of the industry. Because pay-TV is in the initial stages of expansion and development, allowance should be made so that the rate of return is high enough to attract the necessary capital needed for growth. There is still a relatively high degree of risk associated with the acceptance and operation of pay-TV, and the allowable rate of return will have to be set so that it will attract sufficient funds for the considerable investment needed.

The Federal government is attempting to form a regulatory policy for the pay-TV industry that will incorporate three major objectives. Canada's Minister of Communications, Jeanne Sauve, outlined these three objectives in a speech given to the Canadian Cable Television Association in 1976:[22]

1. pay-TV should be regulated so that it provides programming that is not currently being offered by broadcasters

---

[21] Ibid.

[22] Canada, Dept. of Communications, "Pay Television," p. 6 - 7.

2. the production of popular and "high-quality" Canadian programs must be ensured

3. regulation must ensure that programs are produced in Canada for international sale

As far as the first objective is concerned, it appears that it is more of a means to an end rather than a goal or objective itself. More will be said on the rationale behind this "objective" later on. The second objective, a requirement specifically stipulated in The Broadcasting Act which governs the broadcasting system in Canada, is likewise a little ambiguous for it is not clear what is meant by "high-quality" Canadian programs. Also, the question arises that if regulation is needed to ensure the production of "high-quality" programs, then why should these programs be produced in Canada if there is insufficient demand for them? (Would they not be produced without regulation if there was a sufficient demand for them?) Regulation should not be needed to achieve the third objective if the second objective is achieved; the international sale of programs will only occur if "high-quality" Canadian programs are produced.

To achieve these objectives three general options have been proposed. The first option involves individual licensees, the traditional form of regulation in the broadcasting industry. However, if no additional regulation is enacted other than licensees, this option may lead pay-TV into the same situation in which the Canadian motion picture industry now finds itself where foreign interests have acquired a virtual monopoly over exhibition and distribution, and control over the financial resources needed for production. Because a Canadian program distributor currently does not exit, the individual licensees could be forced to deal with foreign distributors thus opening the way for foreign organizations to extract supernormal profits from the industry. Foreign control of

exhibition and distribution would also make it difficult for Canadian producers to obtain exhibition opportunities.

A second option which may circumvent this problem involves the formation of a consortium of cable and broadcasting operators. This possibility has some merit, although steps would have to be taken to ensure that the participating parties do not abuse their position and power in pursuing their own vested interests. The possibility also exists of the government becoming a participant in the consortium.

The third option, an option the government appears to favour, is the establishment of a national pay-TV distributor or network, owned and operated independently from existing cable interests. All licenses to deliver or exhibit pay-TV programming in Canada would be given only under the condition that programs be obtained from this network. This network could either be made a private corporation with extensive monopoly powers, a public corporation, or some mixture of the two. Some type of regulation of the network would be necessary to ensure it doesn't use its monopolistic position to extract monopoly rents from pay-TV exhibitors.

Such a pay-TV network could act as a Canadian program distributor which would considerably strengthen the Canadian program production industry. This option is favoured by the government primarily because it would enhance the production, exhibition, and promotion of Canadian programs. Section 3(d) of The Broadcasting Act requires the broadcasting system to use "predominantly Canadian creative and other resources".

In determining the form which regulation should take, one of the important issues involves whom the network entity should be comprised of and whether the network should be made a public corporation or left as a private entity but given extensive monopoly powers.

Because pay-TV is a new component of the broadcasting system, the necessary skills and experience for successful operation do not reside with any one member of the present system, and thus it may be desirable to bring together in the pay-TV network members from the various parts of the total program and broadcasting system. Pay-TV will require unique programming, promotion, and distribution skills.

Also, despite government intentions to set up a network owned and operated independently from existing cable interests, it may be desirable to allow the cablecasters to have a major position in determining network operations, since cable-delivery is the most feasible form of delivery for pay television. The rationale behind this is so as to not risk damaging the relationships with existing cable customers, and also to avoid losing some of the considerable capital investment required to introduce pay-TV.[23]

As far as the issue of public or private control is concerned, it appears that our society favors public sector control only if private sector control is not feasible. This is desirable not only on ethical grounds, but government ownership and control generally involves more bureaucratic and administrative costs and entanglements, as well as some loss of initiative and creativity. The role of government regulation should not be to replace the market mechanism unless the market cannot perform desirably.

The alternative to the formation of a national network is to allow

---

[23]PTN estimates it may take over $34 million in additional equipment for it to initiate pay-TV service over cable; see CRTC COMMENTS: Pay Television, Comment #28 by PTN Pay Television Network, p. 111 - 5.

for a number of individual networks or programmers that would compete against each other to obtain leases from cable operators to deliver their programming to viewers. Here the separation of the program distributor or network from the cable operators is a more important issue.

Pay-TV and Program Variety

It appears at this point in time that the Canadian government favors a regulatory policy which would considerably restrict the growth and development of the pay-TV industry. Yet the pay-TV industry has some very desirable characteristics when compared to the present "free" television system. This section will expand on one of these desirable characteristics — namely, the ability of pay-TV to better satisfy consumer wants by offering television viewers a programming mix that more accurately corresponds to their preferences.

Explanation of Current Program Uniformity

The massive potential for pay-TV as an entertainment medium is based partially on the need for greater selectivity of entertainment for the North American population, a diversity that is not currently being offered by the TV networks.[24] This increasing demand is being generated by an

---

[24] Support for this viewpoint of a growing gap between audience expectations and TV performance can be found in a major study by the broadcast consulting firm of McHugh & Hoffman. The results of this research are reported in "The viewer's ahead of the medium," Broadcasting (6 June 1977); 35 - 36. The research report claims that: "During the last 10 years, the failure of the TV programers to stay in step with the audience's maturation and to remain sensitive to the societal force and functions of the medium, has caused a serious 'loosening' of the audience commitment to the medium."

expanding population along with the growing affluence of the average consumer (entertainment has a relatively high income elasticity). Also, consumers today have greater amounts of leisure time available to them.

Currently, the existing advertiser-supported television system provides a high degree of program uniformity by most standards. Instead of offering a great variety of programs appealing to a wide range of tastes, the current mix of programming involves a high degree of program duplication aimed at the mass audience. Pervasive throughout the literature on boradcasting is the charge that television provides excessive mass appeal programming at the expense of failing to satisfy minority tastes.

Only a casual glance at any current TV program listing is needed to appreciate the amount of duplication involved in the conventional free TV system. For example, in one randomly chosen week for the city of Edmonton, with six channels available to viewers (three broadcast and three cable), 23 different detective shows were offered 43 times, and 23 different game-shows were offered 101 times (see Appendix A).

If a significant demand does in fact exist for more diversity in television programming, then what is the reason for the current high degree of uniformity? After all, television broadcasters are interested in maximizing audience size which determines profit levels through advertising rates.

The broadcasters themselves argue that television program homogeniety is a result of a lack of sufficient demand for other types of productions (however, demand is only in the sense of numbers of people willing to watch a program at a zero price). They claim that more diversity would be offered if there were identifiable minorities with distinct preferences.

Since too much program homogeneity cannot be explained away by the behavioral motivation of broadcasters, the reasons for excessive program duplication lie elsewhere. Excessive duplication can perhaps be explained by either institutional or technological reasons.

In trying to explain product homogeneity traditional economic analysis is not of much help. Standard economic theory has not yet developed a comprehensive framework to explain product variety and quality choices by firms. Economic theory can explain quantity decisions for fixed quality levels of a given product, but in the real world firms often have to decide simultaneously on what to produce, of what quality, and of what quantity.

On a conceptual level perhaps the best way to approach the problem is via the "characteristics" approach. To begin with, potential product variety is enormous. If a firm is defined as a decision making unit — the decision being whether to produce or not produce a product — it can be seen that the number of "firms" in the economy will exceed the number of real firms (firms with actual physical assets). Some firms will not be in physical existence because they have chosen not to produce. And, since each of these "firms" may consider a large number of products and whether or not to produce them, the potential number of products which might be produced is indeed very large.

However, there are limiting factors to the number of goods actually produced and the different degrees of differentiation of each product. Because of cost constraints imposed by scarce resources, the size of the possible production set is limited (a choice has to be made as to what to produce).

In addition, a further limiting factor is the needs and wants of

consumers. Assuming that a firm cannot create or manipulate needs and wants in consumers, the number and types of needs of consumers not only limits the number of possible goods to produce (needs being reflected in demand), but the intensity of needs makes some goods more worthwhile to produce than o- rs. In other words, needs provide a further restriction on what types of products will be most pofitable for firms to produce under their cost constraints.

One general approach to the issue of product variety which incorporates both of these limiting factors in determining quality and quantity output levels is the "characteristics" approach which was initially developed by Lancaster.[25] This approach treats goods as bundles of "characteristics"; the goods themselves are not assumed to directly provide utility to the consumer, but rather they provide basic characteristics which satisfy consumer needs.

However, in regards to television programming, it is difficult to determine what the actual "characteristics" are of programming from which viewers derive utility. One cannot use general characteristics such as "information" or "entertainment" because one of the crucial assumptions of the characteristics approach as it has been developed is that the characteristics possessed by a product are the same for all consumers and are consumed in the same quantities. This assumption would not be very descriptive of television programming where for example, sports fans will derive more "entertainment" from watching a particular football game than a person who has no interest in sports. Almost all of the

---

[25] K. Lancaster, "A New Approach to Consumer Theory", Journal of Political Economy, (April 1966): 132 - 57.

utility derived from viewing television comes via psychic satisfaction, and thus a large degree of personal subjectiveness is involved, and this is not compatible with the underlying assumptions of the characteristic approach to product variety. Thus, the usefulness of using the characteristics approach to explain television program variety is very limited.

Since it is difficult to implement a general framework for evaluating industrial decisions concerning product variety, a less systematic approach is used here to explain the phenomenon of program uniformity. Various hypotheses will be forwarded to explain program duplication.

One possible explanation of program uniformity arises if one is willing to make two behavioral assumptions concerning viewers and broadcasters. Steiner initially produced the argument that if you combine the assumption that individuals will only view their preferred program type, and if this preferred program type is not available then the individual will choose not to view any program, with the assumption that television stations will seek to maximize audience size, a station in determining its programming may find it more profitable to duplicate existing popular programming and share the current audience rather than to program so as to try to attract previously unserved viewers.[26] The factor which will determine whether it will be more profitable to do so is the particular distribution of preferences amongst viewers.

A simple example can illustrate this point. Assume 75% of the viewing public have detective shows as their first and only choice, and that

---

[26] P. Steiner, "Program Patterns and Preferences and the Workability of Competition in Radio Broadcasting," The Quarterly Journal of Economics LXVI (May 1952): 194 - 223.

of the remaining viewers 15% prefer science documentaries and 10% prefer roller derby. If three television stations are in operation, for any single time period all three broadcasters will choose to offer detective shows which, if the audience for any particular type of program is shared equally among the broadcasters showing it, will still give them a larger viewing audience (25%) than if they were the only ones to offer science documentaries or roller derbies.

Thus, a large degree of homogeneity in tastes amongst consumers and a small number of competing broadcasters is likely to lead to duplication of television programming in an advertiser-supported system. Minority preferences will tend to be ignored in favor of majority preferences. Since viewing time and the number of available channels is limited, the maximization of audience size is likely to occur when the bulk of programming is aimed at the mass audience.

The important question, however, is whether this is one of the major explanations of current program uniformity. Does a large degree of homogeneity in tastes exist amongst consumers and does this lead to programming which caters to the tastes of the majority to the exclusion of minority preferences?

Even if you remove the assumption of viewers watching only their preferred choices, the argument may still be valid. Even if people will, in general, watch their second or third choice if their first is not available, it still may be more profitable to duplicate existing programming as existing programming may be viewed by people who rate it as their second or third choice. However, if you do allow second or third choices the results will be more uncertain as behavior will now depend on such factors as how much the available program alternatives can differ from

viewer's first choice before he will become a non-viewer. A model of viewing behavior is need to analyze the problem.

J. Beebe has developed a computer model of profit-maximizing advertiser-supported broadcasting which allows for lesser preferred program substitutes by varying the distribution of viewer tastes and by considering the implications of preference intensities.[27] He also varies within the model the assumptions made for program costs and channel capacity. By assuming discrete program types and particular viewer preference sets, the predicted program patterns generally reveal competitive tendencies toward program duplication and imitation.[28]

So far we have just considered a one-period model. In a multi-period analysis allowance has to be made for the fact that viewer preferences for program types are likely to change over successive periods once that program type has been offered. Also, those people whose first choice isn't initially offered may form a large enough market that it may be profitable to devote some portion of the programming towards their tastes after the popular programming has been offered (although it is likely that their individual preferences are relatively dispersed, thus making it difficult to cater to all minority preferences).

---

[27] J. Beebe, "Institutional Structure and Program Choices in Television Markets," The Quarterly Journal of Economics 91 (February 1977): 15 - 37.

[28] Ibid., p. 26 However, if one of the assumptions made by Beebe — that all channels simultaneously offering the same program type share equally in the program's total audience — was altered so that duplication of an existing program type would initially receive a smaller proportion of the audience than the existing program (a reasonable assumption), then the tendency towards duplication would not be as great.

These additional factors greatly complicate the problem and make it more difficult to explain program uniformity or diversity over a multi-time period. A multi-period model, however, is likely to predict more diversified programming than a single-period model.

An alternative explanation of program uniformity as set forth by Leonard Ross is that it is essentially a result of the oligopolistic nature of the current TV system. The dominance of the networks in the production and selection of programs tends to:

> ...restrict the entry of truly independent producers and insulate the networks' narrow perception of public taste against outside influence. Indeed, the very existence of networks is also held to preclude access of local and innovative programming.[29]

In a study of program duplication in the U.S., Hall and Batlivala discovered that most of the observed duplication was in fact network in origin.[30] Many people in the television industry feel that commercial television programming is the product of the caution and conservatism of the broadcast and advertising corporations, and not of the viewing audience.

Another related explanation for program uniformity stems from the inability of programmers to accurately discern the wants and preferences

---

[29]Leonard Ross, Economic and Legal Foundations of Cable Television, A Sage Research Paper, (Beverly Hills: Sage Publications, 1974), p. 8.

[30]W. Hall and B. Batlivala, "Market Structure and Duplication in TV Broadcasting," Land Economics 47 (November 1971): 405 - 10.

of the viewing public:

> Programming is anything but an exact science, and the networks are constantly being surprised by the popular acceptance of show X over show Y. Within the commercial complex, the only formula for success is the imitation of an existing successful program. The occasional breakaway success in a new season always represents a departure until imitation turns it into a standard formula.[31]

Due to the risks involved, there may be a greater tendency towards program duplication than is warranted by consumer tastes.

Another reason why the current television system is failing to provide more optimal degrees of program variety, is that only one of the dimensions of demand is predominant in influencing program choice by broadcasters. The following section has more to say on this important factor.

Why Pay-TV Can Offer a Greater Variety in Television Programming

The actual degree to which pay-TV will offer consumers a more diversified menu in television programming is an extremely important factor in developing regulatory policies which will affect the growth and viability of pay-TV. If, in the long run, only minor changes in programming will occur because of pay-TV, the case for pay-TV is considerably weakened.

There are three basic reasons why pay-TV is able to offer programming that would not normally be shown on a conventional "free" TV system, and thus help satisfy some of the need for greater selectivity in television

---

[31] R.O. Moore, "Public Television Programming and the Future: A Radical Approach," in The Future of Public Broadcasting, eds. D. Cater and M. Nyhan (New York: Praeger Publishers, 1976), p. 244 - 45.

entertainment. The first and most important reason stems from the fact that a pay-TV system can more accurately reflect consumer preferences than free television because a pay-TV system allows the magnitude or intensity of viewer preferences to be registered in addition to their direction. Because positive prices allow pay-TV to incorporate another dimension of audience demand — namely, demand intensity — some programs might be offered that could not otherwise be provided. J. Minasian explains:

> ...in an advertising-supported system...the program results reflect an all-or-nothing type of voting since votes take weights of either one (viewer) or zero (nonviewer). In contrast, a subscription system allows proportional representation, since votes take different weights (different prices paid for different kinds of programs) and reveal the voters' subjective evaluations of the program. Therefore, a subscription system can be expected to yield a more diversified program menu than an advertising system, because the former enables individuals, by concentrating their dollar votes, to overcome the 'unpopularity' of their tastes.[32]

A weighted voting system is more accurate in revealing consumer preferences than a purely directional voting system, and individual program prices would allow pay-TV to achieve such a weighted voting system. In addition to audience size, preference intensities as reflected in the shape of demand curves (the responsiveness of audience size to price variations) would influence the programming mix.[33]

---

[32] Jora Minasian, "Television Pricing and the Theory of Public Goods," Journal of Law and Economics 7 (October 1964): 75.

[33] However, the charging of a single positive price would not allow for the registry of the preferences of consumers who would be willing to purchase the rights to view a program at a lower positive price than the one charged — the optimum registry of preferences could only occur through perfect price discrimination.

The type of programs offered by pay-TV that would not be shown on free-TV would tend to be those which are relatively more expensive to produce and those with relatively low price elasticities (where higher prices would do less to discourage consumers from viewing).[34] What type of programs have low elasticities is an important empirical question. The growth of pay-TV over the next several years should generate sufficient data to allow empirical measurement of program elasticities.

Because of the ability of a pay-TV system to charge a positive price, relatively expensive programs whose markets are not large enough in numbers to warrant the programs on an advertiser-financed TV system may be offered by pay television if the market has a demand whose elasticity is such that it can generate sufficient revenues to pay for the programming — resulting in a direct gain to viewers.

There is good reason to believe that revenues raised from a pay-TV system can exceed potential revenues from advertisers in many cases allowing for the production of some programs which would be too expensive to be produced in an advertiser-supported television system. Advertising revenues in the free television system depend on both the amount of advertising time available and the rates paid by advertisers per unit of advertising time. Because advertising time has technical limits and advertising rates are limited by competition with other advertising media, total revenues are limited to some amount which appears to be much less

---

[34] However, as will be seen in Chapter III, pay-TV does have some bias away from the optimum which involves programs with low price elasticiticies, although the bias is less than that associated with a free-TV system, this bias could only be removed through perfect price discrimination amongst all potential viewers.

than the value of television programming to consumers.[35]

A pay-TV system, however, because it can charge a direct positive price to consumers for programs, can raise revenues which more nearly approach the value of programming to consumers (as will be discussed in Chapter III) and which exceed the limit for revenues generated by free TV by a seemingly considerable amount. Consider the following argument involving the U.S. television system:

> ...programs which are currently discarded may have had as many as 15 - 20 million viewers. Network programs do not become 'profitable' before passing the 20 million mark, and popular shows command 30-50 million viewers, according to the rating services. A nonpopular show by current standards, if viewers are willing to pay a quarter on subscription television, needs an audience of less than a million to compete with a current show with 30 million viewers on advertising-supported television.[36]

In other words, a small charge[37] for a program to the same free TV audience would allow a pay-TV system to generate considerably more revenues than could be attained in the advertiser-supported television

---

[35]Noll, Peck and McGowan come to this conclusion based on several different types of evidence; see Noll, Peck and McGowan, Economic Aspects of Television Regulation, p. 22 - 23 and 30 - 32.

[36]Minasian, p. 75 - 76.

[37]A "small" charge may indeed be small. Minasian has this to say about the revenues per family received from popular programs: "One of the networks supplied me with data on the total revenues generated by two successful one-hour programs in the 1964 - 65 season...Revenues generated per family were 1.8 cents (less than a penny per viewer) and 3 cents (slightly more than one penny per viewer)." For Canada, in 1965, Table 3 provides data on the amount paid for a one-minute television advertising message per individual home which actually viewed television at a particular time period (as proven by surveys):

system, and thus shows that would normally be too expensive to produce may become profitable. Of course, price elasticity is an important factor in the amount of additional revenues a pay-TV system can generate as a

TABLE 3

COST OF TELEVISION ADVERTISING, C.B.C. AND PRIVATE STATIONS, SIX MAJOR CANADIAN CITIES, FALL 1965

City	Call Letters	Cost Per Home (¢)	Call Letters	Cost Per Home (¢)
Halifax	CBHT	.275	CJCH	.358
Montreal	CBMT	.400	CFCF	.412
Ottawa	CBOT	.329	CJOM	.476
Toronto	CBLT	.263	CFTO	.449
Winnipeg	CBWT	.419	CJAY	.454
Vancouver	CBVT	.353	CJAN	.535
Six Cities Average		.340		.447
Average Cost Per Hour		.340 ¢/min. x 12 min. = 4.08¢ .447 ¢/min. x 12 min. = 5.36¢		

Source: O.J. Firestone, Broadcast Advertising in Canada: Past and Future Growth (Ottawa: University of Ottawa Press, 1966): p. 112, Table 2-1.

Since the maximum amount of advertising time per hour is 12 minutes, the average maximum revenues per family for any program is 4.08 cents for the CBC stations and 5.36 cents for the private stations.

positive price will attract a smaller audience than if offered at a zero price, assuming television programs are normal goods. However, because it is reasonable to expect that viewers would subscribe for programs at a higher price than a few pennies per family which reflects a corresponding amount of advertiser expenditures, more resources can be expected to be drawn into the industry and thus result in a more diversified mix of programming.

In a pay-TV system the nature of programming is determined primarily by consumer demand for television programming rather than primarily according to the productivity of advertisements. The only way in which the two systems would result in the same resource allocation in programming would be if for every program the net revenues obtained from advertising would be exactly equal to the net revenues generated by viewer payments. Since this is very unlikely to happen, we can expect the two systems to provide a different mix in programming.

Not only do free-TV revenues fall considerably short in capturing the value of the intensity of consumer preferences, but also the one dimension of viewer preference that the free-TV system does incorporate in determining the mix of programming it offers - audience size — may be more accurately measured in a pay-TV system. Program audience size in a free-TV system cannot be measured directly and can only be estimated through the various types of television program ratings. These ratings are based on a sample of the viewing population and thus are subject to some degree of error. However, a pay-TV system would be able to directly account for each

viewer because of individual payment for the different programs offered, and thus an extremely accurate measure of audience size can be obtained.

Another reason why pay-TV might possibly offer viewers increased variety in television programming is because a pay-TV system receives its revenues directly from consumers. As a result the mix of programming chosen, in the absence of regulation, is that which will maximize the profits of selling television shows to consumers, and thus programs are chosen that correspond as closely as possible to consumer preferences given the prevailing cost conditions of programs.

An advertiser-supported television system, on the other hand, receives its revenues from advertisers, and thus maximization of profits will result in a mix of programming that is different to the degree that advertising rates are based on factors other than total audience size. Advertising firms will want to sponsor those programs which correspond to the preferences of only those consumers who are likely to be persuaded by the advertised message. Thus, the preferences of some potential viewers will be partially or wholly excluded in the determination of program mix if some shows are produced so as to appeal to specific consumer audiences.

For example, if golf-balls derive great benefit from television advertising in the form of increased sales, the advertisers of golf-balls will want to sponsor a program aimed at maximizing the audience size of golfers as opposed to the audience size of the general population. Thus, even though soap operas may be more popular amongst potential viewers in terms of viewer preference, golf-shows will be offered instead of soap operas because golf-shows will attract more golfers than soap operas would. Thus, the nature of the programming offered in a free TV system

depends in some degree on the mix of products being advertised.

This mechanism for determining output is not conducive to optimization of welfare for the total viewing audience. Because it is advertisers who finance the programs, it would be more accurate to say that the mix of programming chosen is that which maximizes the profits to be derived from advertising and not that programming which maximizes audience size which is often cited as the case. And the maximization of profits from advertising does not necessarily mean programs that are chosen will directly correspond to the preferences of the whole potential viewing audience, because the product that is advertised affects the type of programming offered. (However, it is true that many advertisers will attempt to only reach the general population and not a particular subset, and thus many programs will be produced so as to maximize total audience size.) In contrast, the nature of programming chosen in a pay-TV system is determined directly by the consumer demand for television programming.

A third reason for the ability of pay-TV to offer additional programming stems from the unique physical characteristics of pay-TV which allow for better exclusivity amongst viewers. For example, "adult" entertainment would be more socially acceptable on pay-TV as compared to the present free TV system because children can be much more easily excluded from viewing it.

For all of these reasons pay-TV can be expected to result in a more accurate reflection of viewer preferences, especially if a per-program charge is adopted, than current commercial television programming. Richard Moore, a television program producer, has this to say about commercial television programming:

> Programming is anything but an exact science, and the networks are constantly being surprised by the popular acceptance of show X over show Y...an audience that is paying directly for the television service would undoubtedly make itself heard.[38]

On the other hand, it should be noted that there are some who think the networks do a good job in selecting programs that are the most appealing to the general viewing audience. The actual performance of the current broadcasting system is probably somewhere in between these two extremes.

Using common sense, one might expect pay-TV to provide more diversity in programming. Pay-TV programming similar to existing programming is not likely as consumers will be reluctant to pay for the same type of programming that is already available at a zero price (unless pay-TV outbids free-TV for programs which cannot be duplicated). The following statement made by a Canadian cable operator gives support to this proposition:

> Telecable Videotron's success, without doubt, depends on its type of programming. Compare, for example, our performance with that of a system distributing only programs originating from over-the-air broadcasting stations. In two years, we have captured 48.5 percent of the metropolitan Montreal market while offering our services at a monthly rate of $8.25. Cablevision Nationale Ltee., after 25 years, has about 35 percent of this same market and its monthly rate to subscribers is $6.[39]

---

[38] Moore, "Public Television Programming and the Future", p. 244-45.

[39] R. Jauvin, "A Prescription for cable TV", Insearch, The Canadian Communications Quarterly (fall 1977): 12.

In the U.S., the FCC in its Fourth Report on pay-TV, concluded on the basis of its seventeen-year inquiry into pay-TV that pay-TV would provide a "beneficial supplement" to the conventional "free" broadcasts.[40] It stated that pay-TV broadcasting "is not duplicative of the programming on free TV and that (it) is desired or needed by at least a portion of the viewing public."

Effect of Greater Variety on Consumer Welfare

Whatever the cause, a high degree of program uniformity does exist in the current television system. What has to be assessed, however, is whether an increase in program variety brought about by pay-TV will significantly increase or decrease total consumer welfare. Increased variety will certainly add to the consumer satisfaction gained from viewing television programming, but is the additional cost in terms of the use of scarce resources worth it; could these resources be used elsewhere in such a manner as to yield greater consumer satisfaction? The value of greater variety will rest on the structure of demand and costs.

One of the difficulties involved in evaluating the effect of pay-TV programming on consumer welfare is the fact that the effect of the introduction of pay-TV will be felt throughout the entire television industry, and all changes must be reckoned with. If pay-TV only offered programs that would not normally be available otherwise, it is fairly obvious that consumer welfare would increase as a result. Consumers will pay for these programs only if they add more to their total satisfaction that the purchase of any other product (assuming consumers are rational).

---

[40] U.S., Federal Communications Commission, Fourth Report and Order on Subscription Television 15 FCC 2d (1968): 473, 483-88.

Thus, no television viewers would be worse off because of these programs but some would be better off.[41] These type of programs would be ones that are marginally unprofitable to the "free" broadcasting system, but are profitable to a pay-TV system because of one or more of the reasons stated in the previous section.

However, if pay-TV includes in its package of programming programs which would be available on our conventional TV system if it were not for the existence of pay-TV, the gain to consumers is no longer unambiguous and a more complicated problem exists. The "free" TV system would now be affected in its selection of programming and we can no longer assume that no viewers would be worse off because of pay-TV (some programs previously offered for free would now be offered for a positive price).

The starting point in the analysis is to evaluate whether pay-TV can be expected to provide a range of programming or supplement existing programming such that society is brought significantly closer to the socially optimal degree of program variety. Optimal variety can be viewed as being determined by two major forces — the demand side and the cost or supply side. On the demand side, more variety will be optimal the more widely distributed the tastes of consumers, and the less the substitutability between products (or, in other words, the narrower the distribution of "characteristics" amongst products). On the supply side,

---

[41] This statement assumes that the loss in advertising revenue resulting from the fact that some consumers will decrease their viewing of free TV because of their viewing of pay-TV programs is not sufficient to affect the programming offered on free TV. Noll, Peck and McGowan also make this assumption based on the premise that networks currently earn substantial profits and thus the resulting fall in advertising revenues can be absorbed without a major change in programming: see: Noll, Peck, and McGowan, Economic Aspects of Television Regulation, p. 36 - 37. More will be said on the probable decrease in advertising revenues in Chapter III.

less variety will be optimal the greater the economies of scale in production.[42]

As far as the distribution of tastes is concerned, pay-TV can be expected to better cater to the tastes of consumers than an advertising-supported system, as explained in the previous section. This is primarily because advertising-supported television aims its programs at the mass audience, as revenues are based almost exclusively on total audience size. Pay-TV, on the other hand, because of its ability to charge a positive price for individual programs, can exploit the tastes of consumers on the lower ends of the distribution scale and thus provide additional variety in programming that is justified on a total welfare basis. The extent of this gain in welfare will depend on the distribution of viewing preferences amongst consumers. The gain will be smaller the greater the homogeneity in tastes. As previously discussed, current program uniformity in the free TV system may be primarily the result of a narrow distribution of preferences amongst consumers.

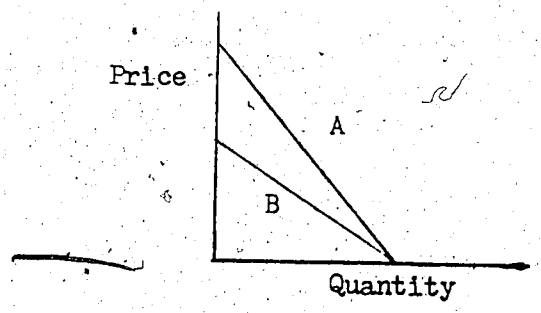
The degree of substitutability between different programs is another important factor in determining optimal variety. The value placed by consumers on variety depends on the relative amounts of satisfaction obtained from preferred and alternative choice program types. Thus, what is required is an empirical evaluation of the amount of additional satisfaction that would be gained by viewers from pay-TV programming that is different than available programming. This is difficult to empirically evaluate, although the eventual success or lack of success of pay-television will shed more light on the amoun- consumers are willing to pay for additional television programming.

[42] Although the economies of scale associated with the joint production of some products may result in more variety being optimal.

The structure of the television programming market is such that pay-TV can result in the provision of a quantity and quality of programming that is closer to the optimum than that programming which would be provided by a free television system. As will be discussed in more depth in Chapter III, the "optimal" quantity and quality of television programming will only be provided if broadcasting revenues are equal to the value of programming to consumers. However, there is no reason to expect advertising revenues in a free TV system to be equal to the social value of programming. Pay-TV, however, can be expected to appropriate a larger fraction of the benefits generated by a program because it obtains revenues directly from viewers by charging a positive price for programs. In such a case, producers' revenues will more nearly reflect consumers' benefits, resulting in the provision of a quantity and quality of programs that is closer to the optimum. There will be a tendency to provide programs up to the point where the cost of programming equals the real value to consumers of programming as reflected in revenues gained from consumers.

In an advertiser-supported television system the only dimension of demand that is important to producers is what the quantity demanded is at zero price. The failure to consider the slope of the demand curve over a range of positive prices may result in the provision of programming which does not maximize total welfare because the intensity of demand is not incorporated in output decisions. Figure 1 depicts the demand for two programs which would generate equal revenues with advertiser support.

FIGURE 1



While an advertiser-supported system would be indifferent between these two programs (assuming equal production costs), a pay-TV system would choose program A, the program with the more inelastic demand. Program A would generate more of both consumer and producer surplus, thus benefiting society as a whole.

One very crude way of determining whether an increase in variety will be of great benefit to consumers is to assess current dissatisfaction among viewers resulting from program uniformity. If the dissatisfaction resulting from the lack of variety is as great as it is often suggested, and viewers express this dissatisfaction by not watching television, then it could be roughly measured. And indeed if this approach is used one might conclude little dissatisfaction exists based on the tremendously large amount of time the average person spends viewing television programs.[43] However, this approach is hazardous because viewers who do not receive their preferred programs may accept existing television programming as simply better than nothing. This argument is reinforced by the fact that at a marginal cost of zero — zero in terms of dollars and not time or sanity — viewers may consume television services even if the resulting satisfaction is very low.

It is even possible that pay-TV can result in a large increase in consumer welfare even if it does not offer a substantially different programming format. Even if duplication is increased, consumers will gain from the increased options available to them. Noll, Peck and McGowan distinguish between viewing options (the number of programs available

---

[43] Average daily television viewing per household is about 6 hours, 11 minutes in Canada; see E.S. Hallman, Broadcasting in Canada, p. 31.

simultaneously to the viewer) and diversity (the number of categories — comedies, westerns, soap operas, etc. — in which programs are offered at a given time).[44] It is argued that viewers value more options (ie. the number of available channels) than diversity because more channels implies a larger number of differing program types and, also, programs within standard categories may not be perfect substitutes for each other and thus viewers will value more options.

---

[44] Noll, Peck, and McGowan, Economic Aspects of Television Regulation, p. 8 - 10.

CHAPTER II

THE RATIONALE FOR REGULATION

The purpose of this chapter is to establish the theoretical basis for the thesis by outlining the conditions under which the government may be economically justified in regulating any particular industry.

If a market left to itself does not attain certain desirable social objectives, then the government may be justified in either replacing the market mechanism as the determinant of industrial output, or modifying the parameters of the market so as to allow for the achievement of these objectives through the functioning of the market — and each of these courses of action will have different economic and political consequences. Thus the issue of regulation boils down into two questions: 1) should the market be regulated to begin with? and 2) if it is to be regulated, how is it best regulated?

The trend in North American regulation over the years has been towards more and more government intervention in industrial markets; each passing year brings more rules and requirements which are imposed upon industry by government. Because regulation has become so prolific it is very important to develop a framework in which to evaluate the rationale for and the costs and benefits imposed by the various forms of regulation.

This thesis is mostly concerned with defining the set of conditions in an industry which might warrant government regulation from an economic standpoint, and then evaluating some of the regulatory issues involving the pay-TV industry.

Although a simple set of rules and guidelines cannot be applied to all regulatory problems, as each industry has its own unique features, valid scientific generalizations can be drawn and useful general guidelines developed. The application of these general guidelines or principles in particular situations must then:

> ...be done on the basis of full consideration of the special characteristics of the industry in question — its technology and other conditions of supply, the nature of its market — and of the varying mix of public purposes, economic and other, that regulation is supposed to serve.[1]

Economic Objectives

The resources which determine our society's capacity to produce goods and services are scarce in relation to the depth and multiplicity of our needs and wants which can be satisfied by the goods and services we produce. Thus, the problem is to choose how to allocate our fixed bundle of resources so they will yield the most satisfaction to society. Efficient resource use is the major objective that economics concerns itself with.

Resource efficiency not only includes the achievement of minimum costs with the given allocation of resources — at one point in time (technical efficiency) and over time (dynamic efficiency) — but also implies optimal product variety (qualitative efficiency), and in the broad sense includes optimal income distribution.

---

[1] A.E. Kahn, The Economics of Regulation: Principles and Institutions, vol. 1: Economic Principles (New York: John Wiley & Sons, 1970), p. 13-14.

The discipline of economics studies the manner in which resource efficiency is brought about:

> It is concerned with the way in which...transacting units or groupings make their economic decisions, and the way in which their several activities are coordinated by economic institutions to make the basic choices dictated by the universal problem of scarcity — what to produce, how and by whom to produce it, and how the product is to be distributed. In Western economies, the coordination is effected through the market system. Therefore microeconomics is concerned with the operation of individual markets — how prices, outputs and distributive shares are determined — and their interrelationships. The criteria of effective performance consist in the desirability of the resulting allocation of resources, the physical efficiency (both statically and dynamically) with which scarce resources are used, and the acceptability of the resulting distribution of income.[2]

As it has developed, regulatory economics is concerned primarily with providing the possible means by which an industry can be led to the achievement of the greatest resource efficiency possible in a regulatory context, with the effects on income distribution usually being ignored. (Regulatory constraints placed on an industry will automatically affect income distribution in some way, but regulatory economics rarely attempts to make policy prescriptions that are directly designed to affect distributive equity. Justification for treating income distribution separately from resource efficiency rests on the assumption that any inequities in income distribution can be alleviated through the taxation system. Lump-sum transfer payments can be used to redistribute income with little or no distortion of resource efficiency.)

It should be kept in mind that the specific performance goals of regulation should be a function of the political process. What

---

[2]Ibid., p. 16.

constitutes "good" performance from an economic viewpoint should be determined by political or ethical criteria, and economists cannot be expected to be any more qualified to make the choice of criteria than anyone else. Economic efficiency labelled as "good" performance implies an underlying value judgement that cannot be made solely on economic grounds.

The role of the economist is to also clearly explain the possible economic costs and benefits, in terms of resource efficiency, of various alternative policies to the decision makers. Regulatory economics is concerned with providing guidance to the policy makers and administrators who develop and apply policies involved with the regulation of industry.

Economic Rationale for Regulation

In this section I will attempt to develop a framework which describes the set of conditions under which government regulation of an industry may be economically justified. The justification will be based on the ability of the regulatory process to improve efficiency in the industry: allocative, technical, qualitative, or dynamic efficiency.[3]

Functioning of the Market Mechanism

Regulation can be economically justified if the economic benefits

---

[3]Allocative efficiency implies (or includes) technical and qualitative efficiency. Allocative efficiency requires scarce resources to be used in such a manner that maximizes the consumer satisfaction gained from those resources — this precludes any waste of resources through inefficient production techniques (technical efficiency), and entails production decisions which are qualitatively responsive to consumer tastes (qualitative efficiency).

of improved performance brought about by regulation exceed the costs of regulation.[4] Improved performance from an economic standpoint in the case of regulation refers to better efficiency: allocative, technical, qualitative, and dynamic efficiency. But the question has to be asked: under what conditions is intervention in a market necessary to bring about this improvement in efficiency?

The partial answer to this question is that intervention may be necessary because market "imperfections" exist. Such an answer implies that the competitive market system with its set of prices perfectly responsive to the laws of supply and demand will result in the most efficient use and allocation of resources and thus maximization of welfare for the entire economy (Pareto efficient). Hence, intervention in a market can only result in improved efficiency if something is preventing the competitive market from functioning properly in the first place:

> ...belief in the invisible hand does not imply that the government has no part to play in the economic system. Quite the contrary. If it is in general true that men, following their own self-interest, act in a way that is of benefit to society, it is, to quote Edwin Cannan, 'because human institutions are arranged so as to compel self-interest to work in directions in which it will be beneficient.' Our task as economists is to help in the devising and improving of those institutions.[5]

---

[4] In addition, because regulation uses scarce resources, an optimal amount of regulation exists depending on the conditions of supply and demand. In other words, regulation should be supplied up to the point where the marginal benefits from regulation equal the marginal costs (assuming that at some quantity the marginal benefits are greater than the marginal costs).

[5] Coase, R.H., "The Economics of Broadcasting and Government Policy" AER Papers and Proceedings, (May 1966), 444.

The competitive market model, while not wholly realistic, is a useful concept in analyzing and prescribing policies for some aspects of our economy:

> For all the great modifications to which market economies have been subjected in practice during the last century, and for all the qualifications that must be attached to the case for such an economy, the competitive market model is still an important measure...descriptive both or reality and of the community's conception of what an ideal economic system would look like.[6]

The economic justifications for regulation will largely be based in this thesis on the ability of regulation to induce an industry to behave in a manner that will more closely approximate competitive market performance.

In a competitive market system prices, costs, and profits reflect relative scarcities and surpluses, and induce the owners of factors of production to use resources that are in the greatest demand relative to supply (allocative efficiency). The prices of goods and services, and the incomes of the factors of production, are the mechanism through which the market system operates. Thus, the decisions as to where resources are allocated and what shall be produced are left to the voluntary decisions of purchasers, whose decisions are guided by product prices and their own wants and preferences.

When the prices of goods and services reflect their true opportunity costs buyers can accurately judge whether the satisfaction received as a result of purchasing any particular product is worth the opportunity cost or sacrifice of the other products forgone.[7]

---

[6]Kahn, The Economics of Regulation, p. 1.

[7]Ibid., p. 66 - 67.

Thus, the greatest possible satisfaction gained from our scarce resources will accrue to consumers if the prices of all products accurately reflect their respective opportunity costs. If prices fulfill this function, consumers will, through their purchase decisions, guide our scarce resources into those lines of production that will yield more satisfaction than any alternative production mix, and thus total satisfaction will be maximized.

Since the true opportunity cost of producing any one product is its marginal social cost, having price equal to marginal cost for all products is needed to ensure the most efficient allocation of resources in the economy (Pareto efficient). If the price for a particular commodity exceeds marginal cost, some consumers who would normally buy the product will refrain from making those additional purchases because the price exaggerates the true opportunity cost or sacrifice if all other products are priced at their marginal cost.

In a perfectly competitive market, competition for customers will induce firms to drive prices down to a product's marginal cost, a necessary condition for the most efficient allocation of resources. (However, as will be discussed later under "second best" considerations, prices equal to marginal costs in any one industry is a necessary but not sufficient condition for an efficient allocation of resources.)

As far as technical (static) efficiency is concerned, in a competitive market firms which fail to operate at the lowest possible unit cost for their given output will incur losses and thus eventually be driven from the market. Competitive firms cannot afford any "X-inefficiencies" or economic waste. Competition for customers will weed out the inefficient firms.

In regards to the issues of qualitative and dynamic efficiency, however, the performance of a perfectly competitive market system is more ambiguous.

There exist some economic arguments against a purely competitive market, and these must be examined also. One argument is made on the grounds of achievement or lack of achievement of dynamic efficiency. Competitive firms may not be willing to bear the risks and costs of technological innovation because they are not protected from competitors who may copy their innovations, and thus they would not be able to benefit, in the long run, from their innovations. Also, competitive firms may not be able to generate enough funds to allocate to research and development. (And, in the end, the rate of technical progress is probably more important than resource allocation efficiency[8]).

On the other end of the scale, it is often argued that a monopolist will tend to be lethargic in its operations and thus also is not likely to become as dynamically efficient as possible. Because a monopolist faces little or no competition, its incentive to introduce technological innovations and increase productivity may be reduced. Competition may be needed to induce innovation designed to achieve better efficiency. Thus, the best solution to the problem of dynamic efficiency lies somewhere in between the two polar extremes of pure competition and pure monopoly.

A second economic argument against a perfectly competitive system is

---

[8]For an explanation see: F.M. Scherer, Industrial Market Structure and Economic Performance (Chicago: Rand McNally, 1970), p. 21.

that optimal product variety (qualitative efficiency) may not be attained — competition may not result in the provision of the socially optimal degree of product differentiation. The effects of market structure on product variety is an important issue when considering regulation of the pay-TV industry because the value of pay-TV to society rests largely on its ability to provide a more socially optimal degree of product variety in television programming.

The theory of product variety is traditionally associated with Chamberlin's theory of monopolistic competition, which predicts a trade-off between increased variety and allocative efficiency. This situation arises because a monopolistic competitor associated with a slightly differentiated product in a competitive industry will eventually produce an output for which average cost is not minimized and price exceeds marginal cost.[9]

Because of the downward-sloping demand curve it faces, a monopolistically competitive market itself cannot be expected to provide optimal product variety. Using the concept of total surplus, M. Spence in a recent article has shown that under monopolistic competition (or monopoly) a firm's selection of product characteristics is likely to be biased away from the social optimum because of a divergence of social and private benefits resulting from possible differences in the marginal and average valuations of q by consumers.[10]

---

[9] Ibid., p. 13 - 15. Scherer provides an explanation of how this comes about.

[10] M. Spence, "Monopoly, Quality and Regulation," Bell Journal of Economics (Autumn 1975): 417 - 29.

The failure of the market arises from the inability of prices to convey information about the value attached to quality by inframarginal consumers; thus the benefits to a firm because of a change in quality may not be a measure of the true social benefits unless the marginal consumer is average or representative, and there is nothing intrinsic to the market that guarantees that the marginal purchaser is indeed representative.

However, much dissatisfaction exists among economists with regards to the way in which the theory of monopolistic competition deals with the problem of product variety. Kelvin Lancaster has developed an approach to consumer behavior that provides an alternative approach to the issue of the social optimality of different degrees of product differentiation. This intuitively appealing approach is different from the traditional treatment of consumer behavior in that it postulates that the utility which consumers attempt to maximize is derived from the properties or characteristics of goods rather than directly from the goods themselves. One of the assumptions used in this approach is that goods, in general, possess more than one characteristic, and many characteristics will be shared by more than one good. Since this approach conceptually views a good as having multiple characteristics, it is better suited to dealing with the problem of product differentiation.

This "characteristics" approach to consumer behavior was first introduced by Lancaster in its general form in 1966.[11] Nine years later Lancaster published another article in which he used this theory to examine the problem of optimal variety in a world in which every consumer

---

[11] K. Lancaster, "A New Approach to Consumer Theory", Journal of Political Economy (April 1966): 132 - 57.

knows exactly what he prefers and how to achieve personal optimality within his constraints.[12] One of the conclusions that emerged from his analysis is that there exists a socially optimal degree of product differentiation which would be achieved under perfect competition, but only if conditions of increasing returns to scale did not exist. (It is very intuitively understandable that conditions of increasing returns will restrict the number and variety of products that it is desirable to supply.)

Thus, on a theoretical level, the effect of market structure on optimal product variety is somewhat inconclusive if not confusing at this point in time. While the competitive model is useful in analyzing the functioning of some aspects of our economy (such as resource allocation), it simply cannot be used in analyzing and evaluating other aspects of the economy (such as qualitative or dynamic efficiency).

Another fault of the purely competitive system from a political and economic viewpoint is its potentially unstable effect on employment:

> It is conceivable that the hair-trigger price adjustments of purely and perfectly competitive markets could intensify tendencies toward instability, making it more difficult to combat the waste and human misery of cyclical unemployment through fiscal and monetary policy measures.[13]

---

[12] K. Lancaster, "Socially Optimal Product Differentiation," American Economic Reivew (September 1975): 567 - 85.

[13] F.M. Scherer, Industrial Market Structure and Economic Performance, p. 21.

Presumably workers prefer stability in employment opportunities to unstability (all other things remaining constant).[14]

As far as this performance goal is concerned, the monopolistic industry organization may be the ideal form of organization since it is more conducive to stability. However, this issue is far from being resolved, and there is some evidence to the contrary. Roger Sherman has surveyed four major empirical studies which explore this issue and he tentatively concludes that concentration is associated with greater fluctuations in total working hours.[15] However, any attempt to explain the relation between employment risk and monopoly power is very difficult because firms will generally make a greater effort to keep more skilled workers in periods of demand instability (usually because such workers possess more of the firm's investment in training), and it happens that firms in more concentrated industries tend to employ more highly skilled workers.[16] Thus, the success of attempts to correlate concentration and employment instability will depend on how well they deal with the problem of systematic differences in worker skills.

---

[14] D.S. Smith has produced some empirical support for the fact that workers must be compensated for employment instability (employment risk) in the form of higher wage rates (see D.S. Smith, "Concentration and Employment Fluctuations," Western Economic Journal, September 1971: 267-77). Smith used the standard deviation in the ratio of weekly production hours to the average weekly production hours in 73 industries to represent employment risk, and found that workers who bear more employment risk apparently insist on a higher average wage rate.

[15] R. Sherman, The Economics of Industry, (Boston: Little, Brown and Company, 1974), pp. 205 - 07.

[16] see L.W. Weiss, "Concentration and Labor Earnings," American Economic Review (March 1966): 96 - 117.

Because the market system does have these potential faults of instability, sub-optimal product differentiation and dynamic efficiency, regulation aimed at removing these faults may be justified even if no static "imperfections" in the market exist. Thus, regulation designed to bring about increased dynamic efficiency or stability, or a more optimal degree of product variety, may be economically justified.

From a social viewpoint, the competitive market system also has its faults. Perhaps the most important criticism arises from the fact that votes in a market system as to where resources are allocated are based on dollars, and thus, because individuals are not equally wealthy to begin with, some individuals will have more influence in the market than others. A competitive economy maximizes the benefits gained from our scarce resources, but these benefits are the benefits which accrue to the individuals with money.

The actual mix of goods and services produced in our economy, given the state of technology and a fixed supply of resources, is dependent on consumer tastes and the distribution of income. Those individuals with relatively high income have a greater influence than low income individuals on output and the employment of resources because they have more "dollars" (representing property rights to the use of scarce resources) to bid for the output of the economy. On the other hand, at least on a theoretical basis, decisions in the public sector are made through the political process where each individual has the same influence through his political vote, and thus government participation in an economy might be justified on the grounds of equity.

Competitive allocations may also involve considerable inequalities in income distribution, and thus while the competitive market may achieve Pareto efficiency, this may not be a socially optimum result. The pattern

of resource ownership in the economy may be such that the resulting Pareto efficient pattern of production and its pattern of income distribution is less acceptable to society than a Pareto inefficient composition of output which results in a different pattern of real income distribution.[17]

Second Best Considerations

Pareto provided conditions necessary for society to reach a position of optimum welfare through resource allocation — conditions which are met in a perfectly competitive economy. However, the marginal conditions necessary for maximum welfare are global conditions which have to be satisfied throughout the entire economic system. There must be conditions of perfect competition in all markets in the economy if a Pareto efficient allocation of resources is to result.

However, in reality all of the marginal conditions cannot possibly be met. Because of the pervasive market "imperfections" in the real world,[18] a Pareto optimum is unattainable. We are thus faced with a suboptimal position about which Paretian conditions give no guidance.

The important policy question then becomes whether the "second best" position calls for a policy aimed at fulfilling as many of the Paretian conditions as possible in order to get nearly to a Pareto optimum.

---

[17] However, this problem can be solved through means other than manipulating the type of market system in operation. The market system can be left to operate and attempts to improve the unequal distribution of wealth can come about through such measures as transfer payments amongst individuals.

[18] Not only are market imperfections prevalent in all markets, but there are large segments of the economy to which the competitive model is not wholly applicable: public utilities, government sectors, households, non-profit sectors.

Are movements toward the competitive norm desirable in terms of resource allocation efficiency? Even if we can never achieve pure and perfect competition in all industries, do we gain by moving towards this ideal and regulating so as to alleviate market imperfections?

In a formal theoretical proof, Lipsey and Lancaster (1956) have demonstrated that if deviations from marginal cost pricing exist anywhere else in the economy, we cannot make the general proposition that rules designed to improve the "competitiveness" of any single sector will be desirable from a resource allocation standpoint.[19] While this approach, and others like it, are negative in nature, attempts to establish more positive general second-best conditions have met with little success.

However, even though it appears that we can make no general predictions of the effect on allocative efficiency of a movement towards marginal cost pricing in any one sector of the economy, this does not necessarily rule out logical solutions to the pricing problem in specific situations. In the words of A. Kahn:

> The existence of pervasive imperfections in the economy greatly complicates the problem of efficient pricing. In the author's view in principle it does not make solution impossible in specific situations, nor does it make it practically impossible in such instances to make the type of informed piecemeal decisions policy makers must inevitably make about how far and in what directions to qualify the basic rule of marginal-cost pricing.[20]

---

[19] But, there is little doubt that in most circumstances markets, even if highly imperfect, allocate resources more efficiently than their feasible alternatives, such as direct administrative controls. The decisions in a market economy are made by the individual buyers and sellers in the market, and the market permits a very sensitive expression of individual preferences.

[20] A. Kahn, The Economics of Regulation, p. 70.

This position taken by Kahn, however, may arise from a defensive attitude resulting from the fact that much of the economic analysis in his book would have little relevancy if in specifi situations it would not be possible to prescribe economically optimal pricing principles or policies. W. Baumol shares an opinion similar to Kahn:

> ...there looms most menacingly the injunction of the theorem of the second best: Thou shalt not optimize piecemeal. But I would argue that in practice this admonition must be softened less otherwise all effective policy be stultified. I would propose, instead, that one should shun piecemeal ameliorative measures that have not been sanctioned by careful analysis and the liberal use of common sense. Many policies may plausibly be expected to yield improvements even though things elsewhere are not organized optimally.[21]

The technological and institutional characteristics of the pay-TV industry are such that it might provide one of those specific situations in which pricing policies can be designed to result in a more efficient allocation of resources. Pricing and efficiency in the pay-TV industry will be discussed in the next chapter.

While the effects on resource allocation are uncertain, other benefits accrue by moving towards the competitive norm — and it appears the government believes these benefits to outweigh the costs. Government regulation in our economy influences the functioning of most private sectors through a variety of means (such as the prohibiting of unfair competition, the providing of market information, the enforcement of contracts, providing tariff protection, subsidies, grants, industrial legislation, and so on). The majority of these influences are designed

---

[21] W.J. Baumol, Welfare Economics and the Theory of the State (Cambridge: Harvard University Press, 1965), p. 30.

primarily to operate at the periphory of the market, maintaining the institutions within whose framework the free market can continue to function. Their role is to remove the imperfections of competition, not to supplant it.[22] Thus, an inherent belief in the desirability of the competitive market system is expressed.

While the government has taken upon itself the role of actually directing and managing the operations in a few industries, the major emphasis of regulation in our economy is felt in attempt to adjust certain parameters of markets to produce desired behavior, while leaving the basic decisions of the market to the private components who make up the industry. In those cases where the government has replaced the market mechanism as the determinant of industrial output, regulation is by and large designed to achieve the same results as would be produced through effective competition.

The history of the North American continent is evidence of the fact that our society as a whole considers the competitive market economy as the ideal economic system ("ideal" at least in some respects), for social as well as economic reasons.[23] Our society profoundly values the institution of private property and the private initiative and control of economic activities.

As we move to a more competitive situation, technical or productive efficiency almost certainly does improve in general. Without competitive

---

[22]Kahn, The Economics of Regulation, p. 2.

[23]The competitive system is socially desirable in the sense that it allows for the dispersion of power, a desirable result to many.

pressures X-inefficiencies often arise. Almost any businessman will tell you that competitive pressures are very successful in motivating firms to produce a better product and at lower costs:

> Competition will weed out the inefficient and concentrate production in the efficient; it will determine, by the objective test of market survival, who should be permitted to produce; it will force producers to be progressive and to offer customers the services they want and for which they are willing to pay; it will assure the allocation of labor and other inputs into the lines of production in which they will make the maximum contribution to total output.[24]

A move to a more competitive industrial scene will also help replenish some of the motivation that is being lost by businessmen as a result of the increasing government intervention in today's society allows for less individual initiative upon which our economy is very dependent.

Also, the movement towards a more competitive market system does not necessarily have to mean a sacrifice in technological innovation as discussed previously. If it is true that the degree of technological innovation decreases as the degree of competition increases, other means can be used to achieve dynamic efficiency, such as the use of patents and tax incentives that encourage research and development.

Actually, the move to a more competitive market system can be expected to increase the rate of technological innovation in certain cases where an industry is characterized by a high degree of monopoly power. F.M. Scherer, in a review of the empirical evidence on technological

---

[24] Ibid., p. 28.

innovation, concluded that:

> A little bit of monopoly power, in the form of structural concentration, is conducive to invention and innovation, particularly when advances in the relevant knowledge base occur slowly. But very high concentration has a favorable effect only in rare cases, and more often it is apt to retard progress by restricting the number of independent sources of initiative and by dampening firms' incentive to gain market position through accelerated research and development.[25]

In a survey of the literature on market structure and innovation, Kamien and Schwartz came to a similar conclusion:

> Little support has been found for the standard hypothesis that R & D activity increases with monopoly power ...A new empirically inspired hypothesis has emerged to the effect that a market structure intermediate between monopoly and perfect competition would promote the highest rate of inventive activity.[26]

Market "Imperfections"

As already discussed, government intervention in a market usually can only result in improved efficiency if something is preventing the competitive market from functioning properly in the first place. These market "imperfections", or situations which may justify regulation, can be conveniently grouped into seven categories:

1) decreasing costs
2) indivisibilities
3) externalities
4) public goods
5) ill-defined property rights
6) equilibrium
7) monopolizing elements

---

[25] Scherer, Industrial Market Structure and Economic Performance, p.377.

[26] M. Kamien and N. Schwartz, "Market Structure and Innovation," Journal of Economic Literature 13 (1975): 32.

It should be kept in mind that not all market imperfections may justify intervention in the market by government. For some imperfections there is no reason why the government or some external agent would be able to alleviate the problem any better than the market itself. For example, factor immobility is certainly a market imperfection, but the cost of making these factors of production "mobile" should be considered as part of the total cost of these factors of production, and thus is internalized into the decisions of the market participants. Factor immobility alone cannot justify government intervention.

Likewise, another market imperfection that is often believed to be a justification for government intervention, but is not really a justification, is imperfect information. The perfectly competitive model unrealistically assumes that perfect or complete information is available to all participants in the economy. However, although information is generally not perfect in the real world, again the cost of obtaining better information is internalized by the market as the benefits from better information accrue to the decision-making units. (Acutually, information should be considered as a market in itself — a demand and supply exists for information. The government would only be justified in entering the information market if conditions such as externalities or indivisibilities existed; the lack of information to market participants in itself does not justify government regulation in an economy, but imperfections in the information market may.)

Decreasing costs

The first of these market "imperfections" refers to the existence of economies of scale which signify that the technology of an industry is such that unit costs for a firm fall as output is increased up to a certain point. These economies can arise from a variety of sources, such as the increased specialization of labor that is possible with greater output.

If scale economies are significant in that costs decrease quickly as output is increased over the relevant range of industry output, the emergence of one or only a few firms in the market is likely, as one or only a few firms the size of the market can produce most efficiently and thus underprice and drive out any competition. And in an industry in which economies of size are significant, large firms are desirable in the sense that activities in the industry can be most efficiently conducted by large firms — a firm which produces all or most of the industry's output is necessary to achieve the lowest costs per unit of output possible with the given demand.

The problem of "natural monopoly" or "natural oligopoly" is more prominent in Canada than in the U.S. because of the smaller market size in most Canadian industries. In many industries the minimum optimal scale of production allows for only a few firms which are large enough to take advantage of all relevant production and distribution economies of size. In other words, the market is so small that a few firms of optimal scale will in many cases comprise most or all of the market.

While one or only a few firms may result in the lowest possible average unit costs, such a structure also implies monopoly power, and

thus the industry can no more be expected to function as a competitive industry would. The marke can no longer be relied upon to achieve the desirable performance s described earlier, and government regulation may be needed to ensu that these goals are attained.

However, decreasing costs may provide justification for government intervention, it is not an entirely sufficient condition in itself for regulation. Each industry that is characterized by decreasing costs has to be individually assessed in determining whether it will perform in a relatively satisfactory manner if left to itself, and there are many factors which have to be taken into consideration before a decision can be made as to whether regulation is economically justified:

> ...the analysis of what is at best a most intricate and intractable problem, (is) that of identifying positions of high levels of market power, assessing their significance in an interdependent market economy and attempting to devise remedies.[27]

From an economic viewpoint, while high concentration levels will be necessary to achieve minimum average costs, high concentration may also lead to X-inefficiencies associated with the lack of competition. Monopoly power in any one industry may also induce non-competitive behavior in those markets which are input suppliers or output buyers to the industry.

From a political viewpoint, industries characterized by decreasing costs are regulated so as to take advantage of any relevant scale economies while avoiding the exploitation of consumers and other industries

---

[27] Canada, Dept. of Consumer and Corporate Affairs, Dynamic Change and Accountability in a Canadian Market Economy (Ottawa: 1976), p. 127.

which interact with the industry. Exploitation is a major concern, especially for industries which provide essential goods and services which have a low price elasticity (electric utilities, telephones, etc.). The potential for the misuse of the monopoly power does exist, and firms may, if unregulated, charge excessive prices, restrict output, or discriminate among buyers — behavior that does not result from a competitive situation.

However, the fact that an industry is heavily concentrated does not imply that firms will automatically restrict output and charge excessive prices.[28] The degree to which a firm with monopoly power, in an attempt to maximize profits, will restrict output to a level below what the output would be in a competitive industry and correspondingly raise prices, depends on numerous factors. The degree to which higher prices will attract other firms into the industry, technological constraints on output expansion, the product's price elasticity, the cross-elasticities of demand for substitute products, and the degree of cooperation between the firms in the industry, are all factors which will influence pricing behavior. Also, prices may be intentionally kept down to the degree a firm adopts a sales or "market share" maximization objective function as opposed to a profit maximization objective.

---

[28] In one interesting empirical study of the relationship between industrial concentration and allocative efficiency by G. Krouse, the regression results from a sample of 115 firms did not confirm the hypothesis that firms in concentrated industries restrict their output and correspondingly raise prices relative to firms in unconcentrated industries. There were no systematic differences between concentrated and relatively unconcentrated firms in output behavior, according to his empirical model; see G. Krouse, "Measuring Allocative Efficiency With Technological Uncertainty," Journal of Finance 31 (May 1975): 685 - 700.

The importance of the product or service and the resulting degree of demand elasticity is a very important factor to be taken into consideration. If the product is deemed a necessity (such as electricity), the resulting inelasticity of demand is more likely to require government regulation to prevent exploitation of consumers than for a luxury good or product with a relatively high demand elasticity. Also, whether the fixed costs of the "natural monopolist" are common or joint in nature further complicates the problem.

It should be kept in mind that the supernormal profit earned from "excessive" prices may or may not be more economically desirable than lower consumer prices. These profits may lead to benefits such as research and development, investment, or larger dividends to shareholders. The advantages resulting from these excess profits have to be weighed against the disadvantages. If a firm can use excess profits in a way that increases welfare more than the increase in welfare to consumers which results from lower prices, society may consider some excess profits to be desirable. In short, a complicated problem exists.

Summing up, if an industry characterized by decreasing costs is to take advantage of the scale economies and operate more efficiently (from a cost standpoint), it should be regulated so as to avoid the inefficiencies associated with monopolist activities and to assure that firm size is the optimum in terms of technical efficiency. To ensure optimal firm size, government franchises or licenses can be used to limit entry into the industry. The pricing problem is a more complicated issue, and is traditionally handled by regulating prices so they cannot exceed cost (including a "fair" return).

Indivisibilities

Somewhat related to the argument of decreasing costs as a rationale for regulation is the presence of indivisibilities in a market. Indivisibilities arise when the individual consumption units of a commodity are smaller than the units of production. The total product cannot be provided by producing smaller products to be sold individually to consumers. Production is "indivisible".

A highway is an example of a product characterized by indivisibilities. The construction of a highway is an all-or-nothing proposition, but the units of consumption are individual crossings along the highway by the individual vehicles who use it.

Production indivisibilities may warrant government regulation if the production size is large relative to the market. Using our example, only one highway is needed between two cities. The construction of a second highway, even if profitable in terms of direct costs and benefits, would use up resources unnecessarily. Thus, as in the case of decreasing costs, regulation may be needed to ensure the optimum quantity of production while ensuring that any resulting monopoly power is not misused.

Also, production indivisibilities give rise to pricing and resource allocation problems. Any positive price charged for a product whose individual consumption units are smaller than the units of production will result in the exclusion of some potential consumers even though there may be no corresponding reduction in resources. In such a case, external intervention may be necessary to ensure pr[illegible]erformance.

Externalities

An important market imperfection which may warrant regulation of an industry is the existence of externalities. Externalities or "third-party effects", as the term implies, occur when an outcome of an activity carried on by one decision maker affects an external party to the initial activity. For example, night flights by jumbo jets produce the unintended externality of causing residents near airports to lose sleep. However, in deciding how many jumbo jets to allow the use of an airport, the costs of lost sleep and annoyance may not be taken into consideration because the residents near airports are external to the decision-making activity. This is an example of external costs or diseconomies, but externalities can also involve benefits or external economies.

In the ultimate sense, as pointed out by James Quirk, externalities arise from the fact that the world we live in is limited in size and natural resources, and thus any activities tend to "crowd in" on the range of activities available to others, whether intended or not.[29] We all live in the same world with a high degree of interaction, and the importance of externalities grows as our population increases and as we crowd ourselves into urban centers.

The essence of externalities is that their costs or benefits are not reflected in market prices, and thus the decision of the unit creating externalities does not take their effect into account. In the case of negative externalities, the decision-making unit will choose an output

---

[29] J.P. Quirk, Intermediate Microeconomics (Chicago: SRA, 1976), p. 318.

that is above the economically optimal output becuase the marginal benefit received by the decision-making unit from an economic activity is dependent on a price which does not reflect the full costs of the economic activity. In other words, the true cost of resources allocated to that particular activity is understated, and as a consequence too many resources are allocated.

Likewise, in the case of positive externalities an output will be chosen that is below the economically optimal output level. (In fact, an industry may not even exist even though it would if the positive externalities it created were somehow internalized.)

It can be argued that social welfare would be increased if decisions are modified to take into account their external effects, and this is why an external agent such as the government can be justified in intervening in a market. The ideal competitive system described earlier rests on the assumption of the absence of any external costs and benefits.

Thus, government regulation of a market may be justified if it can manipulate the parameters of the market so that any externalities are internalized. In other words, regulation should be designed to impose social regulations on an industry or to subsidize or tax an industry so as to bring the full "external" effects into the decision-making process of the responsible parties.[30]

---

[30]The argument is sometimes raised that because it is almost impossible to determine and evaluate all of the externalities attendant to an economic activity, the internalization of any one externality by government intervention cannot be justified. For example, our society and economy is so complex that there are numerous externalities created by a pay-TV industry, and it would be impossible to evaluate and weigh all of these externalities in determining the degree to which the pay-TV industry should be taxed or subsidized to internalize these effects. The proponents of this argument claim that you cannot justify the internalization of, for example, the negative externality of adverse effects on our culture by pay-TV because the value of all externalities taken together may be positive.

However, while in principle this is an economic justification for regulation, the actual assignment of values to external benefits and costs in any particular case is a normative procedure, and thus the political process may also enter in.

Public goods

A pure public good is characterized by the fact that its benefits are available to everyone; no one can be excluded from the benefits of a public good and its consumption by any individual does not reduce the available amount for others. It is thus a special type of consumption indivisibility. The classic example of a public good is national defense. If defense is provided for a country, everyone automatically benefits. Individuals cannot be excluded from its benefits whether they pay for it or not,[31] and consumption by any one individual does not affect the consumption of others.

Because of these characteristics a public good is "unmarketable" even though a significant demand may exist for the good. A private firm would not supply a public good because individuals would not be willing to pay the firm a positive price because they would be able to consume the good at a zero price once it is produced at someone else's expense. Thus, because individuals would not be willing to pay for the good what it is worth to them, it would not be provided unless collective action is taken. Government intervention may therefore be justified if the market fails to supply the good when a sufficient demand exists. Regulation may take the form of general compulsory tax revenues paid to the producers of the public good.

---

[31] This is not exactly true as, depending on where military operations are centered, some parts of a country will benefit more than other parts.

Ill-defined property rights

Government intervention in the economy may also be justified if ill-defined property rights exist. In such a case, overuse of scarce resources may be prevented if they are made "public" and allocated by the government. Examples of ill-defined property rights are ownership of the airwaves, of the water in a river, or the air which surrounds a city.

Fluctuating equilibrium

If competition in a market is possible, but conditions are such that market equilibrium is constantly changing because of highly variable demand or supply, regulation may be warranted if the swings in output and prices are too costly to consumers and producers. Government intervention aimed at stabilizing output and prices, and protecting the equity of firms in the industry, may provide economic benefits which exceed the costs of regulation.

While fluctuating equilibrium is not really a market "imperfection" in a static sense, it certainly can be considered as so in a dynamic sense. The greatest difficulty in evaluating this market imperfection as a justification for government regulation lies in the complicated nature of the problem — there are numerous factors to be evaluated.

A significantly variable equilibrium is most likely to occur in industries characterized by cyclical or random fluctuations in demand. In this case, excess capacity will exist in periods of low demand, and the result may be bitter competition amongst the firms in the industry for the remaining customers. Severe price competition is likely (the degree to which depends on the ratio of fixed to total costs) as firms

will seek a larger customer base over which to spread overhead costs. Firms will only be willing to leave the industry when literally on the brink of collapse, as increases in demand can be foreseen when the cycle swings to a more favorable period.

Destructive competition of this sort may have several potentially undesirable results. One of these undesirable results involves the loss or deterioration of capacity needed when demand recovers. Although this might be desirable if the most inefficient firms are the ones forced to leave an industry in a prolonged slump, in actuality the firms forced to leave will be those which are weakest financially — and those weakest financially will not necessarily be the least efficient producers (although some type of relationship probably exists between the two). For example, a firm may have a greater ability to borrow because it is a branch plant of a large foreign corporation, and thus it may be in a better position to survive a prolonged slump than a newer local firm which may even be more efficient from a technical viewpoint. (However, as F.M. Scherer points out, the loss of capacity because of a slump in business conditions may not be as serious a problem as it appears at first.[32] Scherer explains that seldom does an outright dismantling of production facilities occur after a bankruptcy. Rather, plants are normally purchased by another solvent firm at bargain prices, and that firm is eventually able to restore the plant to operation.)

Another undesirable outcome of fluctuations in demand and their resulting adjustments in firm numbers and size, is the numerous inefficiencies associated with these adjustments. Resources are used in laying

---

[32] Scherer, Industrial Market Structure and Economic Performance, p. 200.

off workers and then rehiring them, in the storage of equipment, the loss in organizational continuity, and so on.

A variable equilibrium is also likely to result in more concentration in an industry. Cut throat competition during a slump may induce smaller or weaker firms to merge with stronger firms, something they would not normally do under more favorable conditions.

Government regulation in the form of some sort of price stabilization program may help remove some of the undesirable results of a fluctuating equilibrium, but such an action has several costs associated with it which must also be evaluated. For example, any tampering with the price mechanism may reduce resource and technical efficiency in the long run. Thus, while the potential certainly exists for the benefits of government regulation to exceed its costs, many factors have to be evaluated in determining whether government intervention is justified in any particular case.

Monopolizing elements

It is in the best interests of the firms in any industry to coordinate their actions rather than to independently compete against each other. Higher profits can be earned if cooperative policies are pursued or if concentration is increased through such means as merger. Therefore, even though coordination is difficult (especially as the number of firms in an industry increases), there will always be an intrinsic tendency in an industry to achieve a greater degree of coordination or monopoly power.

This monopolizing element will exist in all industries, although, because of the many problems faced in coordination, it is likely to be threatening only in the more concentrated industries. Government

intervention may therefore be justified as a preventive measure against this potential usurper of the competitive system.

However, the best solution to this problem lies in the universal application of competition policy to the whole economy, and not through individual regulation of each industry. (Selective regulation would amount to detailed central direction by the government; general policies would leave the essential functions of the economy to the individual private market components.)

Efficiency of the Regulatory Process

The efficiency of the regulatory process must be assessed as part of the procedure in determining whether the government is justified in intervening in markets. Even if the necessary conditions exist, such as significant scale economies or externalities, which would warrant some kind of external intervention in the market, the inability of the regulatory process to efficiently deal with these problems may give rise to the fact that society is better off if the government doesn't intervene in the first place. Because the regulatory process cannot always achieve the ideal theoretical solution, it is possible that regulation may not be justified in situations which may theoretically warrant regulation.

Therefore, what has to be assessed is the efficiency of the regulatory process, and there is considerable evidence which points to the fact that a large part of government regulation is suprisingly ineffectual. The regulatory literature is full of biting criticisms of the regulatory

process:

> The revival of serious discontent and of research on public regulation in the 1960's has now bred several specialties...The new thought and research already amounts to a devastating indictment of regulation as possibly an intractable vehicle of waste and deception.[33]

It is the opinion of many that, in general, instead of encouraging competitive behavior regulation has often allowed and perpetuated inefficiencies.

One of the reasons for this general inefficiency of the regulatory process is that regulation theory and its policy prescriptions must almost always be tempered with practicality, as economically ideal principles fall short of providing workable rules for regulators. As a result:

> ...even the most sophisticated and conscientious effort to apply these principles inevitably involves large doses of subjective judgement and, at the very best, can achieve only the roughest possible approximation of the desired results.[34]

Thus, while the concept of regulation is valid, the means of regulation policy need improvement. A.H. Ende, a regulator with the FCC over twenty-five years has this to say:

> As a former regulator, whose horns may yet be visible, I must confess that I have a strong, adverse visceral reaction whenever I encounter dogmatic attacks upon or denigration of the regulatory processes..,I believe that the concept of regulation is as sound today as it ever was and probably much more important in our highly mechanized and integrated society than ever before. What must be changed, and changed promptly and decisively, is the method by which we regulate and the perspectives we use in carrying out the regulation.[35]

---

[33] W.G. Shepherd and T.B. Gies, eds., Regulation in Further Perspective, p. 1.

[34] Kahn, The Economics of Regulation, p. 182.

[35] A.H. Ende, "Administrative Reform and the Regulatory Processes," in Public Utility Regulation, eds., W. Sichel and T. Gies (Lexington, Mass.: Lexington, 1975), p. 77.

The bureaucractic red tape involved with developing and implementing regulation policies also considerably reduces the efficiency of the regulatory approach:

> ...it...is indirect and time-consuming — hearings and rulemaking procedures can and do take years to conclude and may be appealed to the courts, causing further delays. The delays inherent in regulatory procedures are compounded by budgetary and staffing limitations. Thus, regulatory agencies are scarcely in a position to be fully effective in their dealings with all the entities subject to their jurisdiction.[36]

Another recurring criticism of regulation involves the inability of the regulatory process to adjust and evolve along with changes in the industry:

> However fluid an organization may be in its beginning, it must inevitably adopt certain policies and organizational forms which condition its thinking and limit the range of its policies. Within limits, the regulatory commission may search for what is in the public interest, but it is not likely to find acceptable any solutions which imply fundamental changes in its settled policies.[37]

As an industry changes and evolves the optimal regulatory policy will also change; but the political and regulatory process is structured in such a manner that change is not readily or easily brought about.

R. H. Coase goes on to say that, in addition to this lack of adaptability, the regulatory process often becomes a perpetuator of industry motives rather than a champion of the public interest. (Such behavior is the major postulate of the capture theory of regulation which states that

---

[36] Ibid., p. 72.

[37] R.H. Coase, "The Economics of Broadcasting and Government Policy," AER Papers and Proceedings 56 (May 1966): 442.

regulatory agencies tend to be "captured" by the industry they regulate.) Economic decision-making units are assumed to act in a manner which is consistent with their own self interest. If this is true, then they will allocate resources towards changing, for their own benefit, the institutional rules which govern them. Thus, firms in an industry can be expected to allocate resources towards securing control of governmental machinery to improve their own position, an action which may be contrary to the interest of the rest of society.

An evaluation of the efficiency of the regulatory process is basically an empirical matter, and an overall assessment will not be made here. However, there is sufficient evidence to suggest that the ability of the regulatory agency to regulate in an effectual way may be an important factor for evaluation in determining whether an industry should be subject to regulation in any particular case.

Summary

Table 4 briefly summarizes the cases for government regulation of an industry.

TABLE 4

REGULATION DESIGNED TO IMPROVE ECONOMIC EFFICIENCY

Market Imperfections	Type of Regulation	Benefits of Regulation	Cost of Regulation	Factors to be Evaluated
(i) decreasing costs	entry regulation designed to achieve optimum firm size (-licensing) (-public ownership)	greater technical efficiency	X-inefficiency Costs of legislation and administrative body	-industry concentration -shape and extent of scale economies
	price regulation (price ceiling)	lower consumer prices	-costs of legislation and administrative body -loss of profits -reduction in resource efficiency	-price elasticity -entry barriers -cross-elasticities of demand for substitute products -technological constraints on production -degree & type of cooperation or competition amongst the firms in the industry
(ii) indivisibilities	entry and price regulation	increased resource efficiency	costs of legislation and administrative body	
(iii) externalities	subsidization or taxation, or social regulations	increased resource efficiency	costs of legislation and administrative body	
(iv) public goods	subsidization or government provision of goods	provision of good or service	costs of legislation and administrative body	
(v) ill-defined property rights	government allocation of resources	more efficient use of resources	costs of legislation and administrative body	
(vi) unstable equilibrium	price stabilization	-preservation of capacity	costs of legislation and administrative body sacrifice in longrun efficiency	-ratio of fixed to total costs -industry concentration -degree of variability in cyclical conditions
(vii) monopolizing elements	(best solution to the problem lies in universal application of competition policy)			

* Regulation designed to improve: dynamic efficiency, stability, qualitative efficiency

CHAPTER III

WHY REGULATE THE PAY-TV INDUSTRY?

The purpose of this chapter is to examine the economics of the various arguments as to why or why not the pay-TV industry should be regulated in light of the analysis for justification of government regulation presented in the previous chapter. The first question which should be asked in developing a regulatory policy is: Should the industry be regulated in the first place? The justifications for regulation, if any, will then determine the form which regulation should take.

Economic Arguments for Regulation

Natural Monopoly Argument

The problem of decreasing costs is applicable in some degree to both components of the pay-TV industry--the network entity and the exhibitors (those responsible for delivering pay-TV program packages).

As far as delivery is concerned, pay television is very likely to be offered through existing cable facilities.[1] The cable television industry itself is very capital intensive, resulting in relatively significant decreasing costs up to some point. However, pay-TV will be delivered over <u>existing</u> cable facilities, and thus the true economic costs of a pay-TV system will not include the large capital costs of cable equipment. (Although in reality pay-TV is likely to be assigned its share of the joint

costs involved.)

Table 5, provided to the CRTC by the Pay Television Network (PTN), summarizes the exhibiting costs for various sizes of cable systems (these costs allow 5-year amortization and 15% interest charges, and are based on an intial penetration level of 20%).

TABLE 5

TOTAL COSTS / PAY TELEVISION SUBSCRIBER ($)

Size of Cable System	Annualized Capital Costs			Operating Costs	Total Annual Cost	Capital Cost / Total Annual Cost
	Headend VTRs and Scramblers	Security Devices	(Total)			
1,000	31.58	5.70	37.28	18.76	56.04	.665
3,000	18.53	5.70	24.23	18.76	42.99	.564
5,000	11.12	5.70	16.82	18.76	35.58	.473
8,000	18.19	16.50	34.69	23.36	58.05	.598
15,000	9.70	16.50	26.20	23.36	49,56	.529
25,000	5.82	19.50	25.32	26.01	51.33	.493
50,000	2.91	19.50	22.41	26.01	48.42	.463

Source: CRTC COMMENTS: Pay Television (ottawa, 1976), vol. 1, Comment #28: "Structuring the Introduction and Development of Pay Television in Canada" by (PTN) Pay Television Network, p. F-14.

If the increase in the size of cable system in terms of number of subscribers refers solely to an expansion of the area served by cable (ie. the penetration rate remains constant), then the change in costs represents a movement along the cost curve and thus indicates the extent

of increasing or decreasing costs.

If this is the case, it can be seen from this set of data that the lowest per-subscriber costs occur for a cable system which reaches 5,000 outlets. Costs decrease up to this point, but rise as the size of the cable system increases to 8,000 subscribers. The headend (where programs are originated) equipment costs, which are mostly fixed costs in nature, are primarily responsible for the decreasing costs associated with size, but operating costs and security devices provide substantially increasing per unit costs.[2]

If the increased size of cable system in the data provided by PTN refers to an increase in the penetration rate only (ie. the area served remains constant), then the change in cost may represent shifts in the cost curve and is not reflective of economies of scale in the true static sense. (However, the increase in operating cost with size suggests that these data represent the first case — an expansion of the area served by cable. Installation and servicing costs per subscriber are likely to rise as the size of area increases. If the second case were true — increased penetration rates with the area served remaining constant — it would be

---

[2]According to PTN, the larger the operator, in terms of subscribers, the more stringent are the security requirements (see CRTC, COMMENTS: Pay Television, Comment #28 by PTN Pay Television Network, p. F-10). Thus, larger operators would be impelled to use more sophisticated and expensive security devices. A number of different security systems have been developed, ranging from the relatively inexpensive (low initial capital cost) positive trap device that removes a jamming signal accompanying the pay-TV signal to the relatively expensive but more effective descrambler. However, this source of significantly increasing costs is difficult to explain — there appears to be no reason why larger operators would require better security, and thus the validity of this data is questionable.