

**PRIMARY SCHOOL CHILDREN'S INFERENTIAL PROBLEM
SOLVING IN A COMPUTER GAME CONTEXT**

Seonju Ko

**Child Development and Learning
Institute of Education
University of London**

A thesis submitted for the degree of Doctor of Philosophy (PhD)
1999

ABSTRACT

Computer games are common activities in the nineties and have become a new cultural influence in children's lives. Games labeled 'educational software' are said to be beneficial to the development of children's thinking and learning because they provide opportunity to practise problem-solving skills. However, there is little evidence about what really happens in this respect when children play an educational computer game. Prior to this study, there have been no adequate means for assessing reasoning and problem-solving skills in the context of computer games. The study aims to develop ways to measure and analyze gains in children's cognitive skills acquired through computer game activities. To develop a method of assessing children's reasoning, the game chosen was an inferential problem-solving game called 'Find the Flamingo', one of the 'Safari Search' series (O'Brien, 1985). Different versions of the 'Find the Flamingo' game - computer, board and card games - were given with *if-then* sentences as rules of the game. 282 primary school children took part in this research. Four studies were carried out. **Study 1** compared the effects of specific media on children's performance in the game. No difference was found between the use of computers and traditional game tools such as a board or playing cards. **Study 2** explored developmental trends and individual differences in problem solving with the game. Differences in the curves of performance groups were shown to be stable across games. The production and use of inferences in the process of playing the game were also examined. Children used the inferences with different levels of accuracy according to the conceptual difficulties in the information. **Study 3** explored the impact of guided-planning and timed pausing for reflection on inferential problem solving with a simplified version of the computer game. Children benefited from guided-planning in the training period. **Study 4** examined the development of operative logic of inclusions and exclusions across three inferential tasks and the Flamingo game. The 6-year-old children understood the inclusion rule of multiple possibilities, but they were not able to coordinate the knowledge of inclusion and exclusion to represent more complicated structures. The significant association between the performances of the tasks and of the game even after the control for age allowed the prediction of the Flamingo game performance. Applications of the findings could lead to the design of computer programs that concentrate on specific aspects of problem-solving skills such as planning, and the development of problem-related concepts and operations.

ACKNOWLEDGEMENTS

I love to remember the journey of PhD with these people and to say “Thank you”:

Jaehak and Keunyoung, my husband and my daughter.

Professor Terezinha Nunes and Professor Kathy Sylva who supervised the thesis and led me a deeper understanding of the issues involved and provided critical comments.

Anna Brett has who watched over me from the very beginning of the diploma course to the completion of PhD here in Child Development and Learning group.

Dr. John Versey and Dr. Richard Cowan who taught me psychology.

Eeul Kim who helped with data analysis.

Dear friends, Denise, Ursula, Fred, Kyunghee, Sohyun, Maria, Lily, Angel and Stella.

TABLE OF CONTENTS

LIST OF FIGURES	8
LIST OF TABLES	10
LIST OF APPENDICES	12
Chapter 1. Introduction	14
1. 1. Games.....	15
1.1.1. What is the role of decision-making in games?.....	16
1.1.2. How have outcomes of games been studied?.....	17
1.1.3. What are the special features of computer games?.....	17
1. 2. Problem solving	19
1. 3. The computer game chosen for the research.....	21
1. 4. Structure of the thesis	23
Chapter 2. Literature review.....	25
2.1. Social context: games vs. tasks.....	27
2.2. Development of the concept of a rule in games.....	30
2.3. Contents of rules: affirmative vs. negative	33
2.4. Choice as inference.....	37
2.5. Inferential goal searching.....	45
2.6. Possibility, impossibility & certainty	51
2.7. Gender issues in computer games.....	61
2.8. Methods for investigating children's game activity.....	65
2.9. Drawing the threads together: the study of children's thinking in the context of computer games.....	69

2.9.1. The computer game as a window into children's thinking.....	69
2.9.2. Individual differences and learning	70
2.9.3. Training in planning	71
2.9.4. Prediction for inferential game play.....	71
Chapter 3. Study 1	74
3.1. Aims	74
3.2. Introduction	74
3.3. Method.....	77
3.3.1. Design.....	77
3.3.2. Participants	77
3.3.3. Apparatus	79
3.3.4. Procedure	84
3.3.5. Measures.....	88
3.4. Results.....	89
3.4.1. The game performances.....	89
3.4.2. Performances on the word clue tests	93
3.4.3. Reliability and validity of measurements.....	97
3.4.4. Summary of results.....	98
3.5. Discussion and conclusions.....	99
Chapter 4. Study 2	102
4.1. Aims	102
4.2. Introduction	102
4.3. Method.....	107
4.3.1. Design.....	107
4.3.2. Participants	108

4.3.3. Apparatus	109
4.3.4. Procedure	109
4.3.5. Outcome measures.....	110
4.4. Results.....	114
4.4.1. Analysis of number of moves	114
4.4.2. Analysis of word clue test scores	117
4.4.3. Reliability of measurements	124
4.4.4. Individual differences	125
4.4.5. Learning measured by the completion time and number of moves.....	127
4.4.6. Analyses of choice pattern	129
4.5. Summary of results	134
4.6. Conclusions	136
Chapter 5. Study 3	140
5.1. Aims	140
5.2. Introduction	140
5.3. Method.....	144
5.3.1. Design.....	144
5.3.2. Participants	145
5.3.3. Procedure	146
5.3.4. Outcome measures.....	149
5.4. Results.....	151
5.4.1. Analyses of numbers of moves	151
5.4.2. Analyses of inclusion inferences.....	159
5.4.3. Analyses of inferences from exclusion information	165
5.5. Summary of the results	170
5.6. Discussion and conclusions.....	172

Chapter 6. Study 4	175
6.1. Aims	175
6.2. Introduction	175
6.3. Method.....	180
6.3.1. Design.....	180
6.3.2. Participants	180
6.3.3. Procedure	180
6.3.4. Measures.....	185
6.4. Results.....	188
6.4.1. The Flamingo game.....	188
6.4.2. The Flamingo information card arrangement	188
6.4.3. The Number task.....	189
6.4.4. The Tunnel task	191
6.4.5. Relations between the game and the tasks	193
6.5. Summary of the results	197
6.6. Discussions and conclusions.....	199
Chapter 7. Conclusions and general discussion.....	202
7.1. The findings in relation to the research questions	202
7.2. Limitations and suggestions for further studies.....	212
7.3. Educational implications.....	213
References	217
Appendices	233

LIST OF FIGURES

1.1. A picture of the ‘Find the Flamingo game’ with the rules	22
3.1. Pictures of the computer game	81
3.2. Pictures of the board game	82
3.3. Pictures of the card game	83
3.4. Correct inferences from single clue HOT	86
3.5. Correct inferences form single clue WARM	86
3.6. Correct inferences from single clue COLD	86
3.7. Correct inferences from double clues HOT-WARM	87
3.8. Correct inferences from double clues WARM-COLD	87
3.9. Mean scores across word clue tests	95
4.1. Correct inferences from HOT	111
4.2. Correct inferences from WARM	111
4.3. Correct inferences from COLD	111
4.4. Correct inferences from two HOTs	112
4.5. Correct inferences from a HOT and a WARM	112
4.6. Correct inferences from a HOT and a COLD	112
4.7. Correct inferences from two WARMs	113
4.8. Correct inferences from a WARM and a COLD	113
4.9. Correct inferences from two COLDs	113
4.10. Scores for each of the word clue tests	120

4.11. Scores for word clue tests across age groups	123
4.12. Number of moves for the performance groups across games	126
4.13. Mean number across games for 10-year-old participants	128
4.14. Percentages of single information inference across two age groups	132
4.15. Percentages of double information inference across two age groups	133
5.1. An example of the Boat game	147
5.2. Zone of correct inference from X information	150
5.3. Zone of correct inference from O information	150
5.4. Interactions between age groups and blocks of games	155
6.1. Features of the network of tunnels	177
6.2. Pictures of the network of tunnels	184

LIST OF TABLES

2.1. An overview of studies	73
3.1. Number of participants across the three media and gender	78
3.2. Mean number of moves across the three media and for each gender	92
3.3. Two-way ANOVA (medium x gender) for mean numbers of moves	92
3.4. Mean total scores for the word clue tests across the three media and gender groups	94
4.1. Number of participants across age, gender and medium	108
4.2. Mean number of moves (SD) across age, media and gender groups	116
4.3. Mean scores (SD) across age, media and gender groups	118
4.4. Number of participants, mean number of moves and mean test scores for each performance group	126
4.5. Percentages of moves across three areas from single information	130
5.1. Number of participants across age and treatment groups	145
5.2. Mean number of moves across treatment and age groups	153
5.3. Three-way Analysis of Variance on performance over the practice of games	154
5.4. One-way Analysis of Covariance of training performance	157
5.5. One-way Analysis of Covariance of post-training performance	158
5.6. Mean percentage of correct inclusion inferences across treatment and age groups	160
5.7. Three-way Analysis of Variance on inclusion inference	161
5.8. One-way Analysis of Variance of training performance	163

5.9. One-way Analysis of Covariance of post-training performance of inclusion inferences	164
5.10. Mean percentage of correct exclusion inferences across treatment and age groups	167
5.11. Three-way Analysis of Variance on exclusion inferences	168
5.12. One-way Analysis of Covariance of training performance of exclusion Inferences	169
6.1. Numbers of participants across choice categories	190
6.2. Correlation coefficients between the Flamingo game and the tasks	194
6.3. Results for three fixed order regressions with mean number of moves of the last 4 games as the outcome variable	195

LIST OF APPENDICES

1. An example of the word clue test sheets	233
2. Distribution of the mean numbers of moves in Study 1	234
3. Distribution of the word clue test total scores in Study 1	234
4. Distribution of the mean numbers of moves in Study 2	235
5. Distribution of the word clue test total scores in Study 2	235
6. Distribution of the mean time taken in the first 4 Flamingo games in Study 2	236
7. Distribution of the mean time taken in the last 4 Flamingo games in Study 2	236
8. Distribution of the mean numbers of moves of the first 4 Flamingo games in Study 2	237
9. Distribution of the mean numbers of moves of the last 4 Flamingo games in Study 2	237
10. Distribution of the percentages of correct inferences from HOT in Study 2	238
11. Distribution of the percentages of correct inferences from WARM in Study 2	238
12. Distribution of the percentages of correct inferences from COLD in Study 2	239
13. Distribution of the mean numbers of moves of the last 4 Flamingo games in Study 4	239
14. Distribution of the error scores of the game card arrangement task in Study 4	240

Dedicated to
Keunyoung

Chapter 1. Introduction

Computer games are common activities in the nineties and have become a new culture in children's lives. Children play computer games in and out of the school curriculum. It is a strong tendency in the education policies in many countries to encourage computer use with every age of children (Janssen Reiner & Plomp, 1997; Pelgrum & Plomp, 1991) and home computers are used mainly for game playing (Giacquinta, Bauer & Levin, 1993). New games are often labeled 'educational software'. Although the games are said to be beneficial to the development of children's thinking and learning, there is little evidence about what really happens in this respect when children play an educational computer game. Thus, it is important to develop ways to measure and analyze gains in children's cognitive skills through the computer game activities of the present day.

The aim of the thesis is to develop a framework for analyzing children's thinking in computer games. In particular, the focus is on the development of children's ability to make and use inferences in the context of computer games and how this development differs between individuals. Despite the recent upsurge of interest in computer games, relatively little is known about educational games, and even less about how to examine the development of thinking in an educational game context. To build a framework to investigate computer game activity, two points can be made from research on games in general and on problem solving. The first is that decisions have to be made in games (Colman, 1982; Davis, 1970); the second is that processes of decision making and problem solving often involve inference (Piaget & Garcia, 1991; Manktelow & Over, 1990; Thornton, 1995). Ways of analyzing children's inferences in the context of computer games will be explored in this thesis.

The introduction consists of four sections. In the first section, the role of decision making in games, and the measurements and special features of computer games are defined and discussed. The second section describes how game playing develops into problem solving. The role of concepts in problem solving contexts is discussed. And there will be an introduction to the game chosen for this thesis: its criteria and purposes. The organization of the thesis will be described at the end.

1. 1. Games

Voluntary control has been emphasized in definitions of games (e.g., Demsey, Lucassen, Gilley & Rasmussen, 1993-4; Herron & Sutton-Smith, 1971; Pellegrini, 1995). Players decide what their goals are, plan their moves and attempt to anticipate the other participants' intentions. Since the outcomes may depend upon the choices of the decision-makers involved in games for two or more persons, the players may try to assign probabilities to one another's moves. The player's moves may be aimed directly toward the goal or to obtain more information. In this thesis, games requiring the search for a goal according to certain rules will be studied.

Scholars have classified games according to certain family resemblances. The most basic criterion of classification seems to concern the factors that affect the outcomes of the game: games of skill, games of chance, games of strategy and games of second-guessing. Games of skill, strategy and chance may be classed as games for one person or more, while games of second-guessing must involve two or more decision-makers. The term 'second guessing' is defined as anticipating the moves of an opposing player. This classification is useful because the method of analyzing a specific game can be applied without modification of the fundamental ideas to other games which belong to

the same general class; only relatively few games need, therefore, to be studied in order to understand a much larger number.

1.1.1. What is the role of decision-making in games?

Performances in games are rich sources for the analysis of attributes of decision making. Game players continuously ask themselves, "Which strategy/move should I use next?" Drawing on cognitive and information processing theories, scholars have asserted that

"Human rational behaviour is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor." (Simon, 1990).

Consequently, the decision-making of a game player is bound to her or his understanding of the task, in other words, the rules of the game and the types of information encoded in the course of the game. From the perspective of mathematical game theory (see McKinsey, 1952), randomness and hidden information make the outcome of a game uncertain and this uncertainty is a major characteristic which makes the game interesting and attractive (Malone, 1980). Can young children understand the rules of games and pay attention to information given in the course of a game? What kind of logical inferences do children make? How do individual differences affect game playing? Strategies or goal-directed search patterns are shown in games. A plan, a consciously adopted strategy, means using information to determine further moves towards the goal and to select appropriate moves to bring this about. Do children plan

moves in the game? Winning, losing, or reaching the goal depends on players' strategies and planning as well as on other factors, such as luck.

1.1.2. How have outcomes of games been studied?

Outcome measurements in game activity can be quantitative or qualitative. One approach to measuring gains in skills from playing games is to look at the quantitative changes in performance with practice (Mazur & Hastie, 1978; Walker, 1996). The reduction in solution times with practice has been emphasized, by following individual learning over a large number of trials as the player becomes increasingly competent at executing a skill. As qualitative measurements, types of strategies and categories of judgements are often used. However, there must be caution about the validity of measuring types of strategies, since more- and less-advanced strategies may coexist in an individual and it is not the case that children always apply their most advanced strategy in a task (Siegler, 1989; Sternberg & Davidson, 1995).

1.1.3. What are the special features of computer games?

Firstly, repetitiveness is a distinguishing characteristic of computer game activities. Many children are highly motivated to play against the computer over and over again. Repetition gives players the opportunity to develop cognitive skills related to the game. In learning-from-doing, the players acquire more generalized strategies (Anderson, 1982; Bruner, 1996; Greenfield, 1984). Next, fast feedback and the visual and audio

aspects which the traditional pencil and paper tasks can never offer are motivational factors of this new medium (Carter, 1984; Siegel & Misselt, 1984). In the tradition of Marshall McLuhan's remark that the medium is the message, Kozma (1991) argues that, in a good instructional design, media and method are narrowly integrated and consequently the learner constructs knowledge in interaction with medium and method. Thirdly, proponents of computer games point to a form of preparation for, or initiation into, the more cognitively demanding world of computer technology (Hawkrige, 1990); however, this is not the main concern of this study. Finally, due to the technical advance in computers, it is possible to have automatic recording which tracks each move players have made in computer games.

1. 2. Problem solving

Educational games embody problem solving. Deciding in which boxes to put three crosses to make a straight line in the game of noughts and crosses, which letter will complete a word in “Hangman”, or how to find the way to the exit in a Maze all demand problem solving. Solving a problem is a search that occurs when the means to an end do not occur simultaneously with the establishment of the goal (James, 1890). The journey of problem solving, in almost any instance, has different steps: recognizing that there is a problem and identifying a new goal; planning a strategy to solve the problem; noticing whether or not this strategy works and planning another one if it does not (see Marshall, 1995; Newell & Simon, 1972). Do children recognize problems in a game? Do they plan the use of information before starting a game or in the process of the game? Inferences are a key element in problem solving in games. At each step in problem solving, the child must make sense of the information available and use that information to generate a new understanding of the problem or a new strategy. This involves making inferences or deductions that take the child from what she or he originally knew to some new piece of information or new idea (Thornton, 1995).

Some insight into the question of how children relate a new piece of information to what they already knew has come from neo-behaviourists. In concept formation tasks, a participant must in a first task learn to select one of two (or more) stimuli. She or he is then required in a second task to select one of the previously incorrect choices; i.e. she or he must reverse her or his response to a particular situation. If she or he is trying to infer what makes certain choices positive and certain ones negative, her or his task is to discover which of the discriminable attributes or which combination of discriminable attributes are present in the positive instances and absent in the negative ones. Discovering is, then, a matter of developing intentional hypotheses or descriptions of

concepts and classes identified in clustering, so that instances not previously encountered can be recognized and utilized (Langley, 1987; Mascolo & Fischer, 1999).

Individual differences in performances of concept formation tasks have been reported (Kendler, 1995; Kendler & Kendler, 1959; Kendler, Kendler & Carrick, 1966; Kendler, Kendler & Wells, 1960). In Kendler & Kendler (1959), kindergarten children (5-6 years) were trained on a discrimination problem with two pairs of stimuli which differed both in size (e.g. large *v.* small) and brightness (e.g. black *v.* white). Only one stimulus within a specific dimension was correct (e.g. the large stimulus in the size dimension); the stimuli in the other dimension (e.g. both brightness stimuli) were irrelevant to the solution of the problem. After solving this problem, the children were required to solve a second discrimination problem in which they either had to reverse their previous choice (e.g. select small) or shift to the previously irrelevant dimension (e.g. select black rather than either large or small). The former is termed a reversal shift; the latter a nonreversal shift. With the kindergarten children neither shift was favoured. However, when the children were divided into fast and slow learners, based on their first task performance, it was found that the fast learners mastered the reversal shift better than the nonreversal shift while the opposite was true for the slow learners. How can these individual differences be explained? More direct observation of the learning process itself, rather than its results, has been suggested as a more productive way to analyze individual differences in ability to learn (Resnick & Neches, 1984). The question of how children learn search strategies will be discussed regarding choice as inference in Chapter 2.

1. 3. The computer game chosen for the research

Two criteria for the choice of a computer game for the research were, first, that it can be played by children within a broad age range and, second, the problem solving feature of the educational software. The former criterion is in order to explore what it is that makes younger children able to play with computer games successfully while older children still find them challenging; a question that instantly arises is whether there are developmental trends in understanding the game structure and inferring the target from information. The latter criterion is in order to examine whether children use logic in playing computer games and what they take into account when making decisions about the next moves.

For the purpose of this research, the chosen game was an inferential problem-solving game called "Find the Flamingo", one of the 'Safari Search' series (O'Brien, 1985). It is advertised as educational software and one of the goals of this research is to explore the relation of such software to children's thinking. The goal of the game is to find the location of a Flamingo hidden among 25 cards. The instructions appear on the screen as follows:

Turn over a card to find the Flamingo.

If it's there, you win.

If the Flamingo touches your box sidewise, you are **HOT**

If it touches your box cornerwise, you are **WARM**

If they don't touch at all, you are **COLD**.



Figure 1.1. A picture of the 'Find the Flamingo game' with the rules

The game looks easy and friendly, since the Flamingo must be underneath one of 25 cards, which make a scene on a safari. The Flamingo can be searched for using merely the strategy of random trial and error. However, if the players try to make an effective search, it is cognitively challenging in that information has to be selected in the course of the game, and spatial directions to the Flamingo should be inferred from the information. Interesting research questions raised in this game will be explored in the thesis. Did children understand the logical structure of the *if-then* sentences given as rules of the game and use their knowledge in the course of playing the game? How would children construct and use the new meanings of the words - Hot, Warm and Cold - in the Flamingo game? Hot, Warm and Cold in the game are not degrees of temperature or feelings as used in everyday life, but relations to the location of the goal. Faced with the spatial information of Hot, Warm, and Cold in the game, would the children construe the words as signals of exclusion of impossible locations of the goal or inclusion of possible locations?

1. 4. Structure of the thesis

The thesis is divided into 7 chapters. Chapter 2 provides an account and evaluation of major approaches to explaining children's inference making. Two approaches will be discussed: neo-behavioural and cognitive constructivists' views. Issues concerning the process and components of strategic thinking are considered in more detail. Children's assumptions toward a game, the development of the concept of rule, and understanding inferential goal search will be examined.

Chapters 3 – 6 contain accounts of four experiments, each including the specifics of design and analysis.

Chapter 7 presents a summary of the main findings in relation to the research questions, limitations of this project and suggestions for further research, and the educational implications.

In brief, this thesis explores various aspects of children's making and use of inferences in the context of computer games, and examines ways to analyze their performance. It aims to use the computer game as a window into the everyday thinking of children. It is hoped that findings from studies such as this will be useful in designing educational programs using computer games for children.

Chapter 2. Literature review

This literature review aims to develop a technique for analyzing children's thinking in a computer game context. I will explore various aspects of thinking in game playing: children's representation of the game, functions of their choice behaviour, the process by which judgements are made, individual differences and development, and the need for adult intervention. These six aspects have been also acknowledged as fundamental elements of a theory of thinking (Johnson-Laird, 1983; White, 1993).

Regarding the above elements of thinking, a hypothetical model of decision-making processes is proposed in this study. The model of decision making involves the following process. First, a child is faced with the situation. The player perceives the situation. The product of this perception is an internal representation. In the Flamingo game, the player is presented with many squares making a beautiful safari scene on the first screen and the rules appear on the next. The rules of the Flamingo game say:

Turn over a card to find the Flamingo.

If It's there, you win.

*If the Flamingo touches your box sidewise, you are **HOT**.*

*If it touches your box cornerwise, you are **WARM**.*

*If they don't touch at all, you are **COLD**.*

Surely the player asks herself or himself what is this situation: is winning a chance event or a target which can be achieved logically? It is assumed that the representation of the situation would affect choice making, which is the next process in decision making.

Second, choices are made in a game situation. This situation is conceived as complex, containing a set of choices (at least two and possibly more), each of which involves a set of outcomes (gains or losses of various kinds and degrees). The choices, in turn, are embedded in a context or frame. Each square contains HOT, WARM, COLD or the Flamingo. Each word conveys a spatial relation to the goal, if the child makes a connection between the words and the game rules. Does the player infer from the rules when they make choices? Will all the rules be used for inference? If she or he is not good at using the rules, will use improve as the game goes on?

The third part of the decision process is an attempt to solve the problem. The goal is hidden among 25 squares in the game. The task of finding the Flamingo becomes a problem to solve. How does the player represent the structure of the problem? Will she or he plan moves ahead? The decision is made internally, but the choice leads to actual behaviour that can be directly observed. Since the player's mind cannot be seen directly, her or his actual choice of behaviour in a given situation constitutes a clue, and sometimes the only clue, to what the person's decision processes might be. These three parts of the decision making process are not mutually exclusive. In particular, choices cannot be separated from the problem solving process if the player draws logical inferences from information given in the form of alternative choices.

In my examination of the decision-making processes, I will review and discuss studies of social context, the development of the concept of a rule, contents of rules, choice as inference, inferential goal search, logical certainty and a specific additional issue: girls and computer games. The research questions for each of the experiments of the thesis will be introduced at the end of the chapter.

2.1. Social context: Games vs. Tasks

The first part of the decision process concerns the player's perception of the situation. How does the child play a game if she or he defines it as a chance event? What if the player defines it as problem solving? There is a golden old research study on this issue. Goodnow (1955) carried out an experiment based on a two-choice task. The participant was to predict on each trial which of the two possible events occur when the series of trials has been randomized and so arranged that the probability of event A occurring is 0.7, and of event B, 0.3. Two other cases used an equal probability, 0.5 of A and B event occurring and a probability A of 0.9 with B of 0.1. One group of subjects was instructed so as to induce a 'gambling' set, while another group was instructed so as to induce a 'problem-solving' set. The probability of winning was programmed as the same in the two sets. There was no difference in children's choices between gambling and problem-solving sets for cases 50:50 and 90:10. However, there were significant differences for case 70:30. The players came to predict A (the more frequent event) on all trials in the 'gambling' set, but not in the 'problem' set. In the 'problem' set, the players behaved as if there was some pattern to be found in the game and that regularities were applicable. The 'gambling' set was more conducive to the maximizing tendency (the predictions approximating 100%) than the problem set. Siegel and Golstein (1959) explained the results of Goodnow's study in terms of strategy while confirming the results of her study. They argued that circumstances in which risk or uncertainty occurs, like in the gambling set, were most likely to induce a "maximum gain" strategy: the maximization of the expected frequency of correct predictions. The player came to predict the more frequent event on all trials in the 'gambling' set. These studies showed how the perception of the event altered the behaviour.

Ceci (1990) compared a game context and a problem-solving context in computer activities. Children were instructed to predict by placing a cross on the screen at that location (by moving a joystick), where, on the computer screen, an object would terminate. The object was one of three shapes (square, circle, and triangle), two colours, and two sizes, yielding 12 combinations of features. Squares would go up, circles would go down, and triangles would stay horizontal. Similarly, dark-coloured objects would move to the right, and light-coloured objects would move to the left. Large objects would move on a lower-left to upper-right diagonal, while small objects would move along the opposite diagonal. There were no interactions in the algorithm. Children were given 15 sessions of 50 random trial each; however, their prediction accuracy was only 22%. Next, the same algorithm was used to drive a video game. The three geometric shapes were converted to a butterfly, a bumble bee, and a bird. The same colours and size were used. Instead of placing a cross on the area of the computer screen where they predicted that the object would terminate, children were told to place a "butterfly net" to capture the prey by moving a joystick. Children were awarded points for each correct capture, and sound effects were added to complete the video game context. The change of context resulted in drastically improved performance. After 750 feedback trials altogether, the children's accuracy was near ceiling.

In an extension of this study, a more complex curvilinear algorithm was used to drive the objects. Again, there was a substantial enhancement of performance when the task was embedded in the presumably more motivating video game context, although the overall levels of performance were not as high as those found with the simple additive algorithm.

Relatively few research studies have been conducted on the educational effectiveness of computer games and much of the work is anecdotal, descriptive, or judgmental, rather than investigative (Randel, Morris, Wetzel, & Whitehil, 1992; Walford, 1995).

Nevertheless, some issues have been raised in investigating the effect of educational software on children's learning.

Firstly, research has shown that children get better in using various cognitive skills while they are engaged in the practice of games (Blaye, Light, Joiner, & Sheldon, 1991; Littleton & Light, 1999; Sedighian & Klawe, 1996; Sedighian & Sedighian, 1997; Subrahmanyam & Greenfield, 1994). Children in a game context learn skills that they cannot learn if the activity is defined as a task, for motivation is created by the situation. They behave differently, depending on their assumptions toward a game. If they represent the Flamingo game as a chance event, they make random choices and their behaviour will reflect reinforcement. If they represent it as a problem, their game behaviour becomes problem solving and making inferences from information.

Secondly, observation has been widely made of marked gender differences in the ways in which boys and girls tend to approach computer-based games (Brosnan, 1998; Hughes, 1990). Section 2.7 will review this gender issue in more detail.

2.2. Development of the concept of a rule in games

What if a child starts a game without relevant concepts about rules? Can we predict whether she or he will be a good player? The literature on children's concept of a rule and its relation to game play will be reviewed in this section.

Piaget investigated the question of how children come to understand the rules of games (Piaget, 1932). In accordance with his "stage" theory of the development of children's intelligence, he explains a child's feelings and thoughts about rules in terms of assimilation and accommodation. In the first two stages (up to 6 years), the child assimilates rules unconsciously along with the commands to which he is subjected, taken as a whole. According to Piaget, the simple routine regularities of experience that precede the rules of games imposed by a group of players give rise to the consciousness of rules. In the case of babies, external regularity produces an awareness of "law", or at any rate favours the appearance of motor schemas of prevision. Sources of awareness of the regularity of the environment can be either certain physical events (alternation of day and night, sameness of scenery during walks, etc.) or parental discipline (meals, bed-time, cleanliness, etc.). For older children, however, certain forms of behaviour are ritualized by the child herself or himself (e.g. not to walk on the lines that separate the paving stones from each other on the curb of the pavement). While these motor rules never give rise to the feeling of obligation, certain rules - it does not matter whether they were previously invented by the child, imitated, or received from outside - are at a given moment sanctioned by the environment, i.e. approved of or enjoined.

Piaget (1932) distinguishes three kinds of behaviours in play: motor behaviour, egocentric behaviour (with external constraint) and co-operation. Corresponding to

these three types of social behaviour there are three types of rules in playing games: the motor rules due to pre-verbal motor intelligence, coercive rules due to unilateral respect, and rational rules due to mutual respect.

The child aged between 3 and 7 knows that there are rules, the "real rules", and that they must be obeyed because they are obligatory. But on the other hand, although the child vaguely takes note of the general scheme of these rules (making a square, aiming at the square, etc.), she or he is still hostile to any innovation of rules at this stage.

On the average after the age of 10, i.e. from the second half of the co-operative stage and during the whole of the stage when the rules are codified, autonomy follows upon heteronomy. The rule of a game appears to the child no longer as an external law, sacred in so far as it has been laid down by adults, but as the outcome of a free decision and worthy of respect in the measure that it has enlisted mutual consent.

Corresponding to the development of behaviour, Piaget (1951) classifies play by the structure that characterizes it. He makes three main categories: sensori-motor play (sometimes translated as "mastery play" or "motor practice games"), symbolic play, and games with rules. Piaget suggests two ways to study children's games which have rules: first, to observe the practice of rules, i.e. the way in which children of different ages effectively apply rules: second, to study the consciousness of rules, i.e. the idea which children of different ages form of the character of these game rules, whether it is something obligatory or something subject to their own choice, whether they experience heteronomy or autonomy.

The importance of Piaget's theory of game play (1932, 1951) lies in the fact that he relates children's game play to the development of logic and conceptual thinking. Drawing on empirical evidence, he describes the children's development of the construction of rules. For the sake of the argument, Piaget does not mention the fact

that rules in many strategic games are not subject to the player's choices. The rules are already set up. Thus, the acceptance of the rules is obligatory. For example, in chess, players are not concerned about changing rules but strategic use of the rules.

Of much interest in observing children's game behaviour is children's internalization - understanding - of the game rules and use of it. It is important to distinguish between pure regulation of game behaviour without awareness of following any rule and proper rule-guided behaviour. Not only clinical discourse but also quantitative methods will be applied to analyze children's thinking in this matter.

2.3. Contents of rules: affirmative vs. negative

This section of the literature review looks through the previous research on children's understanding of the logical structure of the *if - then* sentences which are given as the game rules, and the use of their knowledge in the course of playing the game. These rules are stated as affirmative and as negative in the Flamingo game.

*If the Flamingo touches your box sidewise, you are **HOT**.*

*If it touches your box cornerwise, you are **WARM**.*

*If they don't touch at all, you are **COLD**.*

It could be that when she or he makes choices, a player use syllogisms* with these rules. An example would be the player with **HOT**. The reasoning is as follows: "I am **HOT**. The rule says *If the Flamingo touches your box sidewise, you are **HOT***. Therefore, the Flamingo touches my box sidewise." However, it is very unlikely for the player to be able to recall the exact sentences of the rules at each move (Bartlett, 1932; Johnson-Laird, 1983; Markovits, Fleury, Quinne & Venet, 1998).

The rules in use are more like:

*When the clue is **HOT**, the goal is either directly above or below, or right or left from it.*

*When the clue is **WARM**, the goal is one of diagonals from it.*

*When the clue is **COLD**, the goal is not a square touching it.*

* An *if p then q* statement serves as a major premise and the negated or affirmed version of the antecedent and consequent serves as a minor premise.

Research on reasoning has shown that there are differences in information-processing between positive and negative statements. Wason (1959) investigated the mechanism behind the information processing of affirmative and negative propositional sentences with college students. Participants were given four kinds of statements, generated from two kinds of information - positive or negative - with two values: true or false. The four kinds of statements can indicate that something either *is* the case or *is not*. An affirmative statement that is known to be true and the complementary negative statement which is known to be false both imply that something *is* the case. An affirmative statement that is known to be false and the complementary negative statement which is known to be true provide the same information, i.e., that something *is not* the case. The time taken to judge each statement was measured and compared between each kind of information and between values. He measured accuracy of judgements and time taken in completing the task of processing the information. It was found that dealing with negative statements led to more mistakes and took more time than dealing with affirmative statements. Wason supposed that a "pre-existing" set for positive information explained why the false affirmative condition takes longer to process than the affirmative one. The role of a negated constituent, that is a denial of a presupposition, would require more cognitive processes than an affirmation.

Many other explanations have been proposed for the phenomenon (see Evans, 1989). A negated constituent does not always deny a presupposition, but it often identifies a contrast class (Hampton, 1989). Toppino's experiment (1980) is one of the good examples showing that a negated constituent identifies a contrast class. Toppino investigated children's responses to positive and negative queries when two comparing values are both present or one is absent. Kindergarten children were shown two cards - a hypothesis card and a stimulus card - simultaneously. A hypothesis card contained values from two attributes (e.g., blue and circle) and a stimulus card contained only one of these values (e.g., blue). Half of the stimulus cards (proper-subset stimuli) contained only the single value (e.g., blue), which constitutes a proper subset of the

values on the hypothesis card. The other half of the stimulus cards (non proper-subset stimuli) contained two values, one of which *was*, and one of which *was not*, on the hypothesis card (e.g., blue and cross). Half of the children were asked to indicate which value on the hypothesis card was also on the stimulus card (positive queries) and the other half were asked to indicate which value on the hypothesis card was not on the stimulus card (negative queries). The kindergarten children did reasonably well on problems involving the proper-subset stimuli for both positive and negative queries, in that 80 to 90% of the children solved all such problems. However, the results were quite different for those problems involving the non proper-subset stimuli, which included irrelevant features on the stimulus card. Here only 50% of the children solved all problems with positive queries and only 25% solved all problems with negative queries. Corroborating results were obtained in two additional experiments.

However, even though negative queries did pose difficulties for the children, research has shown that pre-training and feedback enhance younger children's handling of them (O'Brien & Overton, 1982; Spiker, Cantor & Klouda, 1985). Spiker, et al, in one of their experiments (Experiment 3), explored factors in producing superior performance: feedback to the children on the correctness of their responses to queries, pre-training with its attendant feedback, or explicit use of labels. They presented 4- and 6-year-old children with sets of cards consisting of relevant and irrelevant information. The unitary stimuli consisted of colour-form compounds made from combinations of one of the colours, red, blue, or green, with one of the forms, square, circle, or triangle. The partitioned stimuli were produced by spatially separating the two components of each unitary compound. A positive query consisted of asking the child to indicate the value on the hypothesis card that *was* also on the stimulus card. A negative query asked the child to name the value on the hypothesis card that *was not* on the stimulus card. Each pair of hypothesis and stimulus cards together with either a positive or negative query constituted a single reasoning problem. Superior responses to the positive query than to the negative were shown. The level of correct performance of the first graders was

as high as above 90% for positive queries. This was much higher than for negative queries. In addition, this difference was found only for the three most difficult tasks. A combination of feedback information and preliminary experience with simple forms of the tasks produced the high performance levels, while the verbal labelling of stimulus components had no effect.

The usage and the meaning of "*If*" change with context (e.g., Wason & Johnson-Laird, 1972; Light, Blaye, Gilly & Girotto, 1989; Light, Girotto & Legrenzi, 1990). "*If-then*" can be used to express the propositional concept of conditional implication and "*If and only if*" can be used to express biconditional equivalence. It was found that understanding the distinction between "*if-then*" and "*if and only if*" improves into adulthood, but tends to be still problematic even for adults (Byrnes & Overton, 1988; see Evans, Newstead & Byrne, 1993; Smith, Langston & Nisbett, 1992 for reviews). In everyday life "*if*" is used for biconditional equivalence as well as conditional implication. Young children employ "*If-then*" statements to express meaningful relations such as temporal order, causality or rule-governed social interactions (French & Nelson, 1985; Scholick & Wing, 1992, 1995).

Studies looking for competence in logical ability have demonstrated young children's sensitivity to semantics. Processing negative queries and statements seems to pose more difficulties than positive. The negated rule "If they don't touch at all, you are COLD" implies that the Flamingo cannot be next to the current location when your choice says COLD. Inference from COLD directs the player to exclude touching squares and to hold ones NOT touching for the goal. The negative constituent in the rule of COLD identifies an opposite class of WARM and HOT that direct the players to infer the inclusion of touching squares. It is predicted that correct inference from COLD will be poor due to the negative constituent in the rule. But feedback may enhance children's performance in inferring from the rule with a negative constituent, as studies have shown.

2.4. Choice as inference

Following the first part of the decision-making process, that is, the representation of the situation, the second part is choice making, which raises a research question of whether choices are dictated by a conscious representation of the problem, or by a likelihood gained through the past experience of reinforcement. Investigations on the issue have been through discrimination (or concept formation) tasks in which two (or more) stimuli are constantly rewarded. For example, choice of a 'large' object is consistently rewarded regardless of its size in a training session. Then the job of the player in a test session is to choose correctly which one the reward will follow. The learning of 'small' is easier than 'white' or 'black' if 'large' has been learned. Why is this the case? It cannot be explained in behaviouristic terms. Whenever 'large' is reinforced, 'small' is not reinforced. A child is less likely to choose 'small' on the next round. Thus, a strict stimulus-response model cannot explain the case.

How can children learn to choose a stimulus which they were not trained to choose? Of great importance in the discrimination tasks is the aspect that participants face successive choice problems and pick up the rules which lie behind the distribution of feedback. Kendler and co-workers (Kendler, Kendler, & Leonard, 1962; Kendler & Kendler, 1970; Kendler, 1995) proposed verbal links between the traditional S-R associations. A link might serve to unite the discrete stimuli "large" and "small" under the concept of 'size is relevant', whereas nonreversal shift or extradimensional shift, that is, choosing white or black in the above example, can be still explained by single unit S-R theory. Kendler and Kendler (1959) investigated the ontogeny of reversal learning. They hypothesized that 5-6 years was the transitional point at which verbal mediation appears; therefore, children who had not yet learned to mediate would be slow learners and would find it difficult to reverse. Conversely children who had

learned to mediate would be fast learners and would find it easy to reverse. Kindergarten children (5-6 years) were trained on a discrimination problem with two pairs of stimuli, which differed both in size (e.g. large v. small), and brightness (e.g. black v. white). Only one stimulus within a specific dimension was correct (e.g. the large stimulus in the size dimension); the stimuli in the other dimension (e.g. both brightness stimuli) were irrelevant to the solution of the problem. After solving this problem, the children were required to solve a second discrimination problem in which they either had to reverse their previous choice (e.g. select small) or shift to the previously irrelevant dimension (e.g. select black rather than either large or small). The former is termed a reversal shift; the latter a nonreversal shift. With the kindergarten children neither shift was favoured. However, when the children were divided into fast and slow learners, based on their first task performance, it was found that the fast learners mastered the reversal shift better than the nonreversal shift, while the opposite was true for the slow learners. However, Kendler & Kendler's assumption that fast learners were mediators while slow learners did not make a verbal link was not proved. Slow learners were not necessarily non-mediators. Furthermore there was no clear demonstration that mediation occurred.

As a method of verifying these hypotheses, Kendler, Kendler and Wells (1960) tested pre-kindergarten age children, since at this age most children should be premeditational and therefore show better nonreversal than reversal shifting. The training task involved only a single dimension (e.g. brightness) with no irrelevant dimension present. Half of the participants were assigned to the treatment group and instructed to tell the experimenter which stimulus was positive and which was negative. They were helped to express the correct word (e.g., black, white, big or little). Each child in the treatment group named both the correct and incorrect stimuli for 10 trials. No significant effect on reversal learning was found attributable to the instructions to verbalize. There was no significant difference in the number of reversal shifts made between the verbal

training group and the control. Overall, nonreversal shift was favoured more than reversal shift.

However, in this study (Kendler, Kendler and Wells, 1960) in contrast to Kendler and Kendler's former study (1959), the training task involved only a single dimension (e.g. brightness) with no irrelevant dimension present. Therefore, the two studies may not be comparable. In the first (Kendler and Kendler, 1959), the children may have learned to avoid the irrelevant cue (e.g. avoid size when brightness was correct) and thus found nonreversal shifting difficult since they then had to select this previously negative dimension. In the second study (Kendler, et. al, 1960), however, nonreversal shifting may have been easy not because the children were younger but rather because they could have no bias against the dimension used in the training task.

To clear up the difference between the two tasks, Kendler and Kendler (1962a) again presented children of 4 and 7 years of age with a pair of stimuli that varied simultaneously in size and brightness. While learning, the participants were required to verbalize aloud the stimuli to which they were responding. One-third of the participants was instructed to say "large" (or "small" as the case may be). The participants were helped to express the correct label by a series of questions. Another third was instructed to say "black" (or "white") in a corresponding way. The remaining third was not required to say anything. After learning the discrimination, all subjects were presented with a reversal shift. The shift was to a stimulus that was small, regardless of the treatment group. Thus, the group that initially described the correct stimulus as "large" had verbalized the relevant dimension. The verbal response of "black" was irrelevant to this reversal shift. It was reported that the younger children profited by making the kind of verbal response appropriate to a reversal shift, while learning inappropriate verbal responses hindered them. With no verbalization the 7-year-old children accomplished a reversal shift much more rapidly than the 4-year-olds. The 7-year-old children did not profit from being trained to make the relevant responses. The

performance of the 7-year-olds in the irrelevant verbalization group was even poorer than that of the 4-year-olds. The experimenters interpreted these results as that the relevant labelling helped those who had not learned the concepts while those who already had ability to form the concepts were disturbed by irrelevant labelling. Children who are not able to represent the situation as, for instance, that big stimuli are reinforced regardless of their colour, can be helped by repeatedly labelling, "Big", "Big", "Big". This labelling may help children to infer that a small stimulus *is* the case when "Big" *is not* the case, under the concept of 'size is relevant in this task'.

Later works by Kendler and co-workers focussed more on representational mediation. Kendler and Ward (1971) presented two sets of conceptually related pictures (e.g., animals versus vehicles) to two groups of kindergarten children. For one group, the reversal experimental procedure was used in that each picture was presented singly, and immediately following the correct choice the picture was removed from view. For the other group, the pictures were presented in a cumulative way. After being sorted correctly, the picture remained in view so that the participant could inspect it while responding to subsequent pictures. The experimenter reported faster reversal shifts in cumulative stimulus presentation than in single presentation. The result was articulated that as the number of previously sorted pictures available for inspection increases, the probability increases that the subject will use a common representation for the instances of each sorting category.

An alternative explanation for performance on discrimination is the attention theory proposed by Trabasso and Bower (1968) and Zeaman and House (1963). They diagnosed young children's difficulty as being unable to ignore irrelevant and redundant information. Their poor performance on discrimination tasks is due to the fact that they are easily distracted by extraneous information.

Can the attention theory account for the effects of overtraining? Zeaman and House (1963) modeled an incremental attention to the relevant dimension and weakening attention to the irrelevant dimension over trials. In contrast, according to Trabasso and Bower's model (1968), the relative strengths of dimensions are set at the outset of a discrimination problem and do not change over trials. It is assumed that, on a trial, the participant only attends to a certain subset of the available dimensions, i.e. the focus of the sample. The stronger a dimension is, relative to the other dimensions, the more likely it is to enter the focus sample and to more of the spaces in the focus sample. When a child makes a correct choice, all the cues that are inconsistent with that choice drop out of the focus sample. On the next trial, there is a certain probability that the space will not be refilled, causing the focus sample to become reduced in size. Thus, extended overtraining should increase the probability that the focus sample will have closed down, so that only the relevant dimension remains. Therefore, according to this theory, as well as Zeaman and House's, overtraining should facilitate sustaining attention on the relevant dimension and reduce attention to incidental information.

The effect of overtraining was compared to that of the one-dimension transfer of training in Casey's study (1976). It was assumed that narrowing down attention to the focus sample could be better achieved by training that involved only one dimension (e.g. colour) than by over-training that involved one relevant dimension (e.g. colour) with an irrelevant one (e.g. form). 4 and 7 year-olds were presented with intra-discrimination tasks, in which one dimension had more than two attributes (e.g., red, yellow and green). The number of the participants who made no errors (i.e., two-cue learners) on the test trials was compared to the number who made one error (one-cue learners) in each condition. The pretraining group, in which the participant had in advance a session involving only a single dimension (e.g., colour) with no irrelevant dimension present, had a significantly greater number of two-cue learners than the overtraining group, in which 20 more trials were given in the initial training session. The training with a transfer problem was more effective than extended training for

focusing attention on the relevant dimension. No age group difference was found. The lack of an overtraining effect was difficult to explain in terms of an incremental learning model. Other investigations (Anderson, 1972; Lycsak & Tighe, 1975) also reported no significant effect of overtraining with young children.

Studies with discrimination tasks have shown that even young children act as if the accumulated experience of the previous reinforcements has an effect on solving the current problem. Despite weakness in the design of discrimination tasks (see Blank, 1968), invaluable implications can be drawn from the studies.

First, the verbal mediation and attentional models differ in their views on the mechanism of picking up the rules behind the disposition of feedback. Verbal (or symbolic) mediation theory emphasizes the role of active agency in making a conscious representation of a problem. In the view of this theory, if the child does not construct the problem of the game by reading the game rules, for example, in the Flamingo game, she or he will gradually build up the representation by examining the disposition of feedback from the former choices. On the other hand, the attention theory gives credit to the neural system in the brain. Incremental attention is given to a stimulus followed by reinforcement, resulting in the brain becoming more sensitive to the dimension. The child makes choices in rule-like fashion as a result of the sensitization of the neural system. In the view of this theory, the child's performance would gradually improve as the practice of game accumulates. But, there is no need to suggest any subjective conscious representation.

Second, the relationship between concepts and choice-making has not been proved. The question of how concepts related to the problem affect the processes of problem solving is of much interest in this study. Children's games depend not only on concrete attributes of concepts like shape, colour or size, but also on relational or abstract concepts. Experimenters in discrimination tasks referred to concepts required for the

ability to label different objects with a common name (e.g. pear and banana as 'fruit'), the ability to select out the relevant physical dimension in problem solving (e.g. size, form, colour) or the use of relational thinking (e.g. being able to compare objects as to size - bigger, smaller). In fact, these definitions of concepts are quite restricted. The understanding of abstract concepts like exclusion or inclusion, or defined concepts like 'diagonal', cannot be readily studied by discrimination tasks. How do abstract concepts or higher-order concepts relate to the processes of problem solving?

Third, the verbal mediation and attention theories shed light on the importance of encoding variables and use of strategies in problem solving. A considerable number of studies has shown the possibility of training young children to attend to and to remember variables in problem-solving. One of the first trainings of youngsters' verbal rehearsal on memory tasks was implemented by Keeney, Cannizzo, and Flavell (1967). From a large number of 6- and 7-year-olds, children who consistently rehearsed or who consistently failed to rehearse were identified. In order to identify these two groups, children were tested on the delayed recall test in which the experimenter recorded children's lip movements. Rehearsers were children for whom lip movements were detected on at least nine out of ten trials; nonrehearsers were those for whom lip movements were absent on a similar number of trials. Half of the first group of children and all of the latter group were trained to rehearse. Children were told to whisper the names of the stimuli over and over until the memory test was given. Several training trials followed in which children were prompted if they failed to rehearse correctly. Several test trials were then given in which children tried to remember from two to five pictures. Rehearsal training was shown to be highly effective. Children who had not rehearsed spontaneously at all did so on more than 75 percent of the test trials. Furthermore, rehearsal had a clear facilitating effect on recall. Before training, children who did not rehearse recalled less accurately than either group of spontaneous rehearsers. After training, differences between groups were negligible.

This response pattern was backed by many later studies (i.e., Ferguson & Bray, 1976; Kingsley & Hagen, 1969).

Fourth, adopting the neo-behavioural model, the methodology of recognizing patterns of children's responses to reinforcements is useful for examining children's strategies in successive decision making situations. Regularity in decision making - strategy - gives clues about the processes involved in learning or attaining a concept. For example, in discrimination tasks, two strategies are easily observed when a current choice is calculated from the feedback the participant has just received. They are lose-shift and win-stay. "Win" designates a correct response on Trial 1; "lose" an incorrect response on Trial 1. There are two response alternatives on Trial 2; the participant may shift her or his response, or she or he may persevere ("stay"). Levinson and Reese (1967) reported that in young children win-stay and lose-shift strategies function separately and that lose-shift strategies are acquired earlier than win-stay strategies. They also reported that with increasing age the two strategies more likely occur in combination. Kendler and Kendler (1970) reported that all the participants - children of kindergarten, grade 2 & 6 and college students - were mostly operating on the basis of a lose-shift strategy, which above 90 per cent of all trials used in their experiment. The kindergarten children did not, on average, stay with the relevant stimulus after a lose-shift response, since they continued to make errors on subsequent trials. The higher the developmental level, the fewer the perseverative errors in the test trials. At college age, the win-stay and lose-shift strategies merged so that there were few, if any, errors occurring after the first correct response. The developmental trend in these strategies supports the mediational role of concepts in problem solving. The next section will examine processes of inference-making in goal-search frameworks.

2.5. Inferential goal searching

The previous section described how inferences are made from information. Concepts for inferences are formed through discrimination and generalization of stimuli. Research on the process of inference and its use in goal-search frameworks will be reviewed here.

Earlier work on this matter was done by Maier (1936). Maier's criterion for inference was the combination of two separate experiences in order to solve a problem. In Experience 1, the child was allowed to explore the whole of the inside of a maze. In Experience 2, the child was led into one of the booths from the outside and was shown a toy which he was allowed to play with for a short while. In the test phase the child was led into one of the other booths, from the outside, and was required to find the toy. Each child was given a series of test trials in which different combinations of starting and goal booths were used.

Inference, according to Maier, would be demonstrated if the incidence of direct routes to the goal booth was significantly above the chance level of 33%. Maier's subjects were 39 children whose ages ranged from 3 years 7 months to 7 years 11 months. He found that very few children below 6 years performed above chance level, and concluded that "the ability to reorganize past experience does not become marked until about 70 months of age."

It is, however, extremely doubtful whether this experiment constitutes a valid test of inference making. The experimental situation was based on Maier's "three tables" experiments with rats (Maier, 1932). The children had to combine their knowledge of the maze with their understanding of where they were at that particular moment.

Children below the age of six failed this task quite consistently, though when older than this they were more successful. In the experiment, there was no check that young children who failed could remember the information that they had to put together. Thus, it was not sure whether the failure was due to children's inability to co-ordinate separate items of information or because they forgot what they had to combine. No conclusion, concerning the young child's capacity for inferential problem solving, may be drawn from Maier's work.

Maier's study did, however, provide the starting point for a more thorough-going experimental investigation of inferential problem-solving in young children by the Kendlers (Kendler & Kendler, 1962b; Kendler, Kendler & Carrick, 1966), who employed a Hullian paradigm that clearly separates the learning of the behaviour segments that ultimately comprise the inferential solution to the problem.

In Kendler and Kendler (1962b), children were thoroughly familiar with the items they later had to combine. Children were shown a box with three panels. The two side panels had a button and pressing this button produces an object which drops into a tray set in its own panel. In one panel this object was a marble, in the other a ball bearing. The central panel was different. Instead of a button it had a hole, and the child must learn that dropping the marble in the hole, but not the ball-bearing, has the effect of producing a small toy. These then are the two items of information, that one side panel produces a marble, the other a ball bearing, and that placing the marble in the central hole leads to a goal (the small toy). These correspond to three behaviour segments: A-B, X-Y, and B-G. To make sure that children remembered the items they had to combine, subgoal training (A-B and X-Y) was given. Children learned to press a button located in each side panel in order to obtain the marble on one side and the ball-bearing on the other. The correct solution, the A-B-G response sequence, was to press A that yielded B and then to insert B into the central aperture, thus obtaining G. The required inferential solution was obtained by 6%, 52% and 92% from kindergarten, 3rd

grader and college students, respectively. Kendler and Kendler, thus, argued that children younger than 7-8 years cannot solve novel problems through the inferential combination of separately acquired behaviour segments.

The Kendlers' explanation of this improvement with age in inferential performance is in terms of their mediational theory of human concept development. According to the Kendlers, the difference between younger and older humans lies in the latter's tendency to make spontaneous, implicit linguistic representations.

Kendler, Kendler and Carrick (1966) enquired whether young children's inference is helped by verbal labelling. In one of their experiments using the same apparatus (a box with three panels) as the previous study of Kendler and Kendler (1962b), they required their participants to produce overt verbal labels for the subgoals during learning of the initial segments. (In the control condition, the subgoals were never labeled.) The incidence of A-B-G solutions among 5 - 6-year-old children was raised to 30% in this labelling condition. Another experiment compared the effects of verbal labelling of each segment between using the same labels and using different labels. Predictions were made that the same-label group would make more inferential solutions than the different-label group at both age levels, but that the effect would be more pronounced among the kindergarten children who were supposed to be developing representational systems, than the 3rd grade group who already did not need representational aid. As predicted, the labelling effects, the greater effect of "same" labelling than of "different" labelling, were obtained from kindergarten children but not third graders.

However, there is the possibility that labelling may have encouraged what may be described as "pseudo-inferential" solutions, i.e. solutions which followed the appropriate A-B-G pattern but which were not governed by an inference. In the Kendlers' experiment, the children were not told how they should go about solving the problem; they were asked to obtain the major goal. Possibly, they were supposed to

understand that the problem required an indirect solution. But whether they actually did so is open to question.

In discussing the Kendler and Kendler study, Bryant (1977) pointed to the arbitrary nature of the problem. Pressing a button produces a marble for no visual reason. Putting a marble in a hole produces a toy, again for no visual reason. Since there is no obvious logic to these artificial association (button-marble, marble-toy), children cannot be logical about combining them. Bryant emphasized the importance of child's interpretation of the nature of particular problems.

Hypothesising the argument developed by Bryant, Hewson (1978) replaced the apparatus of the Kendlers with more familiar and understandable material. He put the items in drawers so that the child simply had to learn which drawer to pull out in order to get a particular marble. This change of apparatus resulted in a big response change. He reported more than 60 per cent of the 5-year-old children succeeded in the task. Thus, it can be concluded that by the age of five children do have the ability to put together at least some kinds of information inferentially.

While the Kendlers' tasks were conducted in a laboratory, Wellman and his colleagues devised naturalistic tasks, involving staged sequences of real-life events, in order to examine young children's logical search. Wellman, Somerville, and Haake (1979) studied 3-, 4-, and 5-year-olds searching eight marked locations on a familiar preschool playground for a missing object, under two different conditions. Each child initially searched the locations under a control condition in which there was no information specifying that the missing item (a calculator) should be at some location(s) rather than others. Next the child participated in a series of eight games at the locations, arranged so that events occurring at two of them defined a critical search area. These events set the stage for a second, logical task, in which the child searched the same eight locations

for a camera that had been used to take his or her picture at the third location and was subsequently discovered missing from the carrying bag at the seventh location.

In the control condition, only 20% of the children went first to a location in what was to become the critical region for the logical task. By contrast, in the logical condition 76% of the children searched first in this region. The central four of the eight locations constituted the critical area; therefore if a child searched end to end, or in a haphazard manner, two of the first four searches would fall in this area by chance. In the control condition the mean number of the children's first four searches in the critical area was 2.05. It rose to 2.84 in the logical condition. There were no differences between 3-, 4- and 5-year-olds on these measures of logical search ability.

However, the 3-year-olds differed from the older children in some respects. In the logical condition, the modal first search choice of every age group was the third location (where the picture was taken). When just those children whose first search was at this location were considered, it was found that the second searches of 3-year-olds were less likely to be in the critical area than those of 4- and 5-year-olds. After these inaccurate second searches, 3-year-olds tended to return to the critical area, but they also showed a greater tendency than older children to repeat a search of a location, in particular of the third location. These findings raised the possibility that the younger children's searches at this location were determined primarily by a strong association of the camera with that place, rather than by the knowledge that that place was the last point in the sequence of events where the camera was present.

To distinguish between the two possibilities, the simple association of the camera with the place and the knowledge that the place was the last place the camera was seen, Haake, Somerville, and Wellman (1980) conducted a second study, on a different playground, incorporating two logical search conditions. One condition was identical to that of the previous study, and the new condition differed only in that the child's

picture was taken at each of the first, second, and third locations. This new condition was designed to establish associations of the camera with locations outside the critical area. If young children's searches were guided primarily by associations, they would be expected to do worse in the new condition than in the original one. The searches of 3- to 9-year-olds were not different under these two conditions and were comparable to the logical search scores in Wellman et al.'s (1979) study. There was also no difference between the performance of younger and older children. Haake, et al. concluded that the children's searches were not guided by associations, and that they were making logical inferences about the location of the missing camera, on the basis of events occurring in the spatio-temporal sequence. This conclusion was supported by the results of a similar study by Anoshian, Hartman, and Scharf (1982).

Previous studies have shown that young children are capable of drawing logical relations between two or more separate items or events. The next question is why the young children in the first study (Wellman, et al, 1979) were significantly less correct than the older children in their second search. Although the naturalistic tasks were presumably easy for young children to comprehend, the search problems were actually introduced incidentally to the child. The young children may have been less able to understand that there was more than one location that might contain the object. This hypothesis will be examined in the next section 2.6.

2.6. Possibility, Impossibility & Certainty

How do children become certain about the position of the Flamingo? An adequate representation of the situation of the Flamingo game requires understanding of not probability, but impossibility and possibility. How do children deal with options of impossibility and possibility? What if one choice turns out to be wrong?

The gain-maximizing strategy and the probability-matching hypothesis explain differently how the player responded to happenings in a game situation. The gain-maximizing strategy portrays the player as an agent who represents the problem and calculates outcomes of moves. The basic assumption is that one can determine the strategies which would best maximize goals in every type of strategic situation (Rapoport, 1990).

The probability-matching hypothesis, on the other hand, suggests that a game player generates and holds multiple expectancies that reflect the probability of a particular happening at a particular situation (Kingstone & Klein, 1991). The hypothesis portrays rather a passive picture of the player. The assumption of probability matching is that children learn consciously or not to match their response ratio (the relative frequencies which they predict for each of the events) to the actual probabilities of occurrence of the events. In this view, strategies are not necessarily consciously formulated or the product of a conscious or rational choice, even in mathematical calculation (Siegler & Jenkins, 1989).

Inhelder (Inhelder & Piaget, 1958) argued that the understanding of possibility is indispensable for hypothetico-deductive or formal thinking. When the child reaches the stage of concrete operations (7-8 years), the decentering process has gone far enough

for the child to be able to structure relationships between classes, relations, and numbers objectively. Operational classifications are constructed at around this age, with the understanding of inclusions and their qualifications. A new series of operational possibilities - disjunctions, implications, exclusions, etc. - arises only after the stage of concrete thought.

However, there are protocols cited by Piaget and Inhelder (1951) showing that the intuition of chance emerges at the pre-operational level. For example,

Roulette experiment: "Mon. (4.11). 'Can we tell where it will stop?'"

'No, because if we say it will stop at blue, and then it goes past blue, we won't know.'"(p. 74.)

Researchers have pointed out (Gellatly, 1987; Murray, 1987; Siegler, 1991) that young children do not view conclusions as being true by reason of logical necessity. Such a failure to distinguish between logically-necessary and empirically-likely outcomes would explain young children's eagerness to verify by empirical means relations that older children and adults view as purely logical matters. It also would explain the seemingly opposite tendency of young children to reach conclusions when the evidence does not logically allow them to do so.

To assess children's sensitivity to logical certainty, Pieraut-LeBonniec (1980) investigated children's discrimination of certain and uncertain (but possible) situations. The experimenter employed a box that had holes of two sizes in its lid. The larger hole allowed marbles of a certain diameter and thin sticks of a smaller diameter to fall into a drawer inside the box. The smaller hole permitted only the thin sticks to fall into the drawer. After children were made familiar with the box, it was placed behind a screen. Children were presented with the following problem: "I am putting something into the large hole. Can you be certain what it is without looking inside the box?" The correct answer was to say that one could not be certain, and that the box had to be inspected

because either the marble or the stick could be placed in the large hole. The experimenter reported that children began to recognise uncertainty in this context at about 9 to 10 years of age, where 45% of all children provided correct responses after feedback. However, most younger children were correct on trials where certain conclusions could be drawn (e.g., when questions were asked about the small hole where only the sticks can go through). The finding that the comprehension of conclusions about certainty emerges earlier than the understanding of uncertain conclusions was confirmed in other studies (Acredolo & Holobin, 1987; Byrnes & Overton, 1986; Holobin & Acredolo, 1989).

Fabricius, Sophian and Wellman (1986) investigated children's inferential reasoning in possible, impossible and certain situations. The task was to find a picture to match a description. There were three kinds of descriptions which led to a conclusion that the case was as described (certainty) or that the case could not be so (impossibility) or that one could not reach a conclusion (possibility under uncertainty). Descriptions contained quantifiers like "all cats in it" or "a duck in it". Significant differences were found in children's ability to distinguish between certain, impossible and possible situations. While 5-year-olds distinguished the impossible situations from the possible but uncertain situations, 3-year-old children only distinguished the certain from the impossible or the possible situations. The younger children failed to determine that negative information (absence of an object) could sometimes support an inference. Even the older children were likely to rely on positive information (presence of object) instead of negative more than half the time in the impossible situation. The study showed two characteristics of inferential problem solving. Firstly, children were more likely to rely on positive information than on negative. Secondly, the comprehension of conclusions of certainty emerged earlier than the understanding of impossibility in a concrete context.

Byrnes and Overton's study (1986) expanded Pieraut-LeBonniec's experiment (1980) to explore content effects on discrimination between certain and uncertain situations. Three tasks were presented: the box task used by Pieraut-LeBonniec (concrete context), an envelope task (causal context: "I am putting a picture into an envelope and the picture I'm putting in the envelope shows what makes the boy cry. Do you know for sure which picture it is or do you have to look inside the envelope to see which picture it is?"), and a formal syllogism task (propositional context: e.g., "If it has rained, then the grass is wet.") Comparisons were made of children's performances between the three tasks. In the concrete context, the Pieraut-LeBonniec box task, the percentage of children at each grade who made no errors on uncertainty trials increased as the age of the children increased, with 13%, 71%, and 83% from 1st, 3rd, and 5th graders being correct, respectively. These correct performance rates were somewhat higher than those reported in Pieraut-LeBonniec. In the causal context, the envelope task, the percentage of children at each grade who made no errors on uncertainty trials were 13%, 71%, and 100% from 1st, 3rd, and 5th graders, respectively. In the propositional context, the syllogism tasks, there was no developmental change for Modus Ponens (e.g., "It has rained, is the grass wet?") and children at each grade performed quite well. Performance on Modus Tollens (e.g., "The grass is dry, hasn't it rained?") dropped significantly between the first and third grades, and between the third and fifth grades. For Affirmed Consequence (e.g., "The grass is wet, has it rained?") and Denied Antecedents (e.g., "It has not rained, is the grass dry?"), the realization that there are other possible causes of grass being wet only enabled children to judge the arguments as uncertain. Although significant improvements occurred for both Affirmed Consequence (e.g., "The grass is wet, has it rained?") and Denied Antecedents (e.g., "It has not rained, is the grass dry?") at each grade, performance on these uncertainty arguments was above chance at only the fifth grade. The understanding of discrimination between certain and uncertain conclusions was mastered in a concrete or causal context (i.e., the box and the envelope task) by the fifth grade, but with respect to a propositional context (i.e., conditional syllogisms), this understanding seemed to

emerge in the fifth grade. These studies demonstrate younger children's appreciation of concrete and causal uncertainty and apparent age differences in performance between concrete and propositional tasks.

While Byrnes and Overton argued that the children's failures were due to the lack of formal operations, Braine and Rumin (1983) suggested that a response bias prevented children aged 6 to 10 years from correct performance on tasks assessing the ability to identify undecidable situations. According to Braine and Rumin, young children have a strong expectation that all problems have unequivocal solutions and this leads them to endorse only one possible solution, although they might actually have been aware of two or more equally viable alternatives.

Hypothesising the argument developed by Briane and Rumin, Acredolo and Horobin (1987) investigated developmental changes in children's avoidance of premature close, which is the tendency to offer only a single solution to any problem that, because of insufficient or ambiguous information, logically permits more than one solution. Children of 1st, 3rd, 5th & 6th grades were presented with a simple computer game in which they had to deduce the possible sizes of one item relative to two others on the basis of visual comparison of the sizes of the two items and a written clue concerning the location(s) of the biggest or smallest of the three items. Some problems had single solutions, whereas others had multiple solutions. On the single solution problems, reasoning errors were less frequent than would be expected by chance even among the first-grade children. On the other hand, no observation was made until 6th grade of voluntary detection of the possibility of more than one correct solution when there were multiple solutions. Acredolo and Horobin pointed out that children almost always commit themselves to a single alternative in choice paradigms, despite an awareness of the viability of other alternatives. The finding of a strong tendency to close on single alternatives among 7- to 9-year-old children was backed up (Horobin & Acredolo, 1989). The experimenters also reported that corrective feedback improved the

frequency with which first- through fifth-grade children detected multiple solutions on reassessment.

Bryant and Roazzi (1992) investigated developmental changes in children's avoidance of premature close in number. 6-, 7-, and 8-year-old children were asked for a number which is bigger than X but smaller than Y. The children were asked to choose a number in one case where there was only one possible correct number and in another case where there were several correct answers. In the task which had one correct answer, even the younger children were quite good at inferring a "critical area", particularly in their first choices. In the other task, the proportion of correct first choices that were immediately followed by a correct second choice was quite high in the seven and eight year olds but a great deal smaller in the six year olds. It seemed that the six year olds had difficulties in holding the "critical area" in their mind and eliminating alternatives one by one. The findings of Bryant and Roazzi's study show that systematic use of inference in numbers develops up to 7 years of age.

However, the results can be thought of as processes derived from children's familiarity of number sequence rather than an abstract understanding of the multiple possibilities. Hypothesizing that without necessarily being able to understand logical relations sufficiently to distinguish consistently between connected premises, young children are able to provide correct responses to some forms of syllogism, Markovits, Schleifer, and Fortier (1989) presented children with logical or illogical syllogisms. The participants were asked to answer "yes", "no" or "don't know" to the syllogisms and to provide justifications for their responses. All the children were reported to produce a very high level of correct responses to the logical problems. However, the 6-year olds and to a certain degree, the 8-year-olds produced similar patterns of responses to the logical and the illogical problems. For the 6-year-old children, 90% of logical problems for which subjects referred to premises in their justification were correctly resolved. In the case of illogical problems, 78% of those with referral to premises received an

equivalent response. For the 8-year-olds, there were significantly more non-equivalent responses when the illogical problems were presented first, whereas there was no significant difference for the logical-first and random presentations. The 11-year-olds did not provide conclusions for the illogical problems, while resolving the logical problems.

The results were also consistent with the notion that much of the correct responding observed may be due to the use of a strategy that relies on the "atmosphere effect", in which the response was chosen as a function of the positive or negative mood of the premises (or some similar strategy that did not involve understanding the relations of necessity between premises and conclusions). The 6-year-olds showed some signs of differentiating logically consistent and logically inconsistent premises, primarily in terms of their responses to the latter although not in terms of their justifications. The experimenters noted the invention of an imaginary link between premises. This form of justification appeared almost exclusively with respect to illogical premises, and only in 8- and 11-year-olds. The invention of a link between premises that did not go together with each other could be an indication of the beginnings of an explicit understanding of the distinction between logical and illogical forms of argument. The 6-year-old children retained the overall tendency to respond similarly to illogical and logical problems, whereas the 8-year-olds showed a definite ability to differentiate the two when presented with the illogical problems first. The 11-year-olds differentiated the two regardless of order of presentation. These results showed the ability to differentiate reliably between premises that are logically related and that are not developed with age. These results are consistent with a variety of others that indicate that there are important developmental steps in the understanding of elementary inferential principles (Byrnes & Overton, 1983; Fabricius, et al., 1986; Pieraut-LeBonniec, 1980)

If children understand elementary inferential principles, the next question is how these elementary principles enable the child to anticipate outcome of her or his choices in a

more complicated logical structure. Piaget and Garcia (1991) investigated the development of operative logic of inclusions and exclusions, using a network of tunnels. Children were asked to choose a path in the tunnels. One tunnel divided into two primary branches, each of which turned into two paths. From each of these four paths two further paths separated. Each of the final paths ended up in one of eight garages. There were windows on each path. A car started at the entrance of the main tunnel and was placed in one garage. The car was tied to a long thin ribbon, which lay between the start and the garage. The task was to determine which garage a car was in by opening as few windows as possible.

Piaget and Garcia proposed 3 stages of development of inclusions and exclusions. Stage 1A is characterised by a lack of inclusions in the direction of the paths' construction. When asked to trace a car hidden in one of eight garages by opening a few windows along the paths, young children aged 4-5 year-old randomly opened the windows nearest to the garages. They said that it was better with those windows because they were closer to the garage.

Children at stage 1B also conducted a semi-empirical and semi-deductive procedure. They started with windows in the middle or prime branches and went up and down across levels, sometimes skipping one or two. Inferences were still invalid.

Children at stage 2, aged around 7 to 8 years, considered the network in its entirety and began to constitute an operative grouping. However, they still could not exclude impossible branches and garages systematically.

Only children aged 10 -12 at stage 3, determined the goal place by opening windows which did not show the ribbon. This negative necessity is the most difficult aspect for the participant to accept. Piaget and Garcia took these findings to reveal the inability of young children to infer the part-whole relation, to group items according to their

classes, and to operate inclusions and exclusions that are prerequisite for a systematic search.

In the Flamingo game, children have to infer the goal place from abstract and rule-defined information. COLD conveys information of the impossibility of a certain area for achieving the goal, whereas HOT or WARM indicate possibility. Inclusions and exclusions would render appropriate representations of a zone of impossible places of the goal and/or a zone of possible ones. To anticipate the results of inferential actions, Piaget and Garcia argued that coordination of inclusions and exclusions has to develop up to a certain level. It is expected that older children would coordinate exclusions of impossible locations of the goal and inclusions of possible locations, whereas young children would have trouble in representing the problem of the Flamingo game.

In conclusion, we can draw two very different pictures of game play. The first picture takes the viewpoint of the gain-maximizing strategy. The player in the 'Find the Flamingo' game, for example, generates a representation of possibility and impossibility, and maximizes the chance of reaching the goal through choices, becoming a player who achieves successes by representing the game structure. Faced with HOT and WARM, the player will map a zone of possible goal places related to the current place. Faced with COLD, she or he will map the zone of impossible ones. The choices will be guided by these inferences. Systematic use of representation requires the exclusion of impossible goal places and inclusion of possible places. The player has to see the logical certainty of predicting a certain goal place through coordinating the areas of possibility and impossibility.

The second picture comes from the viewpoint of the probability-matching hypothesis. The player's choices follow the likelihoods that have developed through accumulated reinforcements. As the attention theory proposes, the child tends to attend to something that is more likely to happen than to any alternative, because of the past reinforcements.

The player does not generate specific and personal representations. She or he makes choices nearer to HOT or WARM and further away from COLD, because experience guides her or him that the target is found in that way. Research on inferential action in the Flamingo game would surely broaden knowledge about the nature of children's thinking in the context of game.

2.7. Gender issues in computer games

One of the main issues relating to computer game play is that males, on average, perform better in visual and spatial skills (Halpern, 1986; Kerns & Berenbaum, 1991; Linn & Petersen, 1985; Okagaki & Frensch, 1994; Petersen & Crockett, 1985) and these differential spatial abilities might be reinforced by male dominance in computer game play (Hoyles, 1988; Hughes, Blackenridge & MacLeod, 1987). According to Halpern (1986), there is "still some confusion about the youngest age at which gender differences in spatial abilities are found" (p.51). In a large study of over 1,800 public school students, Johnson and Meade (1987) used a battery of seven spatial tests tailored to the developmental levels of the children and concluded that a reliable male advantage in spatial performance appeared by age 10. How do these differences in spatial abilities affect each gender's computer game practice?

Questioning the gender difference in favour of males on spatial cognition reported in some studies as possibly the result of exposure to the environment of these games, McClurg and Chaille (1987) investigated the effects of computer games on spatial cognition. It was hypothesized that boys and girls would perform spatial cognitive tasks with similar proficiency after the same amount of practice. Children of grades 5, 7, and 9 played two computer games: 'Factory' and 'Stella 7'. 'Factory' required mental manipulation of three dimensional "products". To replicate a "challenge product" the user must visualize the movement of the product through an assembly line of Punch, Stripe, and Rotate machines. To be successful, the user was required to develop strategies for optimum sequencing. The task analysis involved visualizing a number of transformations in a series, this requires the use of visual memory and constant updating of the visual image. 'Stella 7' required recognition of three-

dimensional objects appearing at different distances, moving at varying speeds and changing orientations, and viewed from different positions.

The Mental Rotations Test (Sheppard & Metzler, 1971) was administered as a pre-test the week before the treatment began and as a post-test the week after treatment ended. The “Factory” group and the “Stella 7” group met for forty-five minutes twice a week for six weeks. During the last two sessions in the sixth week, students were asked to predict the resulting product before the computer “manufactured”, and display it in the “Factory” group. Each week about ten minutes were devoted to students sharing techniques and strategies that they found useful. All children in the “Stella” group and the “Factory” group improved more than children in the control group. It is interesting to note that in this study there was an initial gender difference in spatial skill, with boys performing better than girls; it is not clear from the report whether this difference continued to be present at the end of the study. Miller and Kapel (1985) found a positive effect of similar computer games on two-dimensional mental rotation in seventh and eighth graders.

Subrahmanyam and Greenfield (1994) investigated the effect of computer games on spatial ability. A group of fifth grade children played an action game, “Marble Madness” (Harvey, 1986), which involves the use of the spatial skills of guiding objects, judging speeds and distances of moving objects, and intercepting objects. Another group of the same aged children played a computer word game, “Conjecture” (1986), which does not involve any spatial skills. Children played the games for a total of 2 hr and 15 min. with three sessions of 45 min each on different days. Spatial abilities were measured before and after the experiment. Boys were significantly better than girls during pre-test assessment. However, no gender difference was found in the post-test assessment. The action game was significantly more effective than the word game in improving spatial performance on the post-test assessment; there was no significant interaction of gender with experimental treatment. However, the action

game practice was more effective for children who started out with relatively poor spatial skills. The action game play equalized individual differences in spatial skill performance, including those associated with gender.

It has been found that boys are especially likely to play games requiring aggressive competition (Cunningham, 1995; Linn & Lepper, 1987), whereas girls are turned off by violent themes (Malone, 1981). This tendency might be well associated with the frequency of boys' greater success with electronic games (Griffiths, 1996).

Questioning how the nature of a game affects girls' and boys' playing, Littleton, Light, et al., (1992) manipulated the same game format with different characters to meet each gender group's supposed taste. They presented honeybears in one version and pirates in the other. The task was couched in terms of a quest involving 3 male characters who must go on a journey to collect a missing crown ('Pirates'), or a logically equivalent task involving 3 teddy bears who must go on a journey to collect a missing pot of honey ('Honey bear'). They reported that the girls' performance was highly dependent on the represented task domain - their performance was far superior when using the relatively more 'gender-neutral' software involving teddy bears, while the boys were seen to perform equally well irrespective of whether they were confronted with any version of the characters. On a re-gendered version of the 'Pirates' game with the characters of 'Princesses', it was reported that the girls' performance was rather poor when using the 'Pirates' software and somewhat enhanced when using the 'Princesses' software, and that the performance of the boys was consistently high across both software types (Littleton, 1996). The girls' lower average scores in some types of "adventure" games could be considered to discourage them and make them more reluctant to play computer games.

In Hay and Lockwood (1989), use of strategies on a computer-generated hunting task was examined in boys and girls. 6- to 10-year-old children were introduced to a

computer game that simulated a hunting problem for an animal forager. Girls and boys were equally successful at this stereotypically masculine spatial task, and both girls and boys foraged optimally in terms of the predictions of the ecological theory of optimal foraging. Only one sex difference appeared: girls tended to use more careful strategies than did boys. The excessive caution of girls in spatial tasks was well documented in Linn and Petersen (1985).

In conclusion, research has shown that the nature of the game affects the level of engagement with it differently for boys and girls (Linn & Lepper, 1987; Malone, 1980; Littleton, et al., 1992). It is important for researchers using computer games as a window on children's thinking to be cautious in choosing a gender neutral game as a research tool. More research is needed on how girls and boys differ in their interaction with computer games and how we could help both genders be interested and enhance their cognitive skills while playing computer games.

2.8. Methods for investigating children's game activity

Both behavioural and cognitive approaches have been adopted for examining game behaviour, as the two approaches are complementary to each other. Behaviourists assume that human choice is nothing but the result of respondent conditioning. These scholars focus on links - connections or associations - between observable stimuli and observable responses. The terms "association" and "connection" refer to the phenomenon that links a particular stimulus with a particular behavioural response which an organism makes as a result of exposure to that stimulus (Isaacson, Hutt & Blum, 1965). Skinner (1938) views choice in terms of the competition between individual operants. In a choice situation, Skinner sees each alternative and its outcome as a separate contingency, strengthening or weakening an animal tendency to make each response. For Skinner, choice among alternatives is not an object of study in itself; it is merely the outcome of combining two individual responses to two stimuli.

If we follow behavioural tradition, the objective of analyzing these game behaviours is to look for a link between cue and choice behaviour. "Cue" is here defined as a stimulus that points toward a goal without requiring any necessary conscious knowledge of its significance. One common measure of choice is the behaviour ratio (BR) introduced by Tolman (1938). The numerator of the behaviour ratio is the number of times an animal chooses one alternative and the denominator sums the total choices of all alternatives. The behaviour ratio is like a probability in that it varies from unity (when alternative 1 is chosen exclusively) to zero (when alternative 1 is never chosen). Another common measure of choice is the fraction of two choices, the relative choice (RC).

How does behavioural theory in a game situation explain the cause of the action? It is only the last action that is reinforced in terms of the goal. The analyst does not know what the player's viewpoint is. In game theory, a cognitive analysis is a very important aspect of the game, because it tells the subject what contributes to winning. Some children might not have any idea of what makes for winning. They just do anything. Other children might have an idea that is not as elaborate as adults have. They go close to a HOT and away from a COLD, so they have more chances of winning. Another group of children might operate more systematically. Faced with a HOT, they realize that there are a maximum of four possibilities. The game player needs to distinguish probable, possible and impossible solutions. These are the reasons why game theory adopts a cognitive approach, which studies the subject's thinking about the strategies that she or he develops.

In games, the player finds herself or himself faced with the necessity for making choices at each point. She or he keeps tracking, searching, making judgements based on the information already obtained. Seeing games as goal-directed transforms playing games into problem solving. In the behaviourist view, each attempt to solve a problem consists only of a chain of responses. Problem solving, then, is defined as combining separate items of information in order to come up with a solution that is a completely new one. From the Piagetian and information processing point of view, however, problem solving is searching for an operation which bridges the gap between the state of affairs which is to be found now, and a more desirable goal state. Developmental changes in problem solving are known to occur along a wide variety of dimensions: planning, encoding of the problem, strategies for solving the problem, and the ability to learn from experience with it.

The cognitive effort required to make a decision can be measured in terms of the total number of elementary information processes needed to solve a particular problem (Payne, Betterman & Johnson, 1993). Thus, one way to look at the player's making of

inferences is to understand the amount and types of information (see Payne, 1976; Thorngate, 1980). At one extreme are simple information board procedures. The Flamingo game is a good example of these information boards, which consist of three kinds of spatial information; adjacent above, below or on the right or left, diagonally adjacent and non-adjacent regarding the goal.

It may be assumed that there are two kinds of game players: those who think that getting a goal happens purely by chance, and those who try to work out a string of rules governing reaching a goal. For those who are concerned about the rules, the rules will dictate their choice behaviour. They can develop strategies for playing games. They know what their choices mean. On the other hand, children may still play the game without understanding the rules. If they do not understand the rules of the game, the choice will be random. Some choices will be perhaps reinforced. Winning in a game is generally regarded as reinforcement, since it makes the child more likely to play the game. However, applying the concept of reinforcement to play is not as easy as it looks. When the child plays the Flamingo game and she or he wins, what is reinforced? These questions could be answered by looking at the choice behaviour and listening to what the player says. Inferences about how the child is reasoning are made on the bases of what they do and how they talk, whether autonomously in the game process or prompted by the experimenter - what Piaget called clinical discourse. Is there a relationship between what they say is their strategy and what they actually do in the game? In this research, for studying children's judgements about the possible goal, paper tests were organized. The scores on the paper tests were compared to the performance on the game.

A correct judgement would result in a move that is in accord with the directional cue information provided. For example, a correct move given a COLD cue would be to move to a space far from the present game position. A correct move given a HOT cue would be to a space in contact with the present game position (either to the top, bottom

or sides of the current space). A correct move given a WARM cue would be to a space diagonally touching the space of the current game position.

Each move made in the process of the game was examined to see whether the next move was based on a valid inference from one or two successive bits of the previous information. Two methods of analysis systems were introduced. The first method is called 'single information inference', in which information from the immediately preceding move is counted as the base of inference. The other method is called 'double information inference', in which information from the two preceding moves is counted. These procedures of data analysis allow us to look at the way children integrate the information about spatial relation, and to identify which combinations of relations children find to be most demanding.

Intervention in the game process is based on the assumption that the player's becoming more aware of the structure of the problem would enhance the learning process. The awareness comes from the learners reflecting about their own representation of the problem. To be reflective is to stop and think before committing any action. The guided planning helps to plan before doing. The more knowledgeable person acts as an active helper to stop the learner and encourage them to think beforehand about inferring the goal from the current position. Pausing is another compelling way to improve performance. The relative effects of guided planning and pause were investigated.

2.9. Drawing the threads together: the study of children's thinking in the context of computer games

The thesis reports on four experiments analyzing children's inferential game play in a computer context. The main themes of the research questions in each experiment are as follows: the computer game as a window into children's thinking, developmental and individual differences and learning, the effects of intervention for planning, and prediction of the Flamingo game performance.

2.9.1. The computer game as a window into children's thinking

The first experiment was an exploratory study. Three sets of research questions were raised. The first set of questions concerns the perception of the game and the use of rules. Do children represent the game as a random event or a problem-solving task? Do children make inferences from the game rules for choices in the game?

The second question concerns the measurement of inferential problem solving. Direct observation of the game performance and questioning on making inference from the game rules were employed. The number of moves until the completion of game was counted in the observation. The questioning was carried out with each of the three kinds of spatial information (HOT, WARM, and COLD) printed on the test papers.

The child was asked where the target could be and could not be. Would the number of moves and scores on the tests be correlated?

The third question is whether the computer on which children play the game affects problem solving in the game. Do the computer game effects come from the computer itself or from the game activity? The 'Find the Flamingo' game was presented either on the computer, a board, or cards.

2.9.2. Individual differences and learning

The second experiment was a replication and extension of the first one. It was intended to answer three research questions. The first set of questions concerns individual difference, which are defined in this study as the child's age, gender and inferential ability. Are there developmental differences in the making and use of inferences in the game process? Are young children able to link the game rules and the information during the game process? Do some children consistently perform better than others?

Another research question concerns whether children become more proficient in using a particular form of logical reasoning when it is embedded in a game that they can play repeatedly. The question here is whether children understand a particular type of logical structure and whether they can use the inferences more and more in the course of playing a game.

The third question is whether there are differences in the making and use of inferences from each types of information. Do children make inferences similarly concerning all

kinds of information, or does different information pose different levels of difficulty in making and using inferences?

2.9.3. Training in planning

The third experiment was a training study. The research question was whether children's inferential problem solving can be enhanced by adult intervention in planning. Do children make links between the game rules and the clues? Do children plan to use the inferences in their current position? Are there differences in these between inclusion and exclusion inferences? For these research questions, a simplified game with two clues – inclusion and exclusion without directional cues – was used. A brief training session was held in two conditions: guided planning and pauses. Guided planning actively poses to the player questions about the possibilities and impossibilities of the goal place before making each choice. In the pausing condition, pauses would provide the player with a definite time to plan before making a choice. Inclusion and exclusion inferences were studied in each condition.

2.9.4. Prediction for inferential game play

The last experiment concerns prediction for the inferential game play. Do children understand the inclusion of multiple possibilities? Do children coordinate inclusions and exclusions to find the target with efficiency? For these research questions, the study compared children's performances across the Flamingo game and two inferential

tasks: the Number tasks (Bryant & Roazzi, 1992) and the Tunnel task (Piaget & Garcia, 1991). The Number task was used to investigate the inclusion of a 'critical zone' of multiple possibilities while excluding impossibilities. The Tunnel task was used to investigate the child's anticipatory actions in a structure that necessitates coordination of inclusions and exclusions. Through the examination of the child's responses to the tasks, it was intended to identify the level of the child's operative logic of inclusions and exclusions. The child's level of the operative logic would inform about the way she or he uses inferences from the game rules and allow to predict her or his performance in the game.

An overview of the studies summarizes the aims, rationale, and measurement for each experiment, and the age and number of the participants.

Table 2.1. Overview of studies

	Aims	Rationale	Research Questions	Sample Size	Sample Age	Treatment Group	Measurement
Study 1	Representation of the game Measurements of inferences Comparison of three media	Representation/ Performance Motivation	Inferential problem solving Differences in performance across media	50	8yrs 3m	Computer Board Card	No. of Cards Scores on paper test of Spatial inference
Study 2	Individual Differences and Learning	Inferences in game and task	1. Developmental trends 2. Individual Differences -Backward learning curves	32 55	7yrs 4m 10yrs 4m	Computer Cards	No. of Cards Scores on paper test Inferences from clues
Study 3	Effects of intervention in planning	Planning in use of inferences	Effective intervention in planning on computer	32 31 32	6yrs 5m 7yrs 5m 8yrs 5m	Guided planning Pause Control	Inferences from inclusive & exclusive information
Study 4	Prediction for the inferential game performance	Operation of Inclusion and Exclusion	Correlation between the tasks and the Flamingo game	50	6yrs 10m		Construction of cards Numbers inferred Piagetian stage

Chapter 3. Study 1

3.1. Aims

The present study aims to explore four aspects of children's inferential problem-solving in a computer game context: the representation of the game, the ability to make inferences from the game rules, the measurement of inferential problem-solving in the game, and the effect of the computer on the game play.

3.2. Introduction

In the review of the literature, various aspects of children's game play in the context of computers have been discussed in relation to the framework for analyzing problem-solving in playing computer games: the effects of game and task contexts, development of the concept of a rule in games, the effects of rule content, inferential goal searching, understanding of possibility, impossibility and uncertainty. As exploratory in the series of experiments for the thesis, this study had four purposes. The first three purposes were to explore children's inferential game play in respect of the representation of the game, inferential problem-solving and its measurement. Research questions follow: Do children perceive the game as a chance event or a task? Do children make inferences using the game rules? How can children's inference making be measured? The fourth purpose of this study was to examine whether the computer that the series of

experiments of this thesis used as the main research tool affected children's game behaviour differently from other traditional game tools. Computers give immediate audio and visual feedback that a board or card games cannot offer. Does the immediate feedback aspect of the computer affect the way children play the game?

A problem-solving type game, 'Find The Flamingo', was chosen for the study since many educational computer games belong to the problem solving type of activities. The game is structured by the following rules.

If the Flamingo is there, you win.

If the Flamingo touches your box sidewise, you are HOT.

If cornerwise, you are WARM.

If they don't touch at all, you are COLD.

The game demands various steps in problem-solving: making spatial links between the clues and the goal from the rules stated in an *if-then* form, affirmative or negative, mapping the links in the current situation at each move, and trying one of the possible places while holding in mind possible and impossible positions of the goal. The inferential problem-solving of the game is defined as a compound of two elements: first, the child *makes links between the clues and the Flamingo place based on the game rules*. Second, the child *maps the relations on the current information and position*. The clue HOT specifies the position of the goal with respect to the current location that has intrinsic axes, such as top and bottom, and left and right; for example, 'The Flamingo is at the top, bottom, left or right of this square'. The clue WARM specifies the position of the Flamingo with respect to the current location according to 45 degree axes; for example, 'The Flamingo is at one of the corner squares'. The clue COLD specifies the negative relation between the Flamingo and the current location; for example, 'The Flamingo is not at the adjacent squares'. Inferring from COLD requires the concepts of the impossibility of a certain area for the Flamingo and of the possibility

of the rest, whereas HOT or WARM require the concept of the possibility of certain places for the Flamingo.

What decides whether children will make inferences from the rules when they make choices? This is initially a question about the player's understanding of the game, as a chance event or as a task that can be tackled with logic. For those who understand the game as a problem-solving task and use each clue to work towards the goal, their performance would be enhanced by the ability to make inference from the game rules. For those who cannot make inferences from the game rules and play the game without realizing its problem solving nature, or those who are capable of inference-making but do not use this ability in the game, their performance would be bad.

Inferential problem solving of the game can be measured in two ways. Firstly, the number of moves the player makes until the completion of the game can be measured. The weakness of relying on such measurement of the problem-solving performance is the element of luck in the game. Children may hit the goal without any consideration of the rules. The first clue can give any of these items of information: the Flamingo, HOT, WARM or COLD. In most games, an initial clue of COLD needs more moves than a HOT or WARM clue if the player infers from the game rules. There are two ways to enhance the measurement. One is that each player plays the game more than once. Many games allow a more equal distribution of the element of luck to each player. The mean number of moves of all the games would represent the player's ability more fairly than the number of moves in a single game. A comparison of mean numbers of even and odd sets of games would work as a reliability test. Secondly, the inferential problem-solving can be measured by examining the child's making of possible and impossible inferences from the clues. For this, direct tests can be carried, outside the context of the game. Test scores can also be used as a validity test for the measurement of the number of moves.

3.3. Method

3.3.1. Design

The study had a mixed between-subjects and within-subjects design. It aimed to answer the research questions in these ways:

1. by examining the relation between the game and word clue test performances, and
2. by comparing the performances of the computer, board and card media groups.

Using the first method, the relation between the children's game performance and the task performance could be examined. Significant positive relation in the game and test would suggest the reliability of the measurement of the inferential problem solving in the game situation.

The second method was used to see whether using the computer affected the problem solving performance.

3.3.2. Participants

The 50 children, aged 8 years (mean age = 99.88 months, ranging from 7 years 9 months to 8 years 10 months, S.D. = 4.1 months) were all the children of year 2 in two

primary schools in a middle class area of Inner London. They were allocated into the three media groups in the order they entered the experiment room.

Table 3.1. Number of participants across the three media and gender

Gender\Medium	Computer	Board	Cards	Total
Girls	9	12	10	31
Boys	7	6	6	19
Total	16	18	16	50

3.3.3. Apparatus

The computer, the board and the cards as media all presented children with the same game structure. Each game apparatus had a 5x5 grid to allow 25 word clues. There were two distinguishing features among the three media. First, there was a review key that provided a view of all the moves made in the computer game. Second, the computer game and the board game had a cover picture of a safari, whereas the cards were plain green in front. What the player faced in the beginning of the game was the safari picture for the computer and board groups, and 25 green cards in a wooden frame for the cards group.

a) Computer game

An Apple Macintosh LC with 13 inch high resolution colour monitor was used for the computer group. There was a review key in the left hand corner of the screen to let the child see all the places already tried during the game (see Figure 3.1). The player was encouraged to use the review key after every move. Because the programme is not sold in the UK., this was a new game to the children.

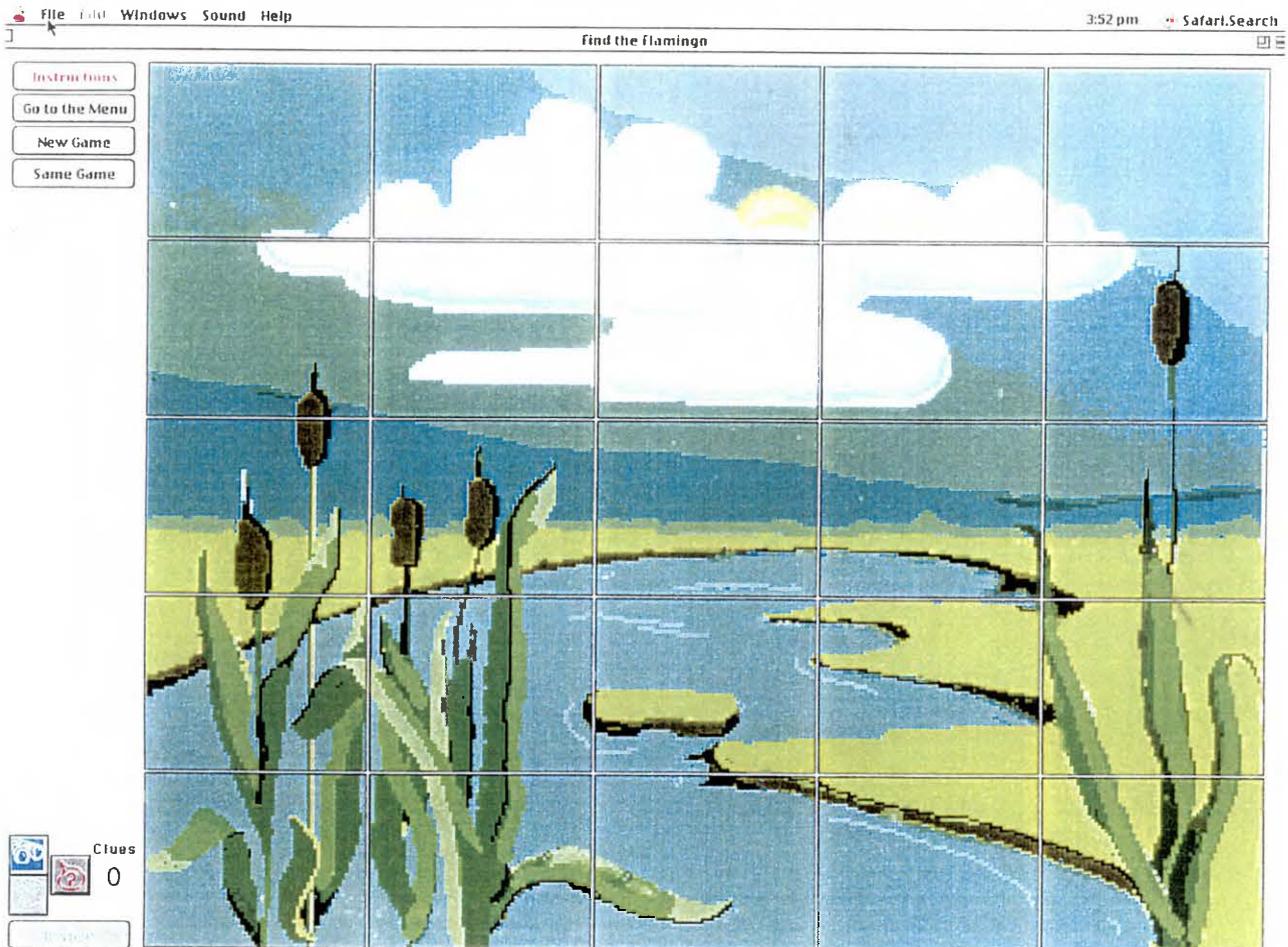
b) The board game

This game has the same structure as "Find the Flamingo" but it was presented in a different format. A transparent plastic double-deck board was designed for the game (see Figure 3.2). It was covered with 5x5 pieces which formed a picture of a safari on

the top deck (a similar display as in the computer game). Each piece had a knob that allowed the child to lift it away. On the lower deck, the game rule-governed word clue sheets were already introduced before the game and withdrawn after the game. As the player lifted a cover piece away, she or he saw a clue word or the Flamingo through the transparent plastic board.

c) The card game

Two wooden boards were used for the convenience of the procedure. Each board (52.5 cm x 43 cm) was made up of 5 rows and 5 columns clearly marked with intercept sticks for laying 25 cards. The cards had a green cover, unlike the computer game and the board game (picture above). Each card showed a word: HOT, WARM or COLD (see Figure 3.3).



	Cold	Cold	Cold	Cold	Cold
	Cold	Cold	Cold	Cold	Cold
	Warm	Hot!	Warm	Cold	Cold
	Hot!		Hot!	Cold	Cold
Clues 25	Warm	Hot!	Warm	Cold	Cold

Figure 3.1. Pictures of the computer game



Figure 3.2. Pictures of the board game

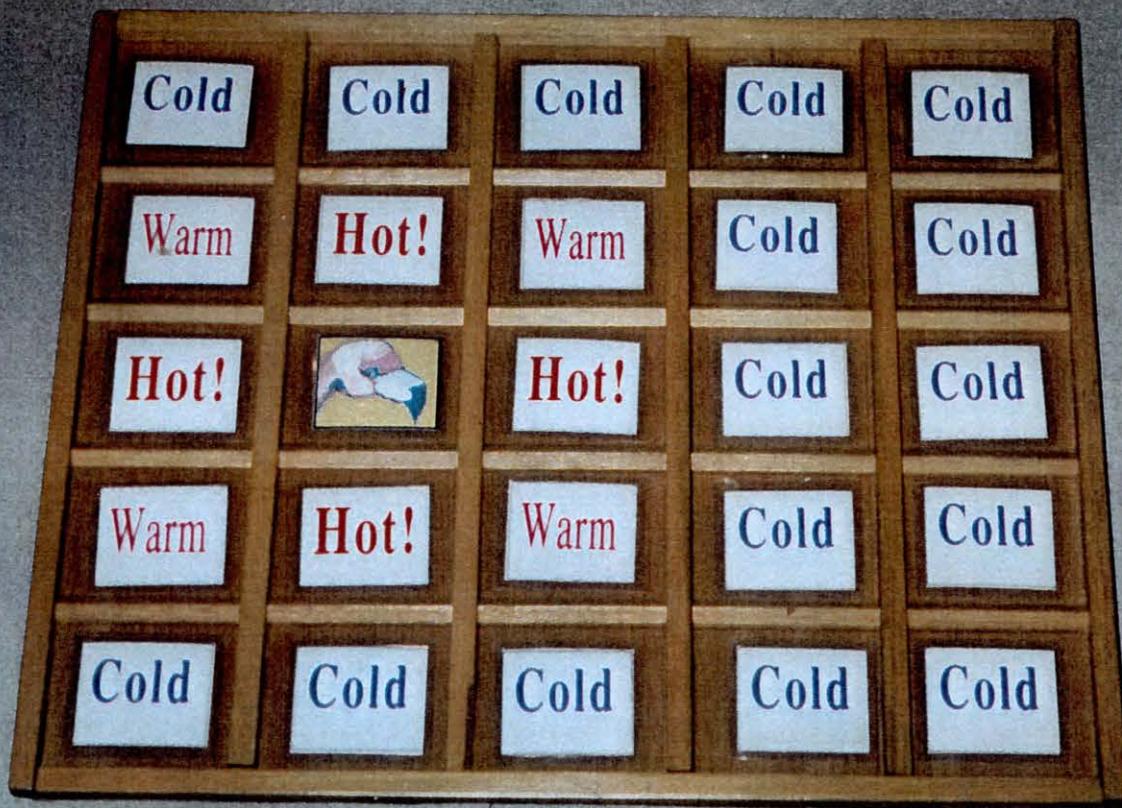
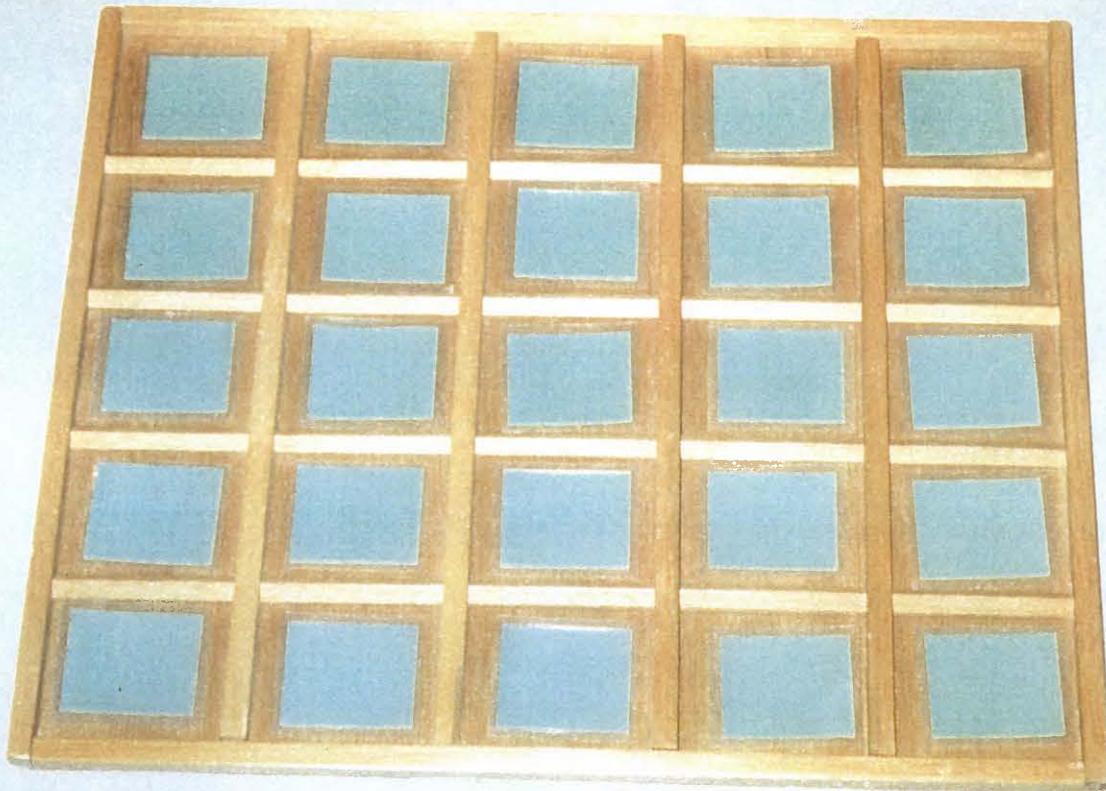


Figure 3.3. Pictures of the card game

3.3.4. Procedure

Children were taken individually to a quiet room in their school and shown the game. Each child was given an explanation: that she or he was going to play a game in which she or he had to find a Flamingo. The experimenter said:

"If you follow the rules, the job will be much easier. Could you read the rules to me?"

The screen that had rules was clicked for the computer group. A sheet on which the following rules was written was introduced to each child in the board and card group.

Turn over a card to find the Flamingo.
If it's there, you win.
If the Flamingo touches your box sidewise, you are HOT.
If cornerwise, you are WARM.
If they don't touch at all, you are COLD.

If the child was not a fluent reader, the experimenter helped by reading the instructions.

After reading the rules, the child was asked to explain them to the experimenter. If the child understood the rules properly, she or he started to play the game. If not, the experimenter explained the rules. Each child played the game four times. Each time, the Flamingo was hidden in any of 25 places. The other two media groups had the same procedure. In the board game, the piece that had been lifted were not put back. In the card game, the cards that had been uncovered were not put back, either.

After the fourth game, each child was asked three questions on each of five test sheets. The five test sheets were shown in a random sequence (Appendix 1). Each of three sheets contained a single clue while each of two sheets contained double clues. The single clues were COLD, HOT, and WARM. The double clues were HOT- WARM, and WARM-COLD. The single clue tests were to test whether the goal place was inferred based on each of the word clues (the game rules). The double clue tests were to test whether two clues (rules) were applied at the same time. An assumption was set that there were differences in making inferences between the game rules, because of differential cognitive demands. The game rules that define the position of the goal in relation to HOT and WARM were stated as positive sentences and specified possible places of the goal. The game rule about COLD was stated as a negative sentence and informs only about impossible places of the goal. On the other hand, tests on multiple clues required the child to apply two or more game rules at the same time. To meet the requirement, the child had to hold the areas of possible or impossible goal places and to find the common ground between the areas. Besides a clue(s), the test sheets also contained a picture of a safari that at the beginning of game on the computer and the board. The card group was given test sheets on which there was a 5x5 matrix, without the picture.

The rationale behind the questions enabled the child to judge the possible and impossible goal places. The first question was to test whether the child made an inference of the goal place from the information. The second question was to test whether the child made multiple inferences. The third question was to test whether the child made an impossible inference. The experimenter asked:

"Where can the Flamingo be?"

"Can it be anywhere else?"

"Or can it be here?" (pointing to an impossible place) (Figures 3.4 to 3.8)

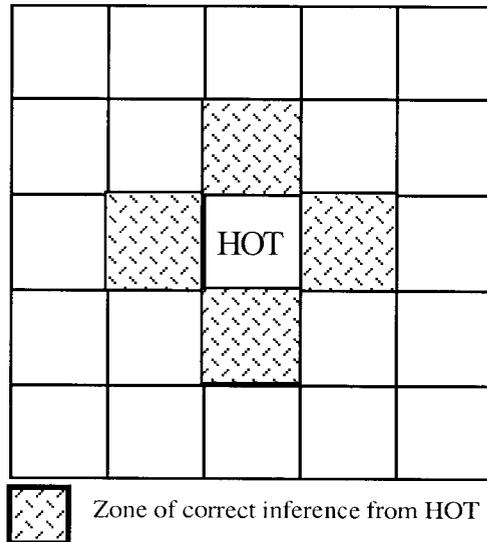


Figure 3.4. Correct inferences from single clue HOT

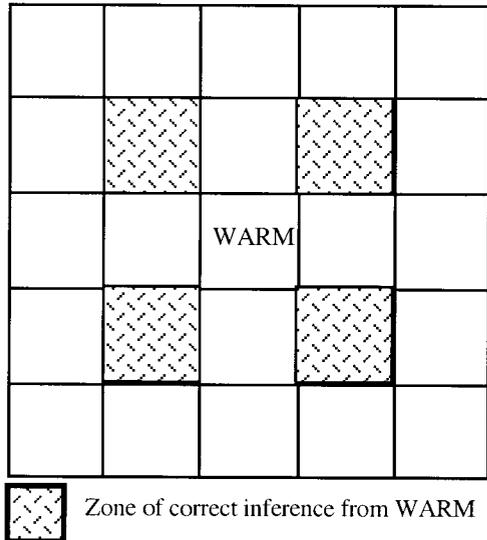


Figure 3.5. Correct inferences from single clue WARM

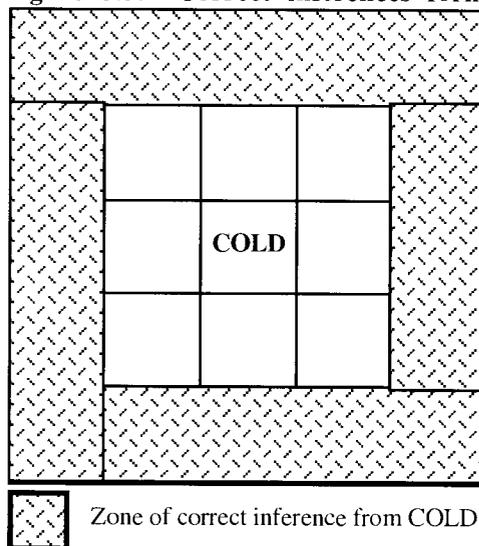
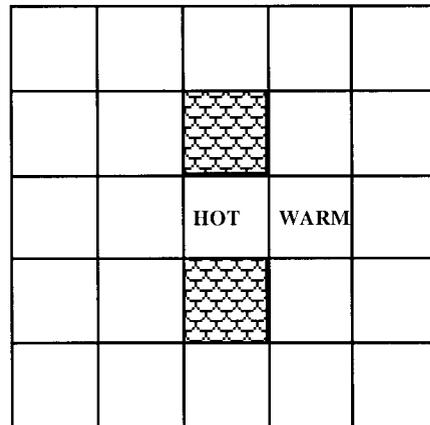
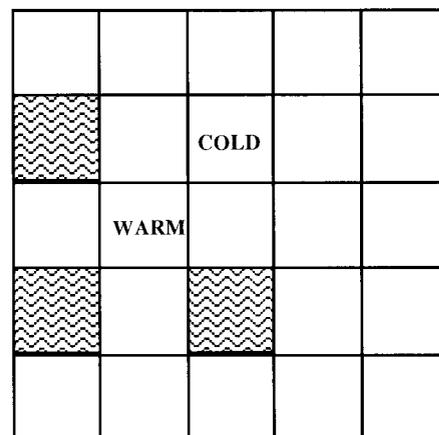


Figure 3.6. Correct inferences from single clue COLD



 Zone of correct inference from double information HOT & WARM

Figure 3.7. Correct inferences from double clues HOT-WARM



 Zone of correct inference from double information WARM & COLD

Figure 3.8. Correct inferences from double clues WARM-COLD

Correct answers to these three questions scored 1 point each. The range of scores for each sheet was between 0 and 3.

If the child was right, she or he was given positive feedback. If not, the next question was asked without feedback.

3.3.5. Measures

Two quantitative measurements were used. First, the number of moves needed to complete the game was counted. It represents the child's inferential problem-solving performance in the game. Second, scores for the paper tests were counted. They represent the child's ability in the test situation to infer from each word clue defined in the game rules.

3.4. Results

The results are explained in four sections. The first section describes the game performances. The criterion for success at chance level of performance is presented. Media and gender differences are analyzed. The second section describes performances in the word clue tests. Inferences from each clue - HOT, WARM and COLD - and of deduction of two clues are presented. In the third section, the reliability of the measurements is presented. The fourth section summarizes the findings of the study.

3.4.1. The game performances

The questions to be answered in this section were whether the children understood the game rules and used them in their choices. In order to answer the questions, the criterion for success at chance level was needed.

3.4.1.1. The criterion for success at chance level

In order to know whether the player made inferences from the game rules or just chose the places without any consideration of the game rules, perhaps at random, the mean number of moves of 4 games was observed. If the player made inferences from the game rules, it would not take as many moves as to find the target as if she or he found it by chance. How many moves would it take to find the Flamingo, if the player made

random moves? Three mathematical assumptions were made. Firstly, every place in the array has the same probability to be picked. Secondly, the number of places chosen until success follows a discrete uniform distribution over the integers from 1 to 25. Then the calculation for these follows. The mean of this distribution is $(25+1)/2$. The variance is calculated through deviance between the square of the mean and the summation of $1^2, 2^2, 3^2 - - -$ and 25^2 divided by the number of places. Thirdly, it happens that with 4 games, the mean is satisfactorily normally distributed. i.e. it follows the normal distribution with mean 13 and variance $52/4=13$. In order to set the criterion of chance number, one-tail test was carried out with the mean, SD and the number of games. This leads to the criterion of the chance number of 7.07 moves, with 95% confidence. If the mean number of moves over 4 games is 7.07 or fewer, I can reject the hypothesis that the player makes moves randomly. There must have been some bias, which is the child's use of the game rules. 36 children's mean number of moves was the same as or below the criterion, out of the total of 50, whereas the other 14 children made more than 7.07 moves. It can be said that the majority of the children did not make moves at random. They used inferences from the game rules in order to make subsequent moves.

3.4.1.2. Media and gender differences in the game performances

First, in order to find out the computer medium effect, performances in the computer game were compared with those on the board and card versions of the game. The mean number of moves of each media group was calculated. Second, in order to see whether there were gender differences, the mean numbers of moves for the girls and boys were

calculated. Table 3.2 presents mean numbers of moves across game media and gender groups.

The distribution of all the participants' mean numbers of moves is shown in Appendix 2. It was normal. In order to verify whether there were significant differences between the computer, the board and cards and between genders, I carried out a two-way between group (medium x gender) ANOVA test (see Table 3.3). There was no statistically significant difference between gender groups nor among the three game media groups, using the computer, the board and the cards. Gender and media effects will be explored again in the word clue tests.

Table 3.2. Mean numbers of moves across the three media and for each gender

Gender\Medium	Computer	Board	Cards	Total
Girls	7.83	6.50	5.58	6.59
Boys	5.96	6.58	6.71	6.40
Total	7.02	6.53	6.0	6.52

Table 3.3. Two-way ANOVA (medium x gender) for mean numbers of moves

Source of Variation	DF	SS	MS	F	Sig. of F
Medium	2	4.406	2.203	0.285	0.7534
Sex	1	0.552	0.552	0.071	0.7905
Medium X Sex	2	17.94	8.97	1.16	0.3228
Error	44	340.195	7.732		

3.4.2. Performances on the word clue tests

Two levels of analysis were carried out: total scores and scores for each of the tests. Firstly, each child's total score on the word clue tests was counted for the ability to make inference from the game rules. Analyses were carried out to see whether there were differences between those who played the game at random and those who used the rules, between the game media groups and between gender groups. Secondly, analyses of each of the tests were made. Analyses of the single word test scores were to assess the accuracy of inference making from each of the clues. Analyses on the double word test scores were to assess the systematic use of multiple possibilities or impossibilities.

3.4.2.1. Group differences in total scores

Total scores on the word clue tests were not normally distributed (see Appendix 3). The distribution skewed a little negatively. The possible maximum score was 15. The mean of total scores was 7.1 (S.D. 3.4). In order to know whether there was a difference in the inferential ability measured through the tests between those who made moves around chance level and those who made significantly fewer moves, total scores for the word clue tests were calculated for the two groups. The mean of the total scores was 7.94 (S.D. 3.41) for those whose mean number of moves was below the criterion; 4.93 (S.D. 4.93) for those who made moves within the criterion. In order to validate the difference between the two groups, a Mann-Whitney U test was carried out because data were not normally distributed. There was a statistically significant difference in

total scores between the children whose game performances were within the criterion of chance and those who were less than the criterion ($z = -2.895$, $p < 0.01$). Thus, there was an association between the criterion of chance and the total score for the tests.

In order to see whether there were game media and gender differences in the inferential ability, group means of the total scores for the word clue tests across media and gender were calculated. Table 3.4 presents the means of the total test scores across the three media and gender groups.

Table 3.4. Mean total scores for the word clue tests across the three media and gender groups

Gender\Medium	Computer	Board	Cards	Total
Girls	7.11	8.42	6.30	7.36
Boys	7.57	5.33	7.00	6.68
Total	7.31	7.39	6.56	7.10

To verify whether there were differences between the three game media groups, a Kruskal-Wallis test was carried out because data were not normally distributed. There was no statistically significant difference between the computer, board and cards groups ($H = .791$, $p > 0.05$). To verify whether there was a difference between gender groups, a Mann-Whitney U test was carried out because data were not normally distributed. There was no statistically significant difference between the boys and the girls ($z = -.48$, $p = 0.63$).

3.4.2.2. Differences between tests

To investigate how accurately the goal place was inferred from each of the word clues, the mean scores for each of the tests was calculated. The maximum possible score for each word clue was 3. To get a score of 3 at each test, the child was able to judge the possible and impossible goal places in relation to each clue. The mean scores for the single clues HOT, WARM and COLD were 1.98 (SD 0.89), 0.98 (SD 1.02) and 2.1 (SD 1.02), respectively. The mean scores for the double clue tests were 1.18 (SD 0.83) for HOT-WARM and 0.86 (SD 0.88) for WARM-COLD, with maximum possible score 3.

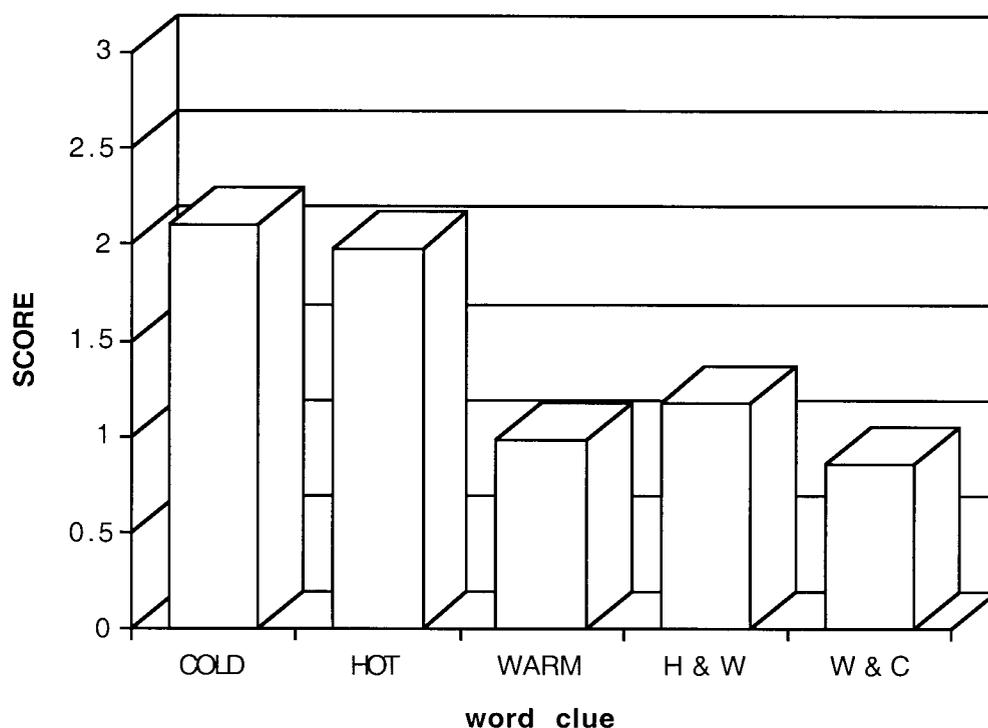


Figure 3.9. Mean scores across word clue item

To verify whether there were differences between test items, Wilcoxon tests were applied. There was no difference between the single clues COLD and HOT ($z = -.889$, $p > 0.05$) while there were significant differences between COLD and WARM ($z = -4.733$, $p < 0.01$) and between HOT and WARM ($z = -4.638$, $p < 0.01$). The scores for COLD and HOT were significantly higher than for WARM.

There was no significant difference in scores between the single clue WARM and the double clue HOT-WARM ($z = -.146$, $p > .05$) nor between the single word clue WARM and the double word clue WARM-COLD ($z = -.973$, $p > .05$). However, there was a significant difference between two double clues HOT-WARM and WARM-COLD ($z = -2.341$, $p < 0.05$). The scores for HOT-WARM were significantly higher than those for WARM-COLD.

3.4.3. Reliability and validity of measurements

In order to verify the reliability of the measurement of the number of moves for assessing children's inferential problem-solving performance, a comparison was made between the mean numbers of moves in the odd and even sets of games. The correlation between the two means was statistically significant ($r = .524, p < 0.01$).

In order to verify the validity of the measurement of the number of moves for assessing children's inferential problem-solving performance, a correlation was made between the number of moves and the scores on the word clue tests. The correlation between the mean number of moves and the total scores for the word clue tests was statistically significant ($r = -.542, p < 0.001$). Those who made a small number of moves in the game had high scores on the word clue tests, whereas those who made a large number of moves in the game had low scores.

In order to verify the validity of the criterion for chance level success as a test of rule use, a comparison was made in the word clue tests between those whose mean number of moves was below the criterion and those who made moves within the criterion. There was a significant difference between the two groups ($z = -2.895, p < 0.01$).

3.4.4. Summary of results

1. The majority of children made fewer moves than the criterion of chance level.
2. There was no significant difference in the mean number of moves between the computer, board and cards groups.
3. There was no significant difference in the mean number of moves between genders.
4. There was a significant difference in total scores for the word clue tests between those whose mean number of moves was below the criterion of chance level and those who made moves within the criterion.
5. There was no significant difference in the total scores for the word clue tests between the computer, board and card groups.
6. There was no significant difference in the total scores for the word clue tests between genders.
7. There were significant differences in scores between the tests. Scores for COLD and HOT were significantly higher than for WARM, while there was no difference between COLD and HOT. Scores for HOT-WARM were significantly higher than for WARM-COLD. In the comparison of the single and double clues, there was no significant difference between WARM and HOT-WARM nor between WARM and WARM-COLD.

8. The correlation between the number of moves and the total scores on the word tests was highly significant. The correlation between the mean numbers of even and odd sets of games was also significant.

3.5. Discussion and conclusions

Study 1 explored four aspects of children's inferential problem-solving in a computer game context: the representation of the game, the ability to make inferences from the game rules, the measurement of inferential problem-solving in the game, and the effect of the computer on the game play. Further research questions were drawn from the findings of the study.

Study 1 has shown that 8-year-old children on average can enact the game as a problem-solving task when the game is structured by the rules, stated as affirmative and negative *if -then* sentences. The majority of the children found the Flamingo with a smaller number of moves than the criterion of the chance number expected with random choices. These children inferred the goal place from the game rules when they made moves. It was evident that they represented the game as a problem-solving task.

Counting the number of moves until the completion of the game is a reliable measurement of the child's inferential problem solving in the game process. There was a high correlation in the mean numbers of moves between the even and odd sets of games. The mean numbers of moves for all four games also significantly correlated with the total score for the word clue tests measuring the ability to infer from the HOT, WARM and COLD clues that appeared in the game and the game rules. The results of each of the word clue tests give hints on how the player might have dealt with each clue

in the actual game process. Scores for the clue WARM were significantly lower than those for HOT and COLD. Scores for the clue of COLD were as high as those for HOT in the word clue tests. These results contrasted with previous findings of the effects of negative constituents of statements on information processing (Fabricius, et al., 1987; Toppino, 1980; Wason, 1959). The negative statement "If *they don't touch at all, you are COLD*" was inferred from as accurately as the affirmative statement that "*If the Flamingo touches your box sidewise, you are HOT*". The negative statement was also inferred from significantly more correctly than the affirmative statement that "*If the Flamingo touches your box cornerwise, you are WARM*". There are two possible explanations for this contrast. First, COLD was inferred from as accurately as HOT, because the children in this study were old enough to handle the cognitive demand of a negative constituent as a sign of exclusion of the adjacent places. Second, it might be that the nature of the information has a crucial effect on inference making, stronger than the form of the statements in which information was given. If this is the case, the question is how the nature of WARM differs from that of HOT or COLD. It will be interesting to see whether the findings of the tests are also the cases of the game process. Would COLD be used correctly as often as HOT? Would WARM be less used than HOT or COLD? The way that children make moves in relation to the clues in the game process will be investigated in the next study. The investigation on use of the clues in the game process will inform further about the strategies that the children use. For example, making moves beyond the chance criterion requires a strategy to maximize the elimination of the impossible places for the goal. Do children exclude all the impossible places indicated by COLD in the game process?

No difference was found between gender groups in the game performance. This result may come from the age of the participants. Earlier research (Halpern, 1986; Kerns & Berenbaum, 1991) showed that gender differences in spatial abilities may appear around 10 years of age in favour of boys. The 8-year-old children in this study might still be too young to have developed differences in spatial abilities.

A further interesting finding was that the media on which children played the game had no significant effects on children's problem solving. The game performances of the computer group did not differ from those of the board and cards groups. In spite of the presence of the vivid visual display and audible feedback, computers did not differ from the traditional game tools in the children's problem solving.

Another research suggestion deriving from this study is to investigate the effect of the media and gender differences with children of a broad age range. It is difficult to generalize how far the computer and gender do not affect children's problem solving in search games. I need evidence about whether these findings are the same for younger and older children as well.

A further research question concerns individual difference and learning. A significant difference was shown in the results of the word clue tests between those who made moves below the chance criterion and those who were within the criterion. Did children start the game with understanding of the problem of the game? Would they learn it in the course of the game? Even though literature on problem-solving has its tradition in the study of individual difference (e.g., Bruner, 1966, 1971), this is an area largely ignored in computer related research. For research on individual difference and learning, more games were allowed to each player in further studies.

Chapter 4. Study 2

4.1. Aims

The aim of Study 2 was primarily to investigate development, individual differences and learning in children's inferential game play. Secondly, it aimed to replicate some of the findings of Study 1. Gender and game media differences were investigated extensively, using two more age groups of children.

4.2. Introduction

This study aimed to replicate and extend some of the findings of Study 1, and to indicate some suggestions for further studies. Replication was needed to check that the findings on game media and gender differences held across two more age groups of children, whose data would eventually allow analysis of developmental changes. As discussed in Study 1, two research questions are considered: whether learning occurs while playing the game, and how inferential game behaviour differs from individual to individual.

I shall assume, within inferential game behaviour, two aspects of cognition which depend on developmental changes. The first is the development of the representation of the problem (presumably structured by the game rules). For example, the structure of

the Flamingo game, and consequently the word clues in the game, are related to higher-order concepts such as possibility or impossibility. A COLD clue conveys information of the impossibility of a certain area for the goal and of the possibility of the rest, whereas a HOT or WARM clue indicates the possibility of certain places. Operations of inclusion and exclusion render the appropriate representations of the zone of impossible places for the goal and the zone of possible ones. From the Piagetian perspective, operation of inclusion and exclusion must develop after a certain age (Piaget & Garcia, 1991). Only children aged 10 years upwards are expected to be able to construe the words as signals for an exclusion of the impossible places of the goal and an inclusion of the possible. Then, they would know which choices are appropriate.

Secondly, as children's development progresses to the systematic use of multiple (more than one) possibility, there will be a change in inferential game behaviour. If the players define a zone of possibility and use the possible places within the zone systematically, their performance will be enhanced. Bryant and Roazzi (1992) studied 6-, 7-, and 8-year-old children in number inference tasks which asked for a number which is bigger than X but smaller than Y. The children were asked to choose a number in one case where there was only one possible correct number and in another case where there were several correct answers. In the task which had one correct answer, even the younger children were quite good at inferring a "critical area", particularly in their first choices. In the other task, the proportion of correct first choices that were immediately followed by a correct second choice was quite high in the seven and eight year olds but a great deal smaller in the six year olds. It seemed that the six year olds had difficulties in holding the "critical area" in their mind and eliminating alternatives one by one. The findings of Bryant and Roazzi's study show that systematic use of inference in numbers develops up to 7 years of age. The Flamingo game requires systematic use of inferences in space. It is uncertain at this point whether inference in space develops after that for numbers.

While the assumption of developmental change in relation to the Flamingo game takes the constructivists' view that the structure of the game is consciously represented and develops with age, the probability-matching hypothesis envisages that the choices of the player are guided by the likelihood gained through accumulated reinforcements. No developmental change is expected in this model. Children would not tackle the task of getting the target logically. They would tend to attend to a HOT or WARM because these words were more likely to be found next to the goal than COLD in the previous games.

To understand what lies underneath choice-making, we need to concentrate on the uses of information. If the players make inference from the game rules, their choices would be coherent with inferences from the information they have. Faced with HOT, their choice will be one of the adjacent places that are above, or below, or to the left or right. Faced with WARM, they will move to one of the diagonally adjacent places. Faced with COLD, the next move will be to one of the non-adjacent places. On the other hand, if the players do not have any concept of the game rules, their choices would not correspond to what the game rules define as the relations between the goal and the information. Thus, the focus of analyses of the game behaviour should be on choices (moves) in relation to information. The study adopted the methodology of recognizing choice (response) patterns in relation to reinforcement of concept formation tasks (Levinson & Reese, 1967; Kendler & Kendler, 1970) for the examination of children's use of inference from information. Each move can be seen as a result of either a correct or an incorrect inference from the previous information. To assess whether a move follows from a correct inference, two analysis methods were applied. The first method will be called 'single information inference', in which the immediately preceding information counts as the base of the inference. The other method will be called 'double information inference', in which the two preceding items of information count as the base for inference. These procedures of data analysis allow us to look at the way

in which children integrate spatial relationship information, and to identify which concepts and operations they find demanding.

The next question is whether children become more proficient in using a particular form of logical reasoning when it is embedded in a game that they can play over and over again. In this study, unlike some other tasks, children became engaged in a cultural practice, a game, which assumes either the use of inference from the rules or the accumulation of reinforcements. The study investigates whether children continue to perform at the same level of proficiency after many games, or whether they come to use inferences more and more. Learning curves reveal how performances change over games. Studying subgroups' learning curves over games and their choice-patterns were two methods for the examination of individual differences used in this second study. Two kinds of players were envisaged. The first group was those who played the game with an assumption that the game is a chance event. Their performances were expected to be at chance level. Choices of this group would not correspond to the game rule. The second group was those who made inferences from the game rules. They were expected to perform well. Their choices would be coherent with the game rules.

The secondary goal of the study was to replicate some findings from Study 1, which showed no difference between genders nor between the game media groups. This study tackles the questions again, with younger and older children than the participants of Study 1. For generalization of the findings of Study 1, it was necessary to test whether or not the computer as a game tool affected the game behaviour of children of various ages. The study also aimed to assess the ability to make inferences from word clues, again through word clue tests.

To sum up, the study primarily explored three aspects of the game play.

- developmental change between 7-year-old children and 10-year-old children;

- individual differences: whether some children are consistently better than others in inferential game play;
 - learning: whether changes occur over games.
-
- Furthermore, the study aimed to replicate other findings of Study 1, concerning:
 - game medium difference between use of a computer and a board;
 - gender difference;
 - reliability of measurements: correlation between the mean number of moves of the odd and even sets of the games;
 - validity of using the number of moves for measuring the ability to make inferences, comparing this with the total word clue tests score;
 - validity of using the criterion of chance number for the test compared to the total word clue test score for the use of the game rules.

4.3. Method

4.3.1. Design

The study had a mixed between subjects and within subjects design. It aimed to answer the research questions in these four ways:

1. by comparing performance between the two age groups, gender groups, and the computer and board groups,
2. by examining the relationships between the game performance and the word clue test performance,
3. by tracing the performance of the ‘good’ and ‘poor’ groups from the first to the last games, and
4. by comparing moves after HOT, WARM, and COLD.

The first method was used to see whether there were developmental changes, gender differences and medium effects in inferential problem solving. Significant differences between the age groups in favour of the older group would suggest that inferential ability develops with age.

The second method was used to replicate the findings of Study 1, that there was a significant relationship between the children’s game performance and the test performance. A significant positive relation between the game and the test would suggest the validity of the measurement of inferential problem solving in the game situation.

The third method was used to see whether the children came to the game with already developed logical skills, or whether they developed their strategies while playing.

Using the fourth method, the children's use of inference from HOT, WARM, and COLD in the game process would be examined. It was predicted that COLD would be used more than HOT and HOT used more than WARM, as the word clue tests in Study 1 had shown.

4.3.2. Participants

The study involved 87 children whose ages ranged from 6 years 10 months to 10 years 10 months. Two age groups were represented: 32 children aged from 6 years 6 months to 8 years 1 month (mean age = 88.3 months; SD 4.01) and 55 children aged 9 years 10 months to 10 years 10 months (mean age = 124 months; SD 3.45). The children were all those in years 2 and 5 in two schools in a middle class area in central London.

Table 4.1. Number of participants across age, gender and medium

	7 years		10 years		Total
	Computer	Board	Computer	Board	
Girls	9	7	11	18	41
Boys	7	9	14	12	46
Total	32		55		87

4.3.3. Apparatus

The computer and the board game media presented children with the same game structure. Each game apparatus had a 5x5 grid to allow 25 items of information to be shown. There is a distinguishing feature between the two media. The computer had a review key that allowed a view of all the moves made in the game. As the player clicked a place on the 5x5 grid, a word appeared in the place for a couple of seconds and returned to the original picture. With the board, the player made a choice on the 5x5 grid, lifted the top piece in the place and uncovered it until the completion of the game.

4.3.4. Procedure

The computer and board groups had the same procedure as described in Study 1, except that each child had 8 trials of the game, and that there were 6 word clue tests. There were three tests with a single clue and another three tests with a multiple clue. The three tests with a single clue, HOT, WARM and COLD, measured the ability to infer the goal from each of the clues. The three tests with a multiple clue measured the ability to infer from two or more clues at the same time.

4.3.4. Outcome measures

Five quantitative measurements were used:

- (a) Number of moves until the completion of the game; this was to measure children's inferential problem-solving ability in the game.
- (b) Single information inference: this was to identify whether the move was an adequate result of inference from the previous information (see examples in Figure 4.1 to 4.3).
- (c) Double information inference: this was to identify whether the move was the result of inferences from two successive moves (see examples in Figure 4.4 to 4.9).
- (d) Time taken until the completion of the game; this was to measure children's proficiency in information handling.
- (e) Scores for the word clue test; this was to measure children's ability to infer from word clues, and a test of the validity of (a).

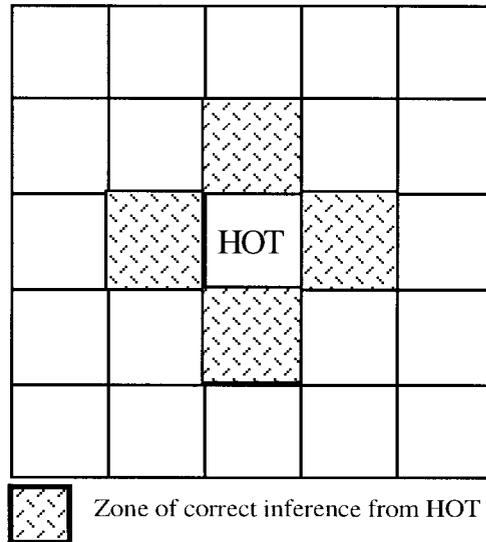


Figure 4.1. Correct inferences from HOT

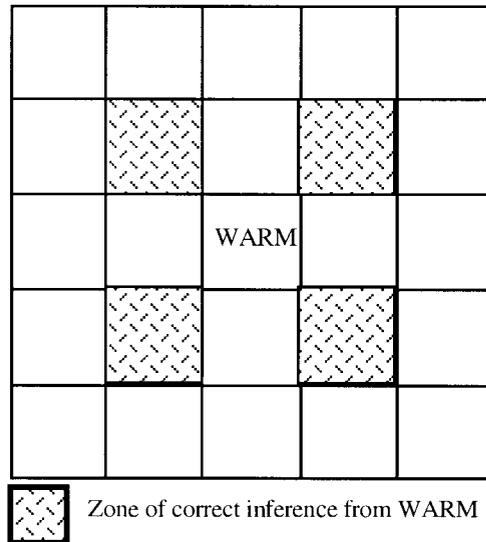


Figure 4.2. Correct inferences from WARM

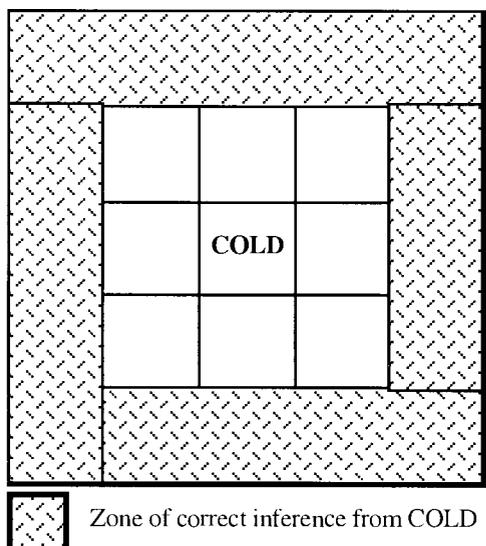
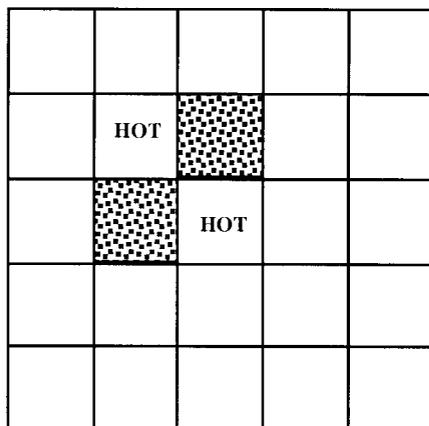
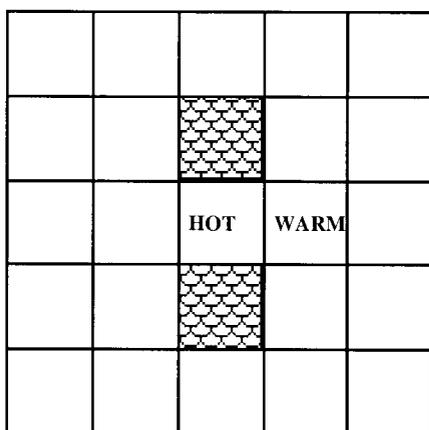


Figure 4.3. Correct inferences from COLD



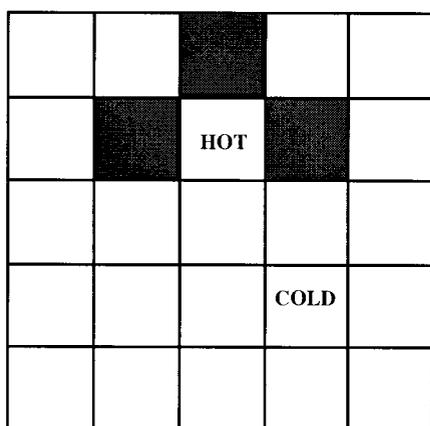
▣ Zone of correct inference from double information
HOT & HOT

Figure 4.4. Correct inferences from two HOTS



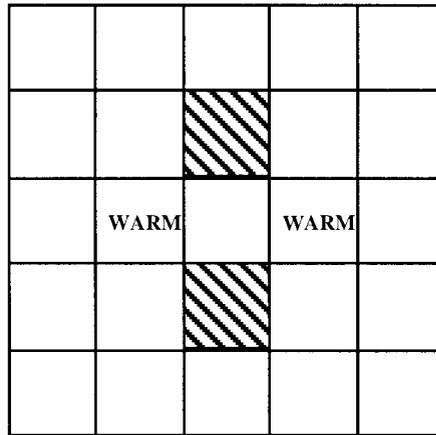
▣ Zone of correct inference from double information HOT
& WARM

Figure 4.5. Correct inferences from a HOT and a WARM



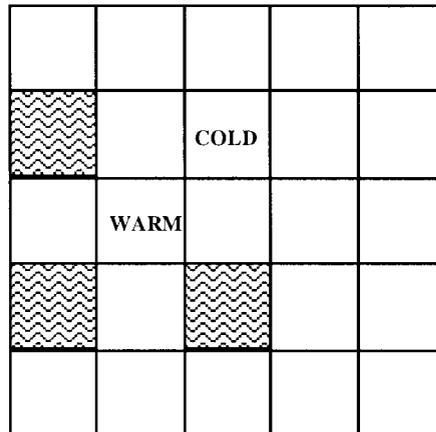
■ Zone of correct inference from double information HOT
& COLD

Figure 4.6. Correct inferences from a HOT and a COLD



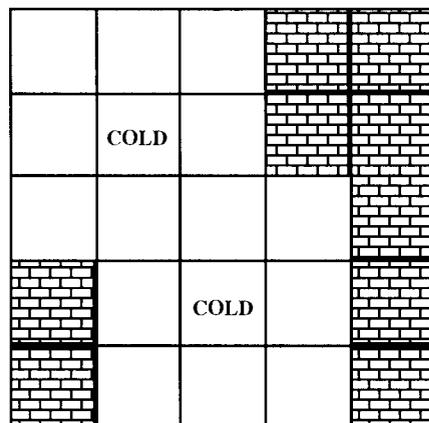
▣ Zone of correct inference from double information
WARM & WARM

Figure 4.7. Correct inferences from two WARMs



▣ Zone of correct inference from double information
WARM & COLD

Figure 4.8. Correct inferences from a WARM and a COLD



▣ Zone of correct inference from double information
COLD & COLD

Figure 4.9. Correct inferences from two COLDs

4.4. Results

The results are given in six sections. The first three sections describe the replication of Study 1: analyses of the number of moves and of word clue test scores, and reliability and validity of the measurements. The next two sections address individual differences and learning over games. The last section describes choice-patterns.

4.4.1. Analysis of number of moves

The questions to be answered in this section were whether the children understood the game rules and used them in their choices. In order to answer the questions, the criterion for success at chance level was needed.

4.4.1.1. The criterion of chance level

In order to know whether the player made inferences from the game rules or just chose the places randomly, the criterion for success by chance was set. It was assumed that, if the player made inferences from the game rules, it would not take as many moves when they were made by chance to find the target. With 8 games, the mean was 13 and variance was $52/8=6.5$ (its calculation was already described in Study 1). A one-tail

test was carried out with the mean, the variance and the number of games. This led to the chance level of 7.07 moves with 99% confidence, or 8.80 moves with 95% confidence. I take 7.07 with 99% of confidence over 8 games, for this number gives the same value as the criterion used in Study 1 over 4 games.

Only when the child's mean number of moves in 8 games was the same as or lower than the criterion, the hypothesis that the child made moves at random was rejected. Each player's mean number of moves in 8 games was counted. 61 children out of 88 children made, on average, fewer moves than the chance criterion. Therefore, the majority of the participants inferred the game rules to find the goal. In order to compare rule use between the age groups, the number of participants who made fewer moves than the criterion of 7.07 was counted for each age group. For 13 out of 32 7-year-old children, the mean number of moves in 8 games was smaller than the chance criterion; while 48 10-year-old children out of 55 made fewer moves than the criterion. There was a significant difference between the age groups in the number of children whose mean number of moves was less than the chance criterion (Chi-square = 4.07, d.f. = 1, $p < 0.05$). In the next section, I will describe differences between genders and between media groups, using the mean number of moves, which were also used to examine differences between these groups in Study 1.

4.4.1.2. Media and gender differences in number of moves

In order to see whether there were no differences between the media groups and between the gender groups, the mean of number of moves was calculated across each

group. Table 4.2 shows the mean number of moves across the age, media and gender groups.

The distribution of the mean number of moves of all the participants was skewed positively (see Appendix 3). In order to verify whether there were differences between the game medium groups or between gender groups, Mann-Whitney tests were applied separately with gender variables or with the media variables because data were not normally distributed. There was no significant difference between the computer and board groups ($z = -.77$, $p > 0.05$) nor between gender groups ($z = -1.27$, $p > 0.05$). The finding of no significant difference between the media groups or the gender groups replicates the finding of Study 1.

Table 4.2. Mean number of moves (SD) across age, media, and gender groups

	7 years		10 years		Total
	Computer n=16	Board n=16	Computer n=25	Board n=30	n=87
Girls	10.55 (4.70)	10.92 (4.70)	6.21 (2.91)	6.00 (3.45)	8.08 (4.35)
Boys	8.29 (4.52)	9.86 (4.40)	5.59 (2.16)	5.03 (2.24)	6.44 (3.52)
Total	9.87 (4.49)		5.58 (2.58)		7.16 (3.97)

4.4.2. Analysis of word clue test scores

Analysis of scores for the word clue tests was carried out at two levels. Firstly, the child's total scores for the word clue tests were counted for the ability to make inference from the word clues. Analysis was carried out to see whether there were differences between the two criterion groups, between the game media groups, and between gender groups. Secondly, analysis of each of the tests was made. Analysis of the single word test scores was to assess the accuracy of inference making from each of the clues. Analyses of the double word test scores were to assess the systematic use of multiple possibility or impossibility.

4.4.2.1. Group differences in total scores

Each child's total score for the word clue tests represents the ability to infer from the rules. The distribution of the total scores was skewed positively (see Appendix 4). The mean score for all the participants was 11.36 (SD 5.04) with the maximum possible score of 18. The mean score was 13.65 (SD 4.01) for those who made fewer moves than the chance criterion; and 6.26 (SD 2.88) for those who made moves above the criterion. In order to verify whether there were significant differences between the two groups, Mann-Whitney U tests were carried, because data were not normally distributed. There was a significant difference between the criterion groups ($z = -6.50$, $p < 0.01$).

In order to see whether there were differences between the age, media and gender groups, the mean scores for each of the groups were calculated. Table 4.3 shows the mean scores across the age, media and gender groups. In order to verify whether there were significant differences between the age, gender or media groups, Mann-Whitney tests were carried with each variable one at a time, because data were not normally distributed. There were significant differences between the age groups ($z = -5.603$, $p < 0.001$). No difference was found between genders ($z = -1.647$, $p > 0.05$) nor between media ($z = -.582$, $p > 0.05$). This study thus replicated the findings of Study 1, that there was no difference in total scores for word clue tests between genders ($z = -.48$, $p = 0.63$) nor between media groups ($H=.791$, $p > 0.05$).

Table 4.3. Mean scores (SD) across age, media, and gender groups

	7 years		10 years		Total
	Computer n=16	Board n=16	Computer n=25	Board n=30	n=87
Girls	7.00 (1.41)	7.89 (3.37)	11.79 (4.73)	13.63 (6.12)	10.37(4.94)
Boys	8.33 (4.39)	6.43 (3.15)	14.82 (3.68)	14.14 (4.02)	12.12 (5.03)
Total	7.5 (3.28)		13.6 (4.5)		11.36 (5.04)

4.4.2.2. Differences between the tests

In order to know how each clue was used to infer the goal place and whether there was systematic use of multiple possibility or impossibility, the mean scores for each of the tests were counted. The maximum possible score for each word clue was 3. To get the maximum score of 3 in each of the tests, the child should have been able to judge the possible and impossible goal places, inferring from the game rules. Mean scores for the single words HOT, WARM and COLD were 2.16 (SD 0.91), 1.60 (SD 1.20) and 2.56 (SD 0.94), respectively. Mean scores for double words were 1.69 (SD 0.99) for the item HOT-WARM and 1.47 (SD 1.15) for the item WARM-COLD, with maximum possible score 3. The mean score for the triple words was 1.86 (SD 0.94) with maximum possible score 3. Figure 4.10 shows scores for each of the tests.

To verify whether there were differences between the tests, Wilcoxon tests were applied, because data were not normally distributed. There were significant differences between the single word COLD and HOT ($z = -3.971$, $p < 0.01$), between COLD and WARM ($z = -5.636$, $p < 0.01$) and between HOT and WARM ($z = -5.093$, $p < 0.01$). Scores for COLD were significantly higher than for HOT, and scores for HOT were significantly higher than for WARM.

There was a significant difference between the two double clues HOT-WARM and WARM-COLD ($z = -2.80$, $p < 0.05$). The scores of HOT-WARM were significantly higher than those for WARM-COLD. There were significant differences between the triple clue and the double clues. Significant differences were found between HOT-WARM-COLD and HOT-WARM ($z = -2.056$, $p < 0.05$), and between HOT-WARM-COLD and WARM-COLD ($z = -3.483$, $p < 0.001$).

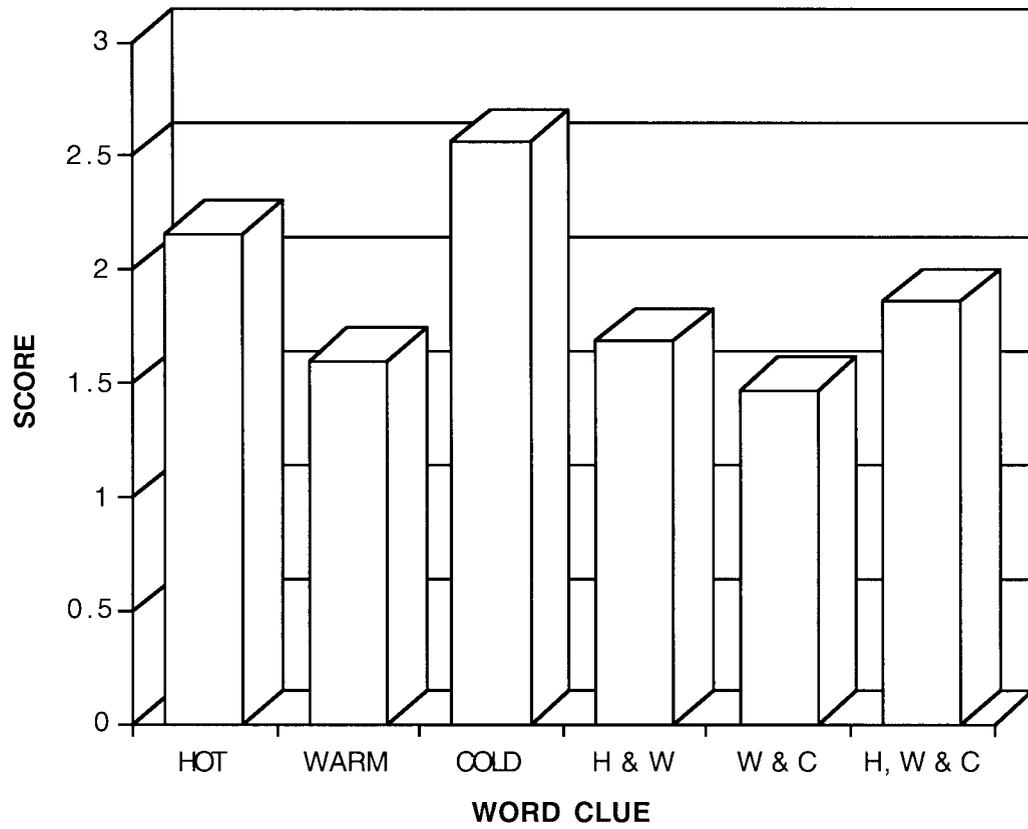


Figure 4.10. Scores for each of the word clue tests

To validate the findings of the word clue tests in Study 1, a comparison was made between Study 1 and 2. In Study 1, COLD was scored the highest. COLD and HOT were scored significantly higher than WARM, while there was no significant difference between COLD and HOT. In this second study, COLD was scored significantly higher than HOT, and HOT was scored significantly higher than WARM. The order of the mean scores for each single word test was identical in both studies. The results of the word clue tests in this study confirmed the levels of difficulties of the single word clues detected in the previous study. The children judged the possible and impossible goal places with differential levels of accuracy across the clues. COLD (the clue of exclusion of impossible places) was used more easily to make inferences than HOT and WARM, which specified a maximum of four possible goal places. These results give more doubt because the cognitive demand of inference making from COLD seems heavier than from HOT or WARM.

4.4.2.3. Age Differences in the word tests

To find out which word clue results had age differences, the mean score for each test was calculated for both age groups. Data were classified only into age groups, because there was no difference in scores on word clue tests between gender or between media groups. Figure 4.11 shows mean scores for word clue tests across age groups.

In order to verify whether there were differences in the scores for each test item between age groups, Mann-Whitney tests were carried out with scores for each item, one at a time because data were not normally distributed. Significant differences were found between age groups for each of the tests: HOT ($z = -4.621$, $p < 0.01$), WARM ($z = -5.18$, $p < 0.01$), COLD ($z = -3.085$, $p < 0.01$), HOT-WARM ($z = -3.895$, $p < 0.01$), WARM-COLD ($z = -4.344$, $p < 0.01$) and HOT-WARM-COLD ($z = -3.547$, $p < 0.01$). The 10-year-old children scored significantly higher than the 7-year-old children in practically every word test. The older group made more accurate judgements, based on the clues about possible and impossible goal places, than the younger group.

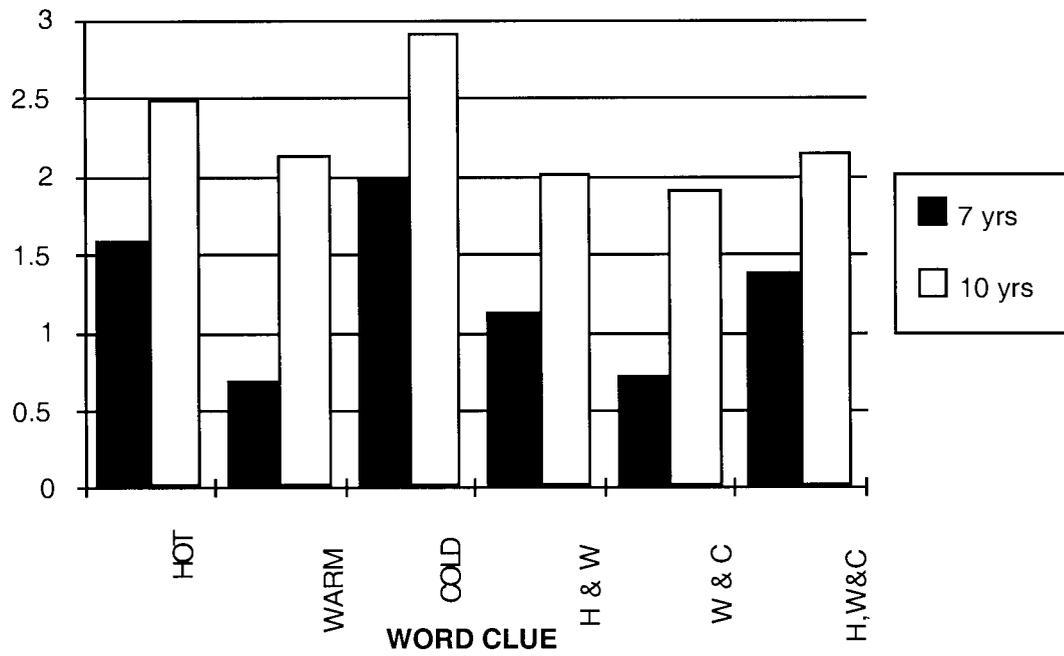


Figure 4.11. Scores for word clue tests across age groups

4.4.3. Reliability of measurements

In order to replicate the findings of Study 1, tests for the reliability and validity of the measurements were applied. To assess the reliability of measurement by counting moves for assessing the child's problem-solving performance, the correlation was calculated between the mean numbers of moves in odd and even sets of games. The correlation between the two means was statistically significant ($r = .634, p < 0.001$).

The correlation between the mean number of moves over 8 games and the total score for the word clue tests was calculated, as a way to test the validity of measurement by counting moves for assessing the child's problem-solving performance. The correlation between the mean number of moves and the score on the word clue tests was statistically significant ($r = -.723, p < 0.001$). Those who had a small number of moves in the game had high scores for the word clue tests, while those who had a large number of moves in the game had low scores.

To validate the criterion for chance level success as a test for the child's use of the game rules, the total scores for the word clue tests were compared between the two criterion groups. There was a significant difference in the test scores between the groups ($z = -6.50, p < 0.01$).

4.4.4. Individual differences

In order to see whether some children were consistently better than others, and whether the difference existed from the beginning of the game, the study used the method of backward learning curves. The method takes final responses for the selection of groups and traces responses from the first to the last games (Zeaman & House, 1963). The children were divided into 2 groups - good and poor - based on the mean number of moves in the last 4 games. Children whose mean number of moves in the last 4 games was smaller than 7.07, the criterion of chance level success over 4 games, were defined as the good performance group. Those who had a mean number of moves around the criterion were placed in the poor performance group. The mean number of moves and the mean scores for the word clue tests of each of the groups were calculated for the comparison between the groups. Table 4.4 shows the number of participants, the mean number of moves and the mean test score of each performance group.

The performances of each group were traced from the first trial to the last. Figure 4.12 shows the learning curves of each of the performance groups. In both age groups, the poor game players who performed around the criterion of success by chance level in the later trials were those who started the game badly. On the other hand, the good performers did better from the start. The patterns of responses found in the final games already existed from the beginning. The 10-year-old 'good' performance group made fewer moves than the criterion of chance level from the beginning. The 7-year-old 'good' performance group made fewer moves than the criterion after 4 games. The group needed the first four games to learn the use of the game rules in their choices. The 10-year-old 'poor' performance group was in overall worse than the 7-year-old 'good' performance group. The fluctuation of the learning curve of the 10-year-old

'poor' performance group might have come from the small group size. The 7-year-old 'poor' performance group did not show any improvement within the eight games.

Table 4.4. Number of participants, mean number of moves and mean test scores for each performance group

Age	Performance	No. participants	Mean No.(SD)	MeanScore (SD)
7 years	Good	13	6.2 (2.48)	9.4 (3.2)
	Poor	19	12.4 (3.78)	6.2 (2.78)
10 years	Good	48	4.7 (0.96)	14.5 (3.74)
	Poor	7	11.4 (2.84)	7.1 (4.18)

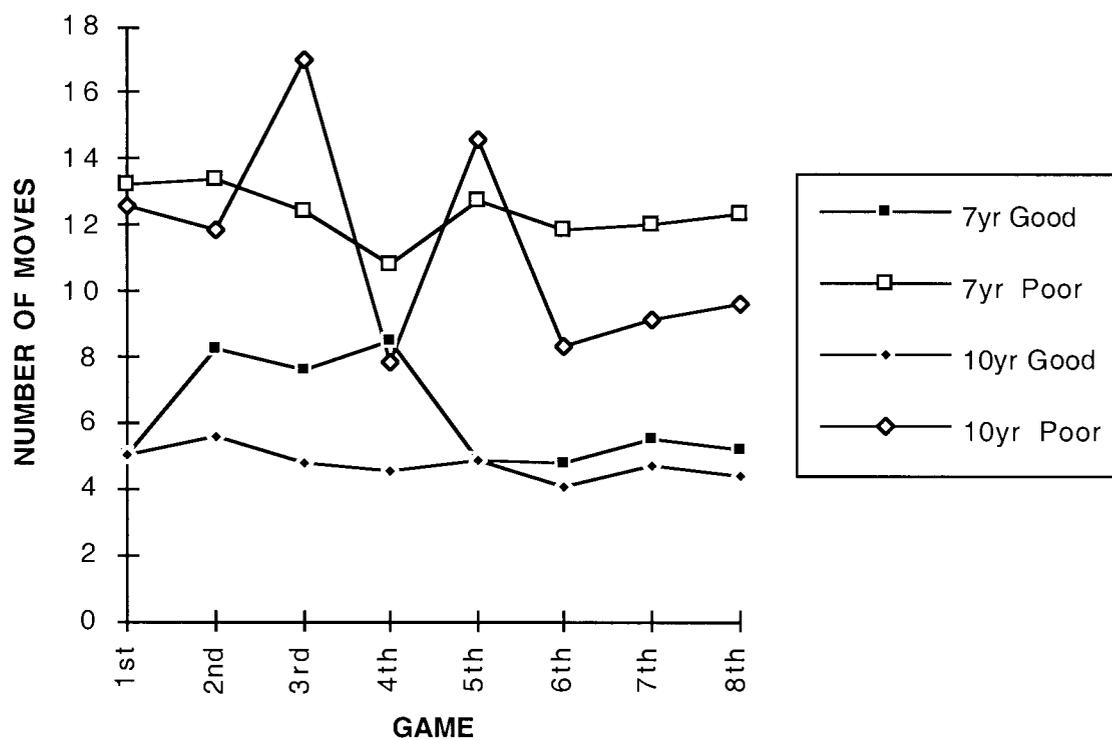


Figure 4.12. Number of moves for the performance groups across games

4.4.5. Learning measured by the completion time and number of moves

In order to see whether the use of the game rules required less time for thinking as the games went on, the completion time for each game was measured. Figure 5.12 shows the mean completion time for the 10-year-old children. These children got much quicker as the game went on. The performance time of the last game was less than half that of the first one. As a turning point, differences from before and after the 5th game were very clear and particularly interesting. Before the 5th game, the decline curve of the performance time was rapid. Children were getting closer to their minimum time.

In order to verify whether there were significant differences before and after the 5th game, games were divided into two blocks (1st – 4th and 5-8th). A Wilcoxon signed-rank test was used, because the distributions of times were non-parametric (see Appendices 6 & 7). There was a significant difference in the time taken to complete the game between the two blocks of games ($z = -2.51$, $p < 0.05$). The phenomenon could be interpreted as that information processing was getting skilled. It has been shown in previous studies that problem solving becomes smooth as practice on the same problem continues (e.g., Case, 1986).

In order to see whether there were significant differences in the numbers of moves before and after the 5th game, games were divided into two blocks (1st – 4th and 5-8th). A Wilcoxon signed-rank test was carried out because the distributions of the mean numbers of moves of the two blocks of games were non-parametric (see Appendices 8 & 9). There was a significant difference in the mean numbers of moves between the two blocks of games ($z = -2.32$, $p < 0.05$).

There were many missing data for the performance time of the 7 year-old-children, due to technical difficulties, so analysis of their data was not carried out.

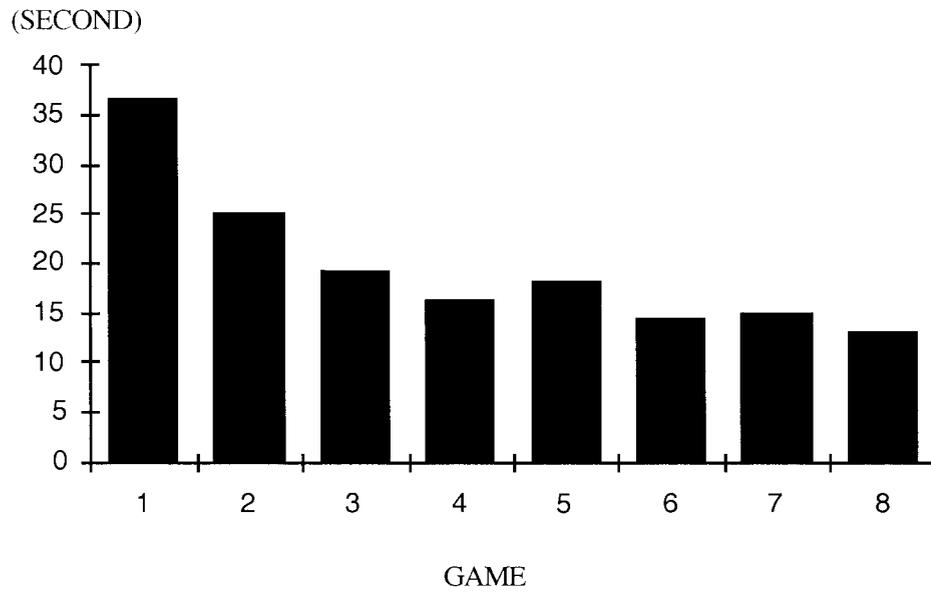


Figure 4.12. Mean time across games of 10 year-old subjects

4.4.6. Analyses of choice pattern

To assess whether each move (choice) followed from a correct inference from information, 'single information inferences' and 'double information inferences' were counted. In single information inferences, each move was judged as correct or incorrect in relation to the immediately preceding information (see Figures 4.1 to 4.3). In double information inference, each move was judged as correct or incorrect in relation to the two preceding items of information (see Figures 4.4. to 4.9).

For single information inferences, each move was categorized according to three areas that were identical to the three relations of the game rules; the horizontally and vertically adjacent area (H/V adjacent area), the diagonally adjacent area, and the non-adjacent area. The numbers of occurrences of each type of information (HOT, WARM and COLD) and the numbers of correct inferences from the information were calculated over 8 games for each of the participants. Because the occurrences of the three types of information varied across each participant, the percentages of correct inferences for the three information types were calculated. The mean percentage of single information inference was 89.1 (SD 15.3) for HOT, 51.0 (SD 35.0) for WARM and 64.9 (SD 26.8) for COLD. Table 4.5 presents the percent of moves into the three areas across the information.

Faced with HOT, most of the moves were made to the horizontally or vertically adjacent places. The children were very good at using inferences from HOT in the process of the game. Faced with WARM, most of the moves were made either to the H/V adjacent places or to the diagonally adjacent places. But moves to non-adjacent places were rare with WARM. It seems that the children coded WARM as referring to adjacent places. Faced with COLD, the majority of moves was made to the non-adjacent places

Table 4.5. Percentages of moves across three areas from single information

	H/V*	Diagonal	No-A~
HOT	89.1	5.9	5.1
WARM	39.1	51.0	9.7
COLD	25.9	9.2	64.9

 Percent of correct Inferences

* H/V ; Horizontally or Vertically adjacent area
 ~ No-A ; Non-adjacent area

The distributions of the percentages of correct single information inferences from HOT, WARM or COLD were not normal (see Appendices 10 to 12). In order to test whether there were significant differences in correct single information inferences between HOT, WARM and COLD, Wilcoxon matched-pair tests were carried out with pairs of HOT and COLD, and of COLD and WARM. There were significant differences between HOT and COLD ($z = -6.34, p < 0.01$) and between COLD and WARM ($z = -4.70, p < 0.01$). HOT was used more correctly than COLD; COLD was used more correctly than WARM. The results did not agree with the findings of the word clue tests: in these, COLD was higher than HOT; HOT was higher than WARM.

In order to find out whether there were differences between the age and gender groups in single information inference from HOT, WARM, or COLD, Mann-Whitney tests were carried separately with HOT, WARM and COLD. There were differences between the age groups in WARM ($z = -4.86, p < 0.01$) and COLD ($z = -5.55, p < 0.01$) in favour of the older groups. No difference was found between the gender groups.

To see how the younger and older groups differed in use of each type of information in the game process, the percentages of correct single information inferences and double information inferences were calculated for each age group. Figure 4.14 shows the percentages of correct single information inferences from the three kind of information for the two age group.

The younger children made correct use of WARM and COLD on less than half of the occurrences. The older children were better at inferring from WARM and COLD than the younger children, although the correct use of these two items of information was not as high as the use of HOT for both age groups.

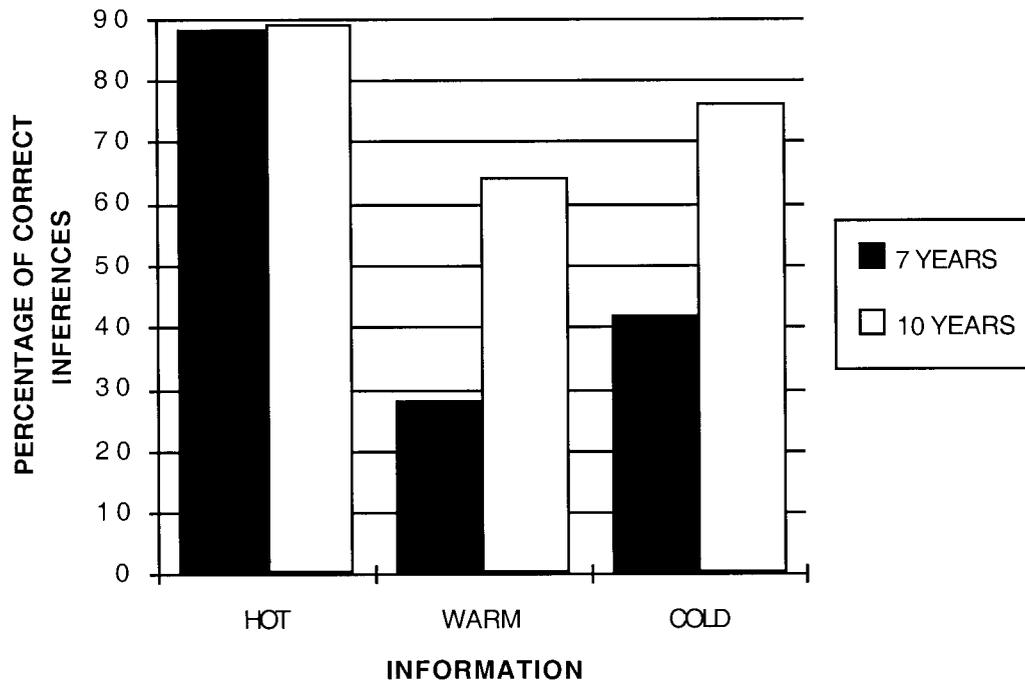


Figure 4.14. Percentages of single information inferences for two age groups (7years and 10 years)

In order to see whether there were differences in inferring from two items of information between age groups, the percentage of correct inferences from two successive pieces of information was calculated for each possible pair of information items. For the 10-year-old children, the percentages of correct inferences from the double information WARM-COLD and those from COLD-WARM were below 60. These two double information inferences were also among the four double information inferences where the 7-year-old children had a percentage of correct inference below 40. WARM and COLD clues still appeared to be more difficult for the younger age group to infer from. Figure 4.15 shows the percentages of correct double information inferences for the two age groups.

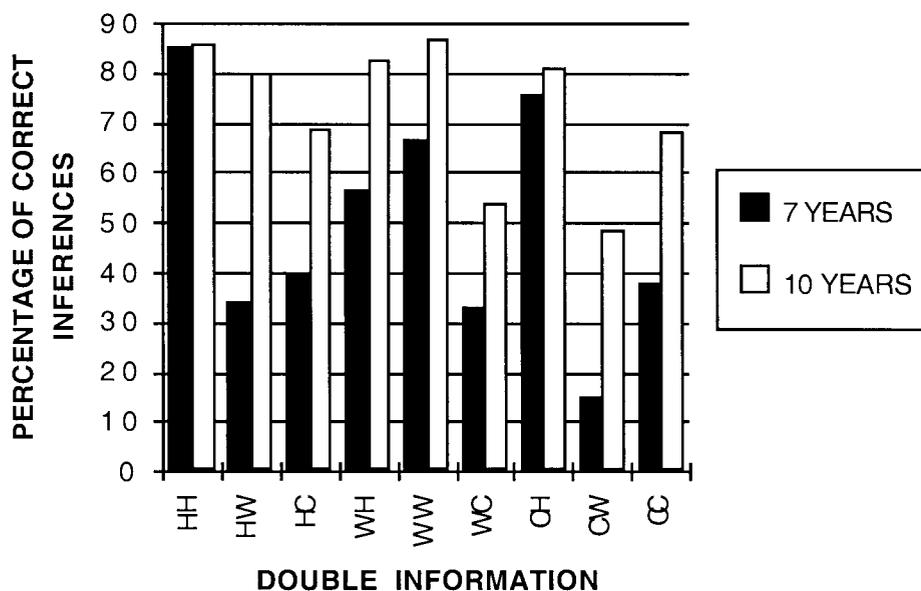


Figure 4.15. Percent of correct double information inferences for two age groups (7 years and 10 years)

4.5. Summary of results

Results will be summarized first with reference to the replication of the previous study, including evidence of developmental change, and secondly with reference to individual differences, learning and choice-patterns.

1. The majority of the 10-year-old children made moves below the criterion, while the majority of the 7-year-old children made moves around chance level. There were significant differences between the age groups in the scores for each of the word clue tests: HOT, WARM, COLD, HOT-WARM, WARM-COLD, and HOT-WARM-COLD.
2. There was no significant difference in the mean number of moves between the computer and the board groups. No significant difference was found in scores for the word clue tests, either.
3. There was no significant difference in the mean number of moves between gender groups. Scores for the word clue tests did not show significant differences, either.
4. There were significant differences in scores between the word clue tests. Scores for COLD were significantly higher than for HOT; HOT was scored significantly higher than WARM. The order of scores of single word tests was identical in Studies 1 and 2. The results of the word clue tests in this study confirm the levels of difficulties of single word clues detected in the previous study.
5. There was a significant correlation between the mean numbers of moves of odd and even sets of games.

6. There was a significant correlation between the mean number of moves and total scores for the word clue tests.

7. There was a significant difference in the word clue test scores between the two criterion groups.

8. Backward learning curves showed that the patterns of responses found in the final games already existed from the beginning of games.

9. There was a significant difference in the mean numbers of moves between the first and last four games. A significant difference was also found in the time taken to complete the game between the two blocks of games for 10-year-old children.

10. Faced with HOT, most of the moves were made to the horizontally or vertically adjacent places. Faced with WARM, the moves were made to the H/V adjacent places nearly as often as to the diagonally adjacent places. But moves to the non-adjacent places were rare with WARM. Faced with COLD, the majority of the moves was made to the non-adjacent places.

4.6. Conclusions

The study investigated individual differences and learning in the inferential game play. Analyses were made in two perspectives on the game play, focusing on the player and the information. One perspective focused on the child who made inferences from the game rules: whether she or he perceived the game as a problem-solving task or a chance event, or how accurately she or he inferred the goal place. The other focused on information in the game frame: how correctly each item of information was used to infer the goal places in the game process. These two perspectives depict the inferential game play more comprehensively than either alone.

Developmental change in inferential problem solving was clearly shown. Most of the 10-year-old children made fewer moves than the criterion for success by chance, whereas on average the 7-year-old children could not achieve this. It took as many moves as working at the chance level would need for the younger children to find the Flamingo. Developmental differences were shown not only in the game situation, but also in the word clue tests. The 7-year-old children usually made poor judgements about the possibility and impossibility of the goal place, whereas the 10-year-old children showed significantly better judgements, inferring from the game rules. Considering these results together with the results from Study 1 which suggested that the majority of 8-year-old children used the game rules, there seems to be a big developmental change in inferential game play between 7- and 8-year-old children.

Analyses of the information correctly used during the game process showed that WARM and COLD led the age differences in the game performance. No difference in inferring from HOT was found between age groups. This may be due to the younger children being too good to leave room for significant improvement. Why was HOT but

not WARM used highly? HOT and WARM have the same meaning: inclusion of possible goal places. Analyses of incorrect inferences showed that the children did not code WARM specifically enough. They regarded WARM as the signal for a move to any adjacent place. Faced with WARM, moves to the horizontally or vertically adjacent places were made nearly as often as to the diagonal places. But moves to non-adjacent areas were quite rare. Was this because keeping two similar clues in mind was too complicated for the children? Or was moving diagonally difficult? There is evidence supporting the second claim that young children have difficulty in drawing diagonal relationship (Bryant, 1974).

Individual differences have been shown to be quite constant over games. The good performance group in the later games included those who started the game with much efficiency, while the poor performance group in the later games included those who started the game without much efficiency. The grouping offered a means of exploring the diversity of inferential ability within children of the same age in a way that is not feasible with reliance on a single mean for the whole group. The 7-year-old good performance group was better than the 10-year-old poor performance group. These variances might stem from the strategies they applied in the games. Those who tended to think about the consequence of a choice, whether one could be better than others, may have performed better than those who merely guessed the goal place.

No game media difference was found, for either the younger or older children. Nor were gender differences found. Because these results replicate the findings of Study 1, I can become more confident that computers do not affect children's inferential problem solving and that the nature of the game is gender neutral. The most impressive observation for the experimenter was that all the children were very eager to play the game on computers. Regardless of whether or not a child managed to solve the problem, with logic or using random choices, she or he reported high satisfaction and joy. Nonetheless, the study showed that the effects of computer activities on learning

did not come from the computer itself, but from the activity. The result points to the importance of the design and use of educational game software. Bearing this in mind, the situational variance should also be explored: how the child's game performance differs when playing alone, with peers or with adults' guidance.

The study contributes to the measurement of inferential problem solving in games. Firstly, it can confidently be asserted that counting the number of moves is a reliable measurement of children's inferential problem solving in the process of game playing. The finding of a high correlation between the number of moves in the games and scores for the word clue tests in Study 1 was supported by Study 2. The correlation between the mean number of moves and scores for the word clue tests was again statistically significant. Furthermore, the correlation between the mean numbers of odd and even sets of games was significantly high in both studies. Secondly, this study advocates considering both single information inference and double information inference when measuring use of the rules in relation to information in games. These measurements are powerful for the task analyses: which information and concepts the task consists of and which element poses a challenge to the player. Despite the limitation in the detection of more advanced and sophisticated use of information, these measurements captured well the phenomena of the game play: how each age group used information.

Two shortcomings of this study come from the design of the experiment. Firstly, it would be more desirable if there had been many games over an extensive period for research on learning with computer games. The study did not allow exploration of the effect of the characteristic repetitiveness of computer games. The 10-year-old participants were already too successful to gain from more of the same games. Their performance of inferential problem solving reached a ceiling within 8 trials. Some studies (Ceci, 1990) have shown that mastery of cognitive skills develops over hundreds of experiences of practice with a game. Secondly, it is desirable to give children an opportunity to play the game after a break. Since some children came

voluntarily to play the game after the experiment and a few actually made improvements at this time, it would have been better to allow more games after an interval. Two reasons were detected for these improvements. One was the realization of the game structure. The other was the consideration of the 5x5 grid as a whole, and the beginning of strategic moves to find at least a HOT or a WARM clue. The most developed use of a 5x5 grid is the division of the grid into 4 square quarters. Choices of the middle place of each quarter are the most efficient way to get HOT or WARM clues. Concerning the concept of the whole-part relation, we will examine more closely the use of logic in inferential problem solving in Study 4.

This study has shown that making and using inferences from the game rules were two different matters in the game. Replicating the findings of Study 1, the results of the word clue tests of Study 2 show that the judgement of COLD as a clue to exclude non-adjacent areas was high, even among younger children. Children were significantly better at making inferences from COLD than from HOT and WARM. However, analysis of correct inferences from each type of information in the game process showed that HOT was used correctly significantly more often than WARM or COLD. The conflict between these findings can be explained in terms of the difference between knowledge and its use. Scores for the tests stand for the ability to make inferences from the game rules, that is to say knowing what inference is, whereas the game performances show the actual use of inferences. Thus, these contrasting findings between the results of the word clue tests and the analysis of information in the actual games may well relate to how children plan to use their knowledge. This assumption was explored in the next study, which investigated the effects of training on the planning of the use of inclusions and exclusions.

Chapter 5. Study 3

5.1. Aims

The study aimed to investigate the effects of 'guided-planning' and 'pausing' in inferential problem-solving in a computer game context. Inclusion and exclusion inferences were explored in the game.

5.2. Introduction

Studies 1 and 2 measured children's inferential problem-solving ability in a game in terms of the moves needed to complete the game, and of judgement scores which were based on possible and impossible places and ascertained through the experimenter's questioning with one or two pieces of information each time. The two measurements that correlated highly in both studies revealed that there were consistent individual differences throughout the practice of the game. Those who played well in the later games were the ones who started the game with a higher performance, and those who played the game around chance success level were the ones who started the game with poorer performances. Furthermore, it was shown that the COLD exclusion information was not used much in the game process, even though children's judgement of the possible and impossible goal places was relatively high with the COLD clue in the test situation where the clue was presented via the experimenter's question and pointing,

and with intervals between the clues. The assumption was that two factors might underlie individual differences, besides the ability to make inferences already shown in Studies 1 and 2. One factor is impulsiveness. The poor performance group might be the more impulsive children. They had a disadvantage because of the hasty nature of their play in the game. Even though it might be that they were capable of making links between the game rules and clues, and between the clues and the spatial relations to the goal, the children neither saw that they were supposed to use the inferences nor realized the advantage of using them. Consequently, they made random choices, which led to their poor performance. The other factor underlying individual differences might be planning. Poor performance could be due to the inability to plan how to use inferences arising from information. To use inferences arising from information, the child has, first of all, to make links between the information and the game rules. Then she or he must be able to anticipate where the goal might be, and then to choose the next move amongst one of the possible places. She or he would fail to use the inference if she or he does not plan the next moves in the way that the inference indicates.

In an attempt to explore ways of enhancing poor performance, this study aimed to compare the effects of guided-planning and pausing in inferential game play. The idea of guided-planning comes from research on problem-solving, in which the more knowledgeable person acts as an active helper to stop the learners from moving too fast and guide them by questioning to think about the goal of the problem and the means-end relationship. This joint plan for the more able and less able, a consciously adopted strategy, will guide the use of the clue as a way to determine further moves towards the goal and to select appropriate moves to bring this about (Gauvain & Rogoff, 1989; Radziszewska & Rogoff, 1988). Mediation tasks illustrate the effects of the more able persons' verbal mediation on the less able children's representation of the problem. For example, Kendler and Kendler (1962) showed that children who were not able to represent a situation, for instance, that big stimuli were reinforced regardless of their colour, were helped by repeating the label, "Big", "Big", "Big". This labelling then

also helped children to infer that under the concept of 'size is relevant in this task', even a small stimulus *is* correct when "Big" *is not* any longer. However, the relevant labelling helped only those who had not learned the concepts, whilst those who already had ability to form the concepts were disturbed by irrelevant labelling. In the Flamingo game, I assume that the good performance group finds the goal through the conceptualization of the clues HOT, WARM, and COLD, as showing spatial relations to the goal: the horizontally or vertically adjacent relation to the goal (HOT), the diagonally adjacent (WARM), and the non-adjacent (COLD). In the guided-planning condition, the experimenter directed the player's attention by asking questions about the possible or impossible goal places in relation to the clues, that is, about their spatial relations to the goal. Through guided-planning, not only is the player guided to conceptualize the variable as the game rules define, but she or he also becomes aware of the benefit of making choices in this way. The effects of the verbal mediation of the variables and individual differences in these effects were expected.

In the 'pausing' condition, the amount of time the child spent between seeing the clue and making the next choice was controlled, because impulsiveness makes the child fail to stop and think before reacting to an unfamiliar situation or situations in such a way that causal relationships or consequences are fully considered. Behavioural psychology literature has often identified the effects of forced delay used as a way to 'calm down' or 'cool' while problem solving (see Kagan, Rosman, Day, Albert & Phillips, 1964; Kagan, Pearson & Welsh, 1966). If children are capable of making inferences from the game rules, the next step is that they have to remember and retrieve the link between the clue and the spatial relation to the goal and use it to guide the next move. 'Pausing' - a delay before the action - gives the player a deliberate opportunity to think before the next move. The effects of 'pausing' seem to depend on individual differences in spontaneous rehearsal. The training of youngsters' rehearsal with forced delay has shown strong effects (Cox, Ornstein, Naus, Maxfield & Zimler, 1989; Keeney, et. al, 1967; Naus, Ornstein & Aivano, 1977; Ornstein, Naus & Stone, 1977). If the children

really know what they are supposed to do in the game, to be forced to delay before making a choice would help them to plan the next move on the basis of the current information. There need be no explicit request for making links between information and the game rules. Pausing is an implicit aid to thinking beforehand, contrasting with guided-planning which explicitly provides questions about the goal places vis-a-vis the current information.

Study 3 aimed to investigate whether guided-planning and pausing can help problem-solving. For these research questions, a software game was developed. The game, called 'Find the Boat', is structured by two rules. Firstly, it consists of inclusion and exclusion information. Secondly, it involves a fair comparison of the two kinds of information. In the Flamingo game, spatial information confounds the two kinds of inclusion: the inclusion of horizontally or vertically adjacent places and that of diagonally adjacent places. In order to concentrate on the comparison between inclusion and exclusion, it is better to separate the inclusion information from the directional cues. Therefore, the Flamingo game was simplified into a game called 'the Boat game'. It has only one kind of inclusion information, without directional clues, and that is 'touching'. The 'touching' information is equivalent to HOT and WARM and the not-touching is to COLD. The probability of error has changed: there are a maximum of eight possible correct responses to the inclusion information in the new game, and four responses for HOT or WARM in the Flamingo game. The zone of possible correct responses is bigger in the Boat game. Even though the numbers of correct responses between inclusion and exclusion information are still unbalanced, it is fairer to make a comparison of inclusion and exclusion in the Boat game than in the Flamingo game.

5.3. Method

5.3.1. Design

The study had a mixed within-subjects and between-subjects design. It aimed to answer the research questions in these ways:

(1) comparing performance on the pre-training, the training and the post-training games,

(2) comparing performance change between three groups: the guided-planning, the pause and the control groups, and

(3) comparing performance between the uses of information of inclusion and exclusion.

The first method was used to examine whether the performances changed overall. This was in preparation for the second method, which was used to see whether there were differences between the treatment groups. Using the first and second methods, the difference between inclusive and exclusive inferences were examined.

5.3.2. Participants

33 children in year 1 (mean age = 6 years 5 months, S.D. = 3.7 months), 36 children in year 2 (mean age = 7 years 5 months, S.D. = 3.8 months), and 26 children in year 3 (mean age = 8 years 5 months, S.D. = 3.5 months) participated. This was the total number of year 1, 2, and 3 children from an inner London school. In the order of their entry to the scene of the experiment, the children were randomly assigned to each of the following three groups: guided-planning, pausing or control groups.

Table 5.1. Number of participants across age and treatment groups

	6 years	7 years	8 years	Total
GUIDE*	11	12	9	32
PAUSE	11	12	8	31
CONTROL	11	12	9	32
Total	33	36	26	95

GUIDE*: Guided-planning

5.3.3. Procedure

Features of the game

The game was programmed on a BBC computer that was attached to a touch screen. There was a 5x5 matrix of cards on the screen. Among these 25 cards, only one was a goal that was programmed to change the position from one game to another game randomly.

The game was administrated to each child individually.

The instructions were the following:

Turn over a card by touching the screen with your finger.
Imagine it is a large blue sea. We need the boat desperately. Only one card has a boat.
If you follow the rules of the game, you can find the boat more easily.
If it's there, we are saved.
If the boat touches you, you are X.
If it does not touch you, you are O.

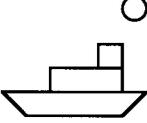
O	O	O	O	O
O	X	X	X	O
O	X		X	O
O	X	X	X	O
O	O	O	O	O

Figure 5.1. An example of the Boat game

The child was asked to explain the rules to the experimenter. If the child understood the rules properly, she or he started to play the game. If not, the experimenter explained the rules. Each child played the game 12 times. Each time, the boat was hidden in any of 25 places. The two experimental groups of children played three blocks of the game. The children had 4 games as a pre-test, the next 4 games as intervention training, and 4 more games as the post-test. In the first 4 games, the children had the opportunity to play the game and get to know the rules. Before the 5th game, the player was told the game rules again.

In the 4 training games, the guided-planning group children were asked to stop before making a next move. They were asked questions by the experimenter. The questions were as follows:

"Can the boat be here (pointing to one of the adjacent places)?"

The same questions were repeated for any three adjacent places.

"Can it be here (pointing to a far place)?"

The same question was repeated for any three non-adjacent places.

"Where do you think the boat is?"

"Did the rules say so?"

In the 4 training games, the 'pausing' group children were given a 10-second 'pause' before making each move until the completion of the game. After the intervention, the child played 4 games more. The control group had 12 games without any intervention except reminding the children about the rules before the 5th game.

5.3.4. Outcome measures

1. Mean numbers of moves in the pre-training, training, and post-training games.
2. Correct inferences from inclusion information (see examples in Figure 5.2) in the pre-training, training and post-training games.
3. Correct inferences from exclusion information (see examples in Figure 5.3) in the pre-training, training and post-training games.

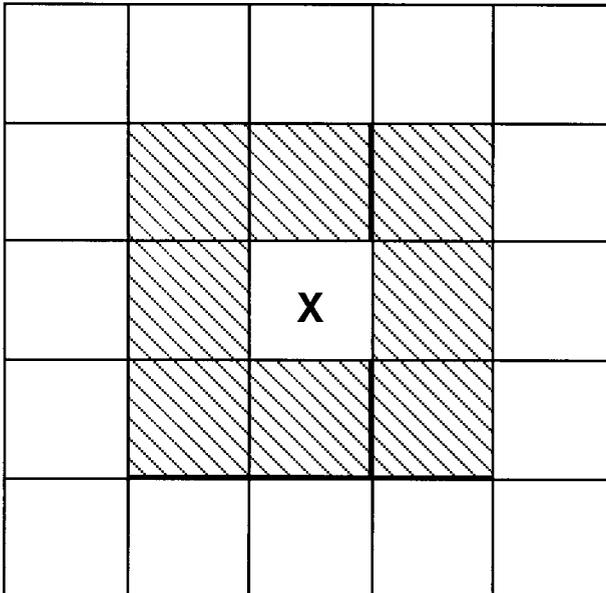


Figure 5.2. Zone of correct inference from X information

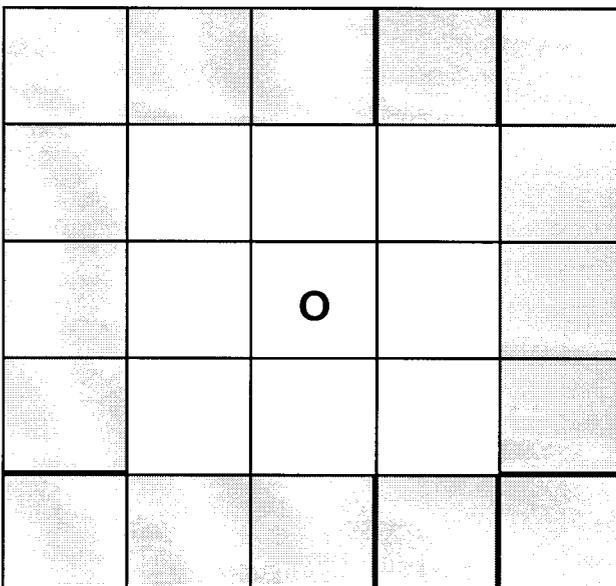


Figure 5.3. Zone of correct inference from O information

5.4. Results

Analysis of the training effect involved two steps. Firstly, it was tested whether there were performance changes before and after the training. Age and gender group differences were tested at this stage. Secondly, it was tested whether there were differences between the results of the treatments (guided-planning, pause and control) during the training and post-training games. These two steps of analysis were based on three sets of outcomes: the number of moves, and inclusion and exclusion inferences. The first part of the results reports analyses of the number of moves. The second part reports analyses of inclusion inferences. The third part reports analyses of exclusion inferences.

5.4.1. Analyses of numbers of moves

In order to see how the children performed, the mean numbers of moves in the pre-, during, and post-training games was calculated for each of the treatment and age groups. Table 5.2 shows the mean numbers of moves in the pre-, during, and post-training games across the treatment and age groups.

The mean numbers of moves were 10.85 (S.D. 4.67) for the pre-training games, 8.28 (S.D. 3.84) for the training games, and 8.12 (S.D. 3.95) for the post-training games. Inspection of Table 5.2 shows that the mean numbers of moves of the pre-training games varied across the treatment and age groups. Thus, the study used two steps of analysis. The first step was a global pre-post performance comparison to see if there

was improvement. Then, the second step was a treatment group comparison, controlling the child's pre-training performance and age, as the pre-training performance suggested the children were not homogeneously distributed across these groups.

In order to test whether there were differences between the pre- and post-training games or between the age and gender groups, a three-way mixed design ANOVA was conducted, using the block of games (the pre- and post-training games) as within-subjects factor and age and gender as between-subjects factors. Table 5.3 shows the results of the Analysis of Variance.

There were significant differences between the pre- and post-training games ($F = 28.51$ with 1 d.f., $p < 0.01$). The mean number of moves in the pre-training games was significantly higher than that in the post-training. There was no gender difference. Interactions were found between the blocks of games (the pre-, during and post-training games) and the age groups ($F = 3.30$ with 2 d.f., $p < .05$). The decrease in the number of moves from the pre-training games to the post-training games was 3.3 fewer moves for the 6-year-old children; 1.0 fewer move for the 7-year-old children; and 4.4 fewer moves for the 8-year-old children. Figure 5.3 shows interactions between the age groups and the blocks of games.

Table 5.2. Mean number of moves across treatment and age groups

	6 years		7 years		8 years		Total							
	Pre	Post	Pre	Post	Pre	Post	Pre	Post						
GUIDE	13.75	8.75	9.27	8.86	10.33	8.81	7.08	7.71	11.17	6.58	7.11	11.74	7.52	8.59
PAUSE	10.55	8.68	8.86	8.86	8.81	7.71	7.48	7.48	12.19	6.75	6.94	10.30	7.81	7.83
CONTROL	12.16	11.07	8.41	8.41	7.92	8.23	7.48	7.48	11.86	9.33	7.92	10.48	9.52	7.92
Total	12.15	9.50	8.85	8.85	9.02	7.63	8.01	8.01	11.72	7.59	7.34	10.85	8.12	8.28

GUIDE*: Guided-planning
 Pre; Pre-training games
 On; During training games
 Post; Post-training games

Table 5.3. Three-Way Analyses of Variance on performance over the practice of games (n = 95)

Sources of Variance	SS	DF	MS	F	Sig. of F
Between group					
Age (A)	134.03	2	67.02	2.85	.063
Gender (G)	.44	1	.44	.02	.892
Interactions					
A X G	4.96	2	2.48	.11	.900
<u>S</u> Within-group error	2094.73	89	23.54		
Within group					
Block of Games* (B)	371.10	1	371.10	28.51	.001
Interactions					
B X A	86.04	2	43.02	3.30	.041
B X G	3.08	1	3.08	.24	.628
B X A X G	20.15	2	10.07	.77	.464
B X <u>S</u> Within-group error	1158.54	89	13.02		

Block of Games*; pre- and post-training games

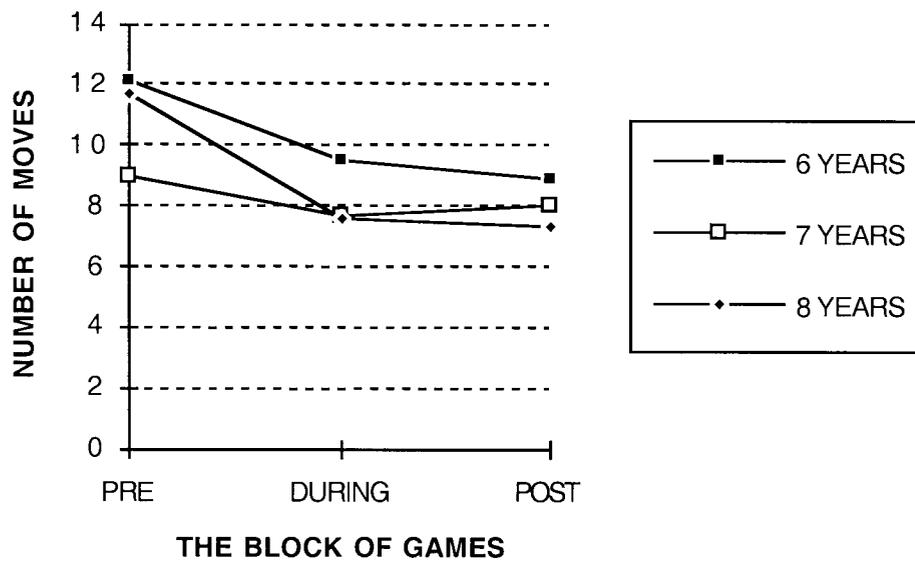


Figure 5.4. Interactions between age groups and blocks of games

In order to test whether there were differences between the treatment groups, a one-way between-subjects design (3 treatment groups) Analysis of Covariance was carried out with the age (in months, for a more precise control) and the pre-training performance as covariates. The participants' age and pre-training performance were calculated in an analysis of the training performance as a way to control sampling error. Table 5.4 shows the results of Analysis of Covariance of the training games.

The training performance was significantly correlated with age ($F = 5.07$ with 1 d.f., $p < .05$) and pre-training performance ($F = 10.13$ with 1 d.f., $p < .01$). The removal of the effects of the covariates (the age and the pre-training performance) showed that there were significant differences between the treatment groups ($F = 3.87$ with 2 d.f., $p < .05$). A post-hoc test (A least significant test) shows that the guided-planning group was significantly better than the control group ($p < 0.05$). No difference was found between the guided-planning and pausing groups or between the pausing and control groups.

In order to see whether the treatment effects lasted in the post-training games, a one-way between-subjects design (3 treatment groups) Analysis of Covariance was carried out with the child's age (in months) and pre-training performance as covariates. Table 5.5 shows the results of Analysis of Covariance of the post-training games.

The post-training performance was significantly correlated with the pre-training performance ($F = 6.55$ with 1 d.f., $p < .05$). Even after controlling the effect of the covariates (the child's age and pre-training performance), there was no significant difference between the treatment groups. The treatment effects shown during the training were washed away in the post-training games. In the next two sections of the results, the effects of the treatments will be analysed with inferences from inclusion and exclusion information.

Table 5.4. One-Way Analysis of Covariance of training performance (n = 95)

Sources of Variance	SS	DF	MS	F	Sig. of F
Covariate					
Age	61.07	1	61.07	5.07	.027
Pre-training	122.01	1	122.01	10.13	.002
Main effect					
Treatments	93.15	2	46.58	3.87	.024
Explained	271.88	4	67.97	5.64	.001
Residual	1084.20	90	12.05		
Total	1356.20	94	14.43		

Table 5.5. One-Way Analysis of Covariance of post-training performance (n = 95)

Sources of Variance	SS	DF	MS	F	Sig. of F
Covariate					
Pre-training	94.59	1	94.59	6.55	.012
Age	49.23	1	49.23	3.41	.068
Main effect					
Treatments	3.32	2	1.66	.12	.892
Explained	165.91	4	41.48	2.87	.027
Residual	1300.81	90	14.45		
Total	1466.73	94	15.60		

5.4.2. Analyses of inclusion inferences

In order to find out whether there were treatment effects on inclusion inferences, correct inferences from inclusion information were calculated for each of the treatment groups. Because the numbers of occurrences of inclusion information varied from child to child, the percentage of correct inclusion inferences was calculated for each child, using the numbers of occurrences of inclusion information and of correct inferences from the information. Table 5.6 shows the mean percentages of correct inclusion inferences across the treatment and age groups.

The mean percentages of correct inclusion inferences were 77.7 (S.D. 20.7) for the pre-training games, 82.3 (S.D. 16.9) for the training games and 87.2 (S.D. 13.3) for the post-training games. Inspection of Table 5.6 shows that the percentage of correct inclusion inferences in the pre-training games varied across the treatment and age groups. The children were not distributed homogeneously across the groups. Thus, the study carried two steps of analysis. The first step was a global pre-post performance comparison to see whether there was improvement. Then, the second step was a treatment group comparison, controlling the child's pre-training performance and age as these suggested that the children were not homogeneously distributed.

To allow the comparisons, data in percentages were converted to proportions and transformed through *acsign*. In order to test whether there were differences between the pre- and post-training games or between the age and gender groups, a three-way mixed design ANOVA was conducted, using the block of games (the pre- and post-training games) as within-subjects factor and age and gender as between-subjects factors. Table 5.7 shows the results of the Analysis of Variance.

Table 5.6. Mean percentage of correct inclusion inferences across treatment and age groups

	6 years			7 years			8 years			Total		
	Pre	On	Post	Pre	On	Post	Pre	On	Post	Pre	On	Post
GUIDED*	78.4	72.0	86.6	80.9	83.7	91.8	82.2	86.1	87.6	80.4	79.4	88.8
PAUSE	73.3	79.9	85.8	86.1	87.8	86.8	54.1	86.9	84.2	73.3	87.3	85.8
CONTROL	73.1	86.3	87.6	90.1	90.5	86.9	71.8	76.7	86.2	79.1	83.1	86.9
TOTAL	75.0	80.3	86.7	85.7	84.8	88.5	69.9	85.2	86.0	77.7	82.3	87.2

GUIDE*: Guided-planning
 Pre; Pre-training games
 On; During training games
 Post; Post-training games

Table 5.7. Three-Way Analysis of Variance on inclusion inferences (n = 95)

Sources of Variance	SS	DF	MS	F	Sig. of F
Between group					
Age (A)	1.03	2	.51	3.48	.035
Gender (G)	.12	1	.12	.81	.370
Interactions					
A X G	.07	2	.03	.23	.793
Σ Within group error	13.15	89	.15		
Within group					
Block of Games* (B)	1.58	1	1.58	13.68	.001
Interactions					
B X A	.40	2	.20	1.74	.181
B X G	.00	1	.00	.03	.854
B X A X G	.07	2	.03	.29	.749
Σ X B Within group error	10.26	89	.12		

Block of Games*; pre- and post-training games

There were significant differences between the pre-and post-training games ($F = 13.68$ with 1 d.f., $p < 0.01$). The mean percentage of correct inclusion inferences in the post-training games was significantly higher than that in the pre-training games. Significant differences were also found between the age groups ($F = 3.37$ with 2 d.f., $p < 0.05$). The mean percentages of correct inclusion inferences of the pre- and post-training games were 80.9 for the 6-year-olds, 87.1 for the 7-year-olds, and 78.0 for the 8-year-old children. On inspection of the means, the 7-year-old children were considerably higher than the other two age groups. No significant gender difference was found.

As a way to control sampling error, the child's age (in month for a more precise analysis) and pre-training game performance were calculated in the examination of the effects of the treatments. A one-way between-subjects design (3 treatment groups) Analysis of Covariance was carried with the age and the pre-training performance as covariates to test whether there were differences between the treatment groups. Table 5.8 shows the results of Analysis of Covariance of the training performance.

The training performance was not correlated either with the child's age or with the pre-training performance. The removal of the effects of the covariate (the child's age and pre-training performance) showed that there was no difference between the treatment groups.

In order to test whether there were treatment effects in the post-training games, a one-way between-subjects design (3 treatment groups) Analysis of Covariance was carried with the age (in month) and the pre-training performance as covariates. Table 5.9 shows the results of Analysis of Covariance for the post-training games.

Table 5.8. One-Way Analysis of Covariance of training performance of inclusion inferences (n = 95)

Sources of Variance	SS	DF	MS	F	Sig. of F
Covariate					
Age	.187	1	.187	1.63	.205
Pre-training	.330	1	.330	2.88	.093
Main effect					
Treatments	.072	2	.036	.31	.732
Explained	.574	4	.143	1.25	.296
Residual	10.34	90	.115		
Total	10.91	94	.116		

Table 5.9. One-Way Analysis of Covariance of post-training performance of inclusion inferences (n = 95)

Sources of Variance	SS	DF	MS	F	Sig. of F
Covariate					
Age	.018	1	.018	.163	.687
Pre-training	.206	1	.206	1.183	.179
Main effect					
Treatments	.141	2	.071	.627	.536
Explained	.375	4	.094	.834	.507
Residual	10.130	90	.113		
Total	10.505	94	.112		

The post-training performance was not correlated either with the child's age or with the pre-training performance. The removal of the effects of the covariates (the child's age and pre-training performances) showed that there was no difference between the treatment groups. The differences between the pre- and post-training performance shown in the results of Analysis of Variance (see Table 5.7) did not come from the treatment. Learning occurred over the practice of the games without the intervention in planning.

5.4.3. Analyses of inferences from exclusion information

In order to find out whether there were treatment effects on exclusion inferences, correct inferences from exclusion information in the pre- and post-training games were calculated for each of the treatment groups. Because the numbers of occurrences of exclusion information varied from child to child, the percentage of correct exclusion inferences were calculated for each child, using the numbers of occurrences of exclusion information and of correct inferences from the information. Table 5.10 shows the mean percentages of correct inferences from exclusion information across treatment and age groups.

The mean percentages of correct exclusion inferences were 46.2 (S.D. 22.5) for the pre-training games, 56.3 (S.D. 24.8) for the training games, and 47.7 (S.D. 25.6) for the post-training games. Inspection of Table 5.10 shows that the percentages of correct exclusion inferences in the pre-training games varied across the treatment and age groups. The children were not distributed homogeneously across the groups. Also, there was the same pattern of improvement during training and regression in the post-training performance, in the guided-planing and pausing groups. But this did not happen in the control group. Thus, the study carried two steps of analysis. The first step was a global pre-post performance comparison, to see whether there was improvement. Then, the second step was a treatment group comparison, controlling the child's pre-training performance and age, as these suggested the children were not homogeneously distributed.

To allow the comparisons, data in percentages were converted to proportions and transformed through *acsign*. In order to test whether there were differences between the pre- and post-training games or between the age and gender groups, a three-way

mixed design ANOVA was conducted, using the block of games (the pre- and post-training games) as within-subjects factor and age and gender as between-subjects factors. Table 5.11 shows the results of the Analysis of Variance.

No difference was found between the pre-and post-training games ($F = .38$ with 1 d.f., $p = .54$). Differences were found between neither the age groups nor the gender groups.

A one-way between-subjects design (3 treatment groups) Analysis of Covariance was carried out, with the age and the pre-training performance as covariates, to test whether there were differences between the treatment groups during the training games. Table 5.12 shows the results of Analysis of Covariance of the training performance.

The training performance was highly correlated with the pre-training performance ($F = 15.6$ with 1 d.f., $p < .001$). The removal of the effects of the covariates (the child's age and pre-training performance) showed that there were significant differences between the treatment groups. A subsequent Student-Newman-Keuls test showed that the guided-planning group scored significantly higher than the pausing group ($p < 0.05$); the pausing group scored significantly higher than the control group ($p < 0.05$).

Analyses of the post-training performance were not necessary because there was no reason to expect treatment effects on the performance. The results of Analyses of Variance (Table 5.11) had already informed that there was no improvement between the pre- and post-training games.

Table 5.10. Mean percentage of correct exclusion inferences across treatment and age groups

	6 years			7 years			8 years			Total		
	Pre	On	Post	Pre	On	Post	Pre	On	Post	Pre	On	Post
GUIDED*	37.7	66.9	51.5	49.4	68.6	48.8	43.6	69.9	56.4	43.8	68.4	51.9
PAUSE	39.2	45.9	36.1	50.3	65.5	50.9	52.4	64.8	57.5	46.9	58.4	47.4
CONTROL	41.9	35.5	36.8	49.3	47.4	48.6	53.2	43.1	45.8	47.8	42.1	43.8
TOTAL	39.6	49.4	41.5	49.7	60.5	49.4	49.6	53.1	59.0	46.2	56.3	47.7

GUIDE*: Guided-planning
 Pre; Pre-training games
 On; During training games
 Post; Post-training games

Table 5.11. Three-Way Analysis of Variance on exclusion inferences (n = 95)

Sources of Variance	SS	DF	MS	F	Sig. of F
Between group					
Age (A)	.40	2	.20	2.22	.115
Gender (G)	.19	1	.19	2.11	.150
Interactions					
A X G	.07	2	.04	.40	.671
<u>Within group error</u>	7.99	89	.09		
Within group					
Block of Games* (B)	.01	1	.01	.38	.537
Interactions					
B X A	.01	2	.00	.16	.855
B X G	.01	1	.01	.51	.479
B X A X G	.02	2	.01	.48	.621
<u>X B Within group error</u>	2.18	89	.02		

Block of Games*; pre- and post-training games

Table 5.12. One-Way Analysis of Covariance of training performance of exclusion inferences (n = 95)

Sources of Variance	SS	DF	MS	F	Sig. of F
Covariate					
Age	.090	1	.090	9.43	.313
Pre-training	1.37	1	1.37	15.6	.001
Main effect					
Treatments	2.40	2	1.20	13.7	.001
Explained	3.70	4	.923	10.56	.001
Residual	7.89	90	.088		
Total	11.59	94	.123		

5.5. Summary of the results

Analyses of the numbers of moves show:

1. There were significant differences between the pre- and post-training games. Interactions were found between the blocks of games and the age groups. The 6- and 8 year-old children learned more than the 7-year-old children through the practice of games.
2. The training performance was significantly correlated with the child's age and pre-training game performance. After the removal of the effects of the age and pre-training performance, there were significant differences between the treatment groups during the training games. The guided-planning groups performed significantly better than the control group.
3. The post-training performance was significantly correlated with the pre-training performance. After the removal of the effects of the child's pre-training performance and age, no significant difference was found between the treatment groups in the post-training games. The guided-planning group did not perform better in the post-training games than in the pre-training games. Nor did the other two groups.

Analyses of inclusion inferences show:

4. There were significant differences between the pre- and post-training games and between the age groups. The performance of the 7-year-old children was highest among the three age groups during the pre-and post-training games.

5. There was no difference between the treatment groups in the training games.
6. No difference was found between the treatment groups in the post-training performance.

Analyses of exclusion inferences show:

7. There was no difference between the pre- and post-training games.
8. The training performance was highly correlated with the pre-training performance. After the removal of the effects of the covariates (the child's pre-training performance and age), there were significant differences between the treatment groups in the training. The guided-planning group performed significantly better than the pausing group; the pausing group performed significantly better than the control group.
9. The post-training performance was significantly correlated with the pre-training performance. After the removal of the effects of the covariates (the child's pre-training performance and age), there was no difference between the treatment groups in the post-training performance. Compared with the training performance, the proportions of correct exclusion inferences of the guided-planning and pausing groups declined in the post-training games.

5.6. Discussion and conclusions

The study investigated the effects of guided-planning and pausing on inferential problem solving in a computer game that provided inclusion and exclusion information. In an attempt to enhance poor performance, the children had a brief training session using questioning or pauses, which were presumed to have a function of encouraging the planning of use of inferences from the game rules. In the guided-planning condition, the experimenter asked questions about the possibility and impossibility of the goal in relation to the clues; in the pause condition, pauses were inserted into the game as a deliberate opportunity for children to think how to use inferences. Analyses of children's responses were carried out in three ways: analyses of the number of moves for a general picture of performance, then more specifically analyses of inclusion and exclusion inferences. These inference analyses provided explanations for the general picture of performance.

Interesting response patterns emerged at several points. Firstly, learning occurred over the practice of games. There were significant differences between the pre- and post-training games. The children, on average, made two point six fewer moves in the post-training games than in the pre-training games. The 6- and 8-year-old groups made a big improvement from the pre-training games to the post-training games, whilst the 7-year-old group did not. Significant differences were found in inclusion inferences between the pre- and post-training games. The percentages of correct inclusion inferences increased as the game went on. The children made 9.5% more inclusion inferences in the post-training games than in the pre-training games. Where did this learning come? Did the intervention in planning lead to this change? This will be discussed in connection with the next finding.

Secondly, the treatment effects in the post-training games contrasted with those during the training. The guided-planning group needed significantly fewer moves to complete the game than the control group during the training. But there was no difference between the treatment groups in the post-training performance. Significant differences were found in exclusion inferences during the training. The guided-planning group was significantly better at inferring from exclusion information than the pausing group; the pausing group was significantly better than the control group. But the difference disappeared in the post-training performance. The finding that the treatment effects were shown only during the training was harmonious with the results of verbal rehearsal training on memory tasks in which children had verbal rehearsal training on the test items which then were remembered (Ferguson & Bray, 1976; Keeney, et. al, 1967; Kingsley & Hagen, 1969). It was reported that children performed better during the training, but the effect of the training did not last in the posttest. Children did not use the rehearsal strategy that they learned during the training. Why did the children in the boat game stop planning to use exclusion inferences that they had learned through the guided-planning training, and the usefulness of which they then saw? The performance during the training did not really reflect the child's development of exclusion inference. The child only appeared to learn concepts of O as information that the goal was non-adjacent to the present place. Piaget's theory of development might explain the underlying reason. It seems that the children at these ages give an interpretation to the information of O (assimilation), but knowledge is not truly appropriated because it has not been internally reconstructed (accommodation). The post-training performance clearly showed that the child actually did not acquire the concept as her or his own.

Thirdly, uses of inferences differed greatly between inclusion and exclusion information. The mean percentages of inclusion inferences were 77.7 in the pre-training games, 83.4 during the training, and 87.2 in the post-training games. In contrast, the mean percentages of exclusion inferences were 46.2 in the pre-training

games, 56.2 during the training, and 47.7 in the post-training games. Even though the intervention enhanced exclusion inferences by up to 10 percent during the training comparing with the pre-training games, exclusion inferences were still used less often than inclusion inferences. The higher use of inference from inclusion than exclusion information in this study is further supported by the findings of Studies 1 and 2, in which HOT, inferring to horizontally or vertically adjacent places, was used more often than COLD, inferring to the non-adjacent places. These results suggest that operations dealing O in the Boat game or of COLD in the Flamingo was only partially developed. Because operations dealing with the concept of O or COLD was not fully developed as a clue that the goal was non-adjacent to the present place, cognitive demands in holding on to the zone of goal-impossible places and choosing one of the goal-possible places were too heavy to carry out without an external help. This line of thinking is supported by the results of memory tasks, showing that if the child has categories (proper concepts) to link test items, performance is enhanced (Bjorklund, 1985; Bjorklund & Douglas, 1997). Exclusion and inclusion concept learning and inference of multiple possibility and impossibility are not separable but interrelated. The hypothesis that operations dealing with the concept of COLD in the Flamingo game were only partially developed was tested in the next study that investigated the development of operative logic of inclusions and exclusions.

Chapter 6. Study 4

6.1. Aims

This study aimed to investigate children's developmental levels for logical inference and to predict their performance on the Flamingo game. An assumption was made about possible explanations of the developmental and individual differences found in the previous studies 2 and 3. This assumption concerned developmental stages for logical inference. If there was an operative logic, this would enable the child to perform across the Flamingo game and other inferential tasks. It was intended to test whether there were significant correlations between the Flamingo game and two other inferential tasks, which were similar in both requiring inclusions and exclusions, but differed in the complexity of their requirements. If there were correlations between the game and the tasks, these would allow the prediction of the game performance.

6.2. Introduction

Children's ability to draw a logical conclusion based on indirect evidence or experiences has been shown to develop with their ages in various contexts (Bryant & Roazzi, 1992; Haake, et al., 1980; Piaget & Garcia, 1991; Pieraut-LeBonniec, 1980; Wellman, 1985). Study 2 also showed that inferential problem solving in the Flamingo game develops with age. In addition, the patterns of responses found in the final games already existed

from the beginning. The poor game players, who performed around the criterion of success at chance level in the last games, were those who started the game poorly, whereas the good players did better from the start. These findings lead to the assumption that the child's developmental level of logical inference plays a role in determining whether the child understands the logical structure of the Flamingo game. If this is correct, then it should be possible to predict how good a player a child will be on the basis of the child's performance in a task that assesses the ability to make the same type of inferences.

In order to play the Flamingo game well, the children need, firstly, to keep critical areas in mind, defined by inclusion and exclusion inferences. The game clue of HOT identifies the adjacent squares which are above, below, and to the right or left of the target. The clue of WARM identifies the squares diagonally adjacent to the target. The clue of COLD identifies all the squares non-adjacent to the target. Inferring the goal from these clues necessitates inclusions and exclusions of multiple possibilities and impossibilities and holding the critical areas of possibilities. Secondly, the children need to plan their choices. They need to plan their moves into the critical areas. Would they remain in the zone of possibilities, given that the first choice was not right?

There is a task, discussed in the literature, which investigates this basic level of understanding of inclusion and exclusion. In the Number task, used in Bryant and Roazzi's investigation on children's concept of equivalence (1992), the inclusion of numbers between two boundaries is straight forward. The child is asked to choose a number, for example bigger than 3 and smaller than 5. In one instance, there is only one possible correct number, and in the other instance there are two possible correct numbers. In the case of two possible correct numbers, inferring the "critical areas" of two possibilities requires the child to include numbers within the boundaries that the experimenter sets, while excluding others. Giving negative feedback to the first numbers, the task examines whether the child's later choice(s) would remain in the

critical zone. Children's responses to the task reveal whether or not they are able to infer the zone of possibilities in numbers and to hold it. No planning of a sequence of actions is required for the number choice. The Number task requires inclusion and exclusion in a number context (e.g., 1, 2, 3), whereas the Flamingo game does so in a spatial context (e.g., above, below, left, right). Despite the difference, I still expect that the ability to include is the same in each case.

In order to play the Flamingo game efficiently, the understanding of inclusion and exclusion is not enough. The players need to be able to coordinate several inclusions and exclusions in order to plan to use their understanding. They need to anticipate the possible and impossible areas for the target and plan the moves into the critical area. For the prediction of these action-based inclusions and exclusions, a more sophisticated task would be necessary. A Piagetian task, the Tunnel task (Piaget & Garcia, 1991), was used to investigate children's anticipatory actions in a structure that necessitates coordination of inclusions and exclusions. In this task, the player must choose among the paths, which are branches in the form of a tree; the trunk T divides into two primary branches, A1 and A2. Each of these in turn branches out into two paths, B1 and B2 for A1, and B3 and B4 for A2. From each of these four Bs two further paths separate: C1 and C2 for B1; C3 and C4 for B2; C5 and C6 for B3; and C7 and C8 for B4. Each of the final paths ends up in one of eight terminals (G) (Figure 6.1). A thin ribbon tied to the target is contained in hollow tunnels into which the child can peep, and check the ribbon through windows. Since she or he is instructed to open as a few windows as possible, the child is supposed to use the information gained from opening them to include and exclude paths and thus to determine what terminal the target is in.

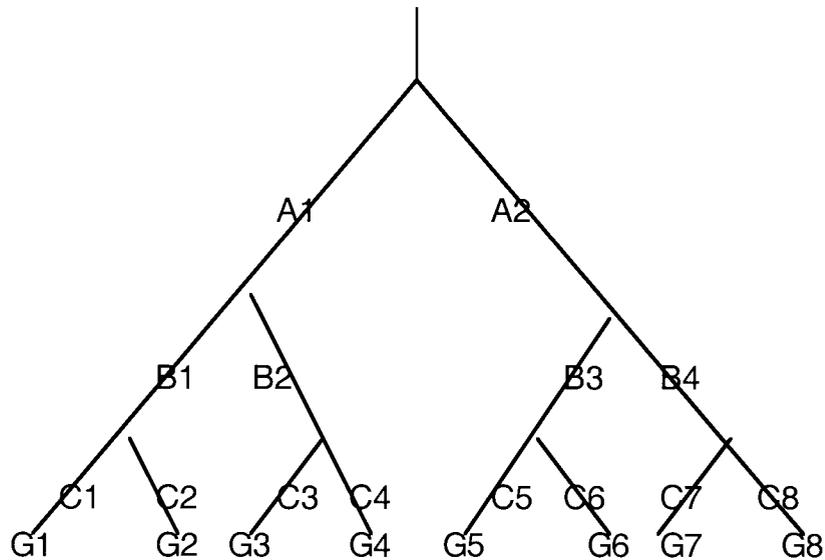


Figure 6.1. Features of the Tunnel

The dichotomous structure of the task sets the rules of the possible and impossible target places implicitly: if the ribbon *is* (positive incidence) in A1, then it is also in either B1 or B2; If the ribbon *is not* (negative incidence) in A1, then it *is* in A2 ($\sim A1$) and also either B3 or B4 (they are branches of A2). In order to represent the logical structure of the task, coordination of operations of inclusion and exclusion is necessary. The development of inclusions of grouping (classification) enables the children to represent the task as a whole: an A consists of (includes) Bs; a B consists of (includes) Cs; a C consists of (includes) Gs. Children's choices indicate their understanding of the logical structure of the task.

The most economical way to solve the Tunnel task is to plan moves (start at A first: then, B: then, C) because of the hierarchical structure which necessitates coordination of inclusions and exclusions. Children need to group places in relation to their level, or the possibility and impossibility of locations of the target. The most economical way to find the Flamingo also requires them to group places as possible or impossible for the target, and to coordinate the places. The previous studies of this thesis showed that the children's performances in the game depended on their understanding of the game clues

(inclusion of possible goal places and exclusion of impossible goal places), and how they planned to use this understanding in their choices.

Study 4 examines children's inference-making in very different contexts: the Flamingo game and the Number and Tunnel tasks. The Flamingo game card arrangement task was also used, in order to examine children's understanding of the logical structure of the Flamingo game. The children were asked to arrange the game cards as the game rules say. The children's performance was of particular interest when they did not understand the logical structure of the game.

Two hypotheses were set:

Firstly, the operative logic of inclusion and exclusion is the cognitive basis for children's understanding of the Flamingo game. It was predicted there would be significant correlations between the Flamingo game and the three inferential tasks: the Flamingo game card arrangement task, the Number task, and the Tunnel task.

Secondly, the Number and Tunnel tasks are similar in their requirements of the understanding of inclusion and exclusion, but they differ in the complexity of their requirements. Different levels of the game play were supposed. Some children might rely on the simple strategy of inclusion, as the Number task requires. More sophisticated players might consider the coordination of inclusions and exclusions and anticipate the results of moves, as the Tunnel task requires. It was predicted that the performances on both the Number and Tunnel tasks would be better predictors on the Flamingo game performance than either alone.

6.3. Method

6.3.1. Design

The study had a within-subjects design. It aimed to examine correlations between the Flamingo game and the three inferential tasks: the Flamingo game card arrangement task, the Number task and the Tunnel task.

6.3.2. Participants

The participants were 50 children aged from 6 years 4 months to 7 years 7 months (mean age = 82.9 months, S.D. = 3.71 months), who comprised all the year 2 children in one primary school in a middle class Inner-London area.

6.3.3. Procedure

Each child worked on three different tasks and eight Flamingo games in a quiet corner of the school library. The three tasks were randomly allocated in turn, but the Flamingo game always followed the Flamingo game card arrangement task. It took 25 to 30 minutes for each child to complete all the activities.

6.3.3.1. The Flamingo game card arrangement task

The aim of the Flamingo game card arrangement task was to measure the children's understanding of the logical structure of the game. Each child was asked to read aloud the rules, which appeared on the computer screen, then to tell the experimenter the game rules. When the game rules were not properly explained, the child was asked further questions:

"If the Flamingo touches a side of the card that you have, what does the card say?" The question was asked to make sure the child could verbalize the answer, "Hot".

"If the Flamingo touches a corner of the card that you have, what does the card say?" The question was asked to make sure the child could verbalize the answer, "Warm".

"If the Flamingo is not next to the card that you have, what does the card say?" The question was asked to make sure the child could verbalize the answer, "Cold".

The child was given a set of cards: a Flamingo, 6 HOTS, 6 WARMS and 10 COLDS. The instruction was:

"Arrange the cards as the game rules say. If you need more cards, take more." No feedback was given.

6.3.3.2. The Flamingo game

The aim of the Flamingo game was to measure inferential problem solving. After the game card arrangement task, the game started.

6.3.3.3. The Number task

The aim of the Number task was to see whether the children understood inclusion and exclusion. The task consists of 2-choice and 3-choice questions. In the 2-choice Number question, the child was told:

"I am thinking of a number. It is bigger than 3, but smaller than 6. Can you tell me what it is?"

In order to see whether she or he still held the zone of multiple possibilities, the child was given negative feedback on the first answer. But, the second answer got positive or negative feedback, depending on its correctness.

In the 3-choice Number question, the child was told:

"Can we do a number game once more? I am thinking of a number. It is bigger than 4, but smaller than 8. Can you tell me what it is?"

In order to see whether she or he still held the zone of multiple possibilities, the child was given negative feedback for the first two numbers. The third number got positive or negative feedback, depending on its correctness.

6.3.3.4. The Tunnel task

The Tunnel task was used as a measure of coordination of inclusions and exclusions.

Apparatus of the Tunnel task

The apparatus of the Tunnel task was a network of branching tunnels leading to caves in one of which a toy man was hidden (dimension: 30.5 x 24 inches, or 77.5 x 61 cm; height: 2 & 3/4 inches, or 7 cm; see Figure 6.1). The "tunnel" was positioned before the child with the big tunnel at the top and the small tunnels expanding towards the bottom. The task was to determine which cave the man was in. For this purpose, the man was tied to a long thin ribbon, which ran between T and G after the toy man was hidden. Each path A, B and C was provided with a small rectangular window that the child could open to check to see if the ribbon was in the branch.

The child was told a story.

"Do you know the story of 'Snow White and seven dwarfs'. One morning, a small man, one of the seven, went down to a mine to find some gold. You can see the entrance here. He went down, down, and down. He neither turned around nor went backwards. Look carefully how the tunnel divided into two and two again. He left the ribbon all the way to the mine he was working. Oh, it's dinner time. It's your job to find him and tell him to come back home. Open the windows. You can check the ribbon inside the tunnel. Try to find out by opening as few windows as possible. Open the lid of the cave where the man is only when you are sure."

Each child did the task four times.



Figure 6.2. Pictures of the Network of Tunnels

6.3.4. Measures

6.3.4.1. The Flamingo game

The mean number of moves over the last 4 games was counted. The first 4 games were not analyzed, in order to allow these games for planning children's moves.

6.3.4.2. The Flamingo game cards arrangement task

One point was given for each of the errors HOT and WARM. The error COLD was not counted, because COLD cards in the wrong places had been put in the places for HOT or WARM, which already counted as an error if HOT or WARM cards had not been put there. The maximum possible error score was 11 (5 HOTs and 6 WARMs) and the minimum error score was 0.

6.3.4.3. The Number task

The children were categorized into three groups. The first group was those who did not understand the inclusion. They did not infer any correct number between the two boundaries. Those who were wrong for the first number(s) but correct for the later number(s) were assigned to this group because their responses were treated as correct

by chance. The second group was those who inferred the first number(s), but were not sure about the next number(s) when the first ones got negative feedback. The third group was those who clearly showed the ability to infer the multiple possibilities and hold the critical area. Those who inferred two correct numbers in the 2-choice question, and three correct numbers in the 3-choice question were assigned to this group.

6.3.4.4. The Tunnel task

Each child's performance on the last two trials of the task was categorized into one of the four developmental stages based on the Piagetian criteria. The first two trials of the task were not counted, in order to allow each child to have trials of adjusting to and understanding the task demands.

The characteristics of the four categories are as follows:

Stage 1A (empirical)

For the children who randomly tried only levels C or G, their choices of Cs were perhaps guided by the short perceptual distance of Cs from Gs. Making links between C and G was not treated as an inclusion, because one C corresponds to one G. These children did not understand the inclusion. Here is an example from Stage 1A.

Timothy tried C8, C6, C5, C4, C3, C2, and C1. He found the ribbon at C1 and concluded G7.

Stage 1B (partial inclusion)

Those who started with C, but tried moves up and down across levels, were assigned to the partial inclusion group. Those who started with Bs, and those who started with As but skipped one level, were also assigned to this group. Their crossing of levels was regarded as showing a sign of knowledge of inclusion, even though the inclusions were not completed to link the whole system together. Here is an example of Stage 1B.

Amelia began from B1 and went to B3. As they were empty, she opened A2, where she saw the ribbon. Skipping the Bs, she deduced C7 and C8 and found the man in G8.

Stage 2 (concept of part-whole) and Stage 3 (concept of negative necessity)

Those who started with A without skipping levels or turning back, went to B, then to C, and concluded at a G were assigned to these groups. These children were regarded as being linked subgroups of one group. Stages 2 and 3 differed in responses after noticing positions where there was no ribbon. Those who moved to the other side of the same level by noticing positions where there was no ribbon were assigned to Stage 2. Those who proceeded to the lower level on the other side of the branch were assigned to Stage 3. Here is an example of Level 2.

Luca didn't see the ribbon at A1, then tried A2. He proceeded B4, C7 and C8 and concluded at G8.

The second judge categorized each child's response into the classifications independently.

6.4. Results

The results comprise five sections. The first four sections describe univariate statistics for each of the tasks: the arrangement of the Flamingo game information cards; the findings of the Number tasks; and the findings of the Tunnel task. The fifth section reports on the associations between the performance in the Flamingo game, the arrangement of the Flamingo game information cards, the Number task, and the Tunnel task.

6.4.1. The Flamingo game

The mean number of moves in the last 4 game was 8.28 (SD 3.71). Of the 50 children, 24 children (48%) made fewer moves than the criterion of chance level success of 7.07, with 95 % credibility.

6.4.2. The Flamingo game card arrangement

In order to know how the child understood the game structure, the arrangement of the game cards (a Flamingo card, 5 HOTs, 6 WARMs and 10 COLDs) was observed. Only 7 participants had no trouble arranging the cards as the game rules say (the

arrangement group). The other 43 participants had trouble doing so (the non-arrangement group). Most of the children were well aware that COLDs were not adjacent to the goal. Only two children arranged COLDs next to the Flamingo card. Most of the children arranged HOTs next to the Flamingo card. They put HOTs in all the places adjacent to the goal, not only at its sides but also at corners. They arranged WARMs non-adjacent to the goal. These children were not aware that WARM should be diagonally adjacent to the goal. It was clear that most of the children did not understand completely how the game was constructed. The arrangement errors were scored according to the number of HOT and WARM cards in the wrong places. The maximum possible error score was 11 (5 HOTs and 6 WARMs) and the minimum error score was 0. The mean error score was 6.06 (SD 3.14) for the whole group.

6.4.3. The Number task

In order to see whether children inferred and held the 'critical area' of multiple possibilities, they were categorized into three groups.

(A): those who did not infer any correct numbers, and those were incorrect for the first number, but correct for later number(s), were assigned to this group. These children did not understand inclusion.

(B): those who chose the first number within the critical zone but were incorrect for the later number(s) in either questions (or both questions) were assigned to this group. These children had trouble holding the 'critical area' of multiple possibilities.

(C): those children who only gave possible numbers in both 2- and 3-choice questions. These children, who clearly understood the inclusion of multiple possibilities, were assigned to this group.

Table 6.1. Number of participants across choice categories

Category	A	B	C	Total
Number of Participants	8	12	30	50

Of the 50 children, 7 children were incorrect for any numbers. Only one child was incorrect for the first numbers but correct for the last number in the 3-choice question. This response was treated as correct by chance, because if he had understood the inclusion, his first choice could not have been outside the critical zone while his later number was within the zone. This child was assigned to group A, that did not infer any correct numbers. There were 12 children who were correct for the first number(s), but incorrect for the later number(s) in either 2- or 3-choice questions (or in both questions). There were 30 participants who were correct for all the numbers in both the 2- and 3-choice questions. There was a high correlation between the 2- and 3-choice questions for those who inferred all the correct numbers (Spearman's $\rho = .684$, $p < 0.01$).

6.4.4. The Tunnel task

This section of the results reports the categorization of children's responses to the Piagetian stages of the operative logic of inclusions and exclusions, and the agreement in the categorization between two raters.

6.4.4.1. The Piagetian developmental stages

General observations about performance in this task were collected initially by examining the starting points, the order of levels of moves and children's responses after discovering the absence of the ribbon were used. Firstly, the frequency of the starting points was 16, 16, 14 and 4 children respectively for each of the levels A, B, C and G. If the child had the whole-part concept and represented the Tunnel structure as a whole, her or his starting point would be on level A. Starting at A was necessary for an economic search in a structure that linked three hierarchical levels. If the child did not realize the relations between the levels, her or his moves would have been guided by the perceptual distances from Gs. The child would start with C or G, because they were near to a target place. If the child partially represented the Tunnel structure, her or his responses were somewhere between the cases described earlier. Her or his starting level might be anywhere and the consequent moves would not be systematic. Thus, the later moves were analyzed in relation to the starting levels.

Secondly, the majority of children starting with level A went down to level B and then to level C without turning up and down across levels (12 children). These children considered the network in its entirety and constituting an operative grouping. They

were at either Stage 2 or 3. If the child understood that if it is not A1, then it is $\sim A1$, that is the concept of negation, she or he would go one level down on the other side, without necessarily checking the windows at the same level. This is the definition of Stage 3, characterized as negative necessity. Among those who started with level A and went down without turning up and down across levels, none avoided the branches at the same level if there was no ribbon. Thus, no participant was identified as at Stage 3 and the 12 participants were at Stage 2 (concept of whole - part).

There were 16 children who started at level B. These children showed the ability to include, but their inclusions were not coordinated systematically. Two patterns clearly emerged in the responses starting at level B. One pattern of responses was made by those who moved down without opening the windows at level A (9 children). They started with B and then went straight down to C. The other was of those who started with B and went up to A, to B, and then to C (7 children). Along with the 5 children who started with C but went up and down in their later responses and the 4 children who started at level A and skipped one level or turning up and down across levels, those who started at level B were assigned to Stage 1B. There were 25 children at Stage 1B, which was characterized as partial inclusion.

The majority of responses starting with C tried neither B nor A (9 children). Along with the 4 children who randomly tried Gs, these children were categorized as Stage 1A, which was characterized as empirical searching without inclusions and exclusions. There were 13 participants were at Stage 1A (empirical).

6.4.4.2. Reliability of the measurement

To test the reliability of categorization of the Piagetian developmental stages, two raters' categorizations were correlated. The agreement between the two raters was 96%.

6.4.5. Relations between the game and the tasks

It was predicted that there were specific and significant correlations between children's Flamingo game performances and those on the three tasks. Three sets of analyses were carried out. First, correlations across the Flamingo game performances and those on the three tasks (the Flamingo game card arrangement task, the Number task, and the Tunnel task) were obtained. Because distributions of the mean number of the last 4 games of the Flamingo game (Appendix 13) and the error scores of the card arrangement task (Appendix 14) were non-parametric, Spearman correlation tests were carried out. Table 6.2 shows indices of correlation between the game and the tasks.

Correlation coefficients of .426, -.296, -.299, and .282 were found between the Flamingo game and the game card arrangement task, the game and the Number task, the game and the Tunnel task, and the Tunnel task and the Number task, respectively. These correlations were all significant. Those who had high error scores for the game card arrangement had high mean numbers of moves in the game; those who scored low in arrangement errors had low mean numbers of moves. Those who inferred and held the multiple possibilities in the Number task performed well in the Flamingo game; those who did not infer the multiple possibilities performed poorly. Those who coordinated the whole structure of the Tunnel had good performances on the game: those who did not represent the structure had poor performances on the game. Those who inferred the multiple possibilities and held them in the Number task were those who coordinated inclusions and exclusions.

Table 6.2. Correlation coefficients between the Flamingo game and the tasks

	1	2	3	4
1 Flamingo game (number of moves)	1			
2 Flamingo card arrangement (error score)	.426*	1		
3 Number task (correct number category)	-.296*	-.089	1	
4 Tunnel task (developmental stage)	-.299*	-.222	.282*	1

*P < 0.05

The second analysis was a series of fixed order multiple regressions, to see whether the children's mean number of moves in the last four Flamingo games was predicted from performances of the three tasks after controlling for age. The reason for carrying out the regressions was that age might have explained the correlation between the game and the tasks. Thus, it was necessary to control for age. The first step variable was the child's age. The last (second) step variable was the performances on the Flamingo card arrangement, the Number task or the Tunnel task. Table 6.3 shows results for three fixed order multiple regressions, with mean numbers of moves in the last 4 Flamingo games as the outcome variable.

Performance on the three tasks significantly related to children's Flamingo game performance after controlling for age. Thus it can be said that the correlations between the Flamingo game and the tasks were not explained by the child's age. So the answer to the question that this section examined was that there was an operative logic of inclusions and exclusions which enabled children to perform the inferential game and tasks. In summary, the results support the idea that children's operative logic is significantly related to children's making inferences and use of them.

Table 6.3. Results for three fixed order regressions with mean number of moves of the last 4 Flamingo games as the outcome variable

	B	SE B	beta	R ² change	F change
Step					
1. AGE	-.075	.138	-.076		
2. CARD ARRANGEMENT	.410	.150	.377	.137	7.43**

** p < .01

	B	SE B	beta	R ² change	F change
Step					
1. AGE	.001	.144	.001		
2. NUMBER TASK	-1.44	.688	-.295	.085	4.37*

*p < .05

	B	SE B	beta	R ² change	F change
Step					
1. AGE	.001	.144	.001		
2. TUNNEL STAGE	-1.59	.722	-.306	.094	4.86*

*p < .05

The third set of analyses was to investigate how far the Flamingo game performance was predicted by performances on both the Number and Tunnel tasks. It was hypothesized that the Flamingo game performance could be predicted better by including both performances on the simple Number inclusion task and the more complicated Tunnel task in a regression equation than by either alone. The hypothesis was tested by stepwise regressions, with mean numbers of moves in the last 4 Flamingo games as the outcome variable. The first step was the child's age. The second step was performance on either the Number task or the Tunnel task. The third step was performance on either the Tunnel task or the Number task, depending on what was the second step. There was a significant correlation between performances on the Flamingo game and on the Number task (R^2 change = .094, $p < .05$) when the Number task performance was put in the regression equation before the Tunnel task performance. When the Number task factor was removed, the Tunnel task stage did not significantly predict the performance on the Flamingo game. There was also a significant correlation between performances on the Flamingo game and on the Tunnel task (R^2 change = .083, $p < .05$) when the Tunnel task performance was put in the equation before the Number task performance. When the Tunnel task factor was removed, the Number task performance did not significantly predict the performance in the Flamingo game. These results come from the high correlation between the Number and Tunnel tasks.

6.5. Summary of the results

In the Flamingo game card arrangement,

1. The majority of children had trouble arranging the game cards as the game rules require. However, there were some indications that children understood the game structure partially; that is, most of the children did not arrange HOT in the non-adjacent places or COLD in the places adjacent to the goal.

In the Number tasks,

2. In the 2- and 3-choice Number tasks, the majority of participants inferred all correct numbers. They did not have trouble holding a 'critical zone' of possibilities in the numeral contexts.

3. The correct responses were highly correlated ($r=.684$, $p < 0.01$) between the 2- and 3-choice Number tasks.

In the Tunnel task,

4. The development of the majority of participants was characterized by partial inclusion. The children in this study had trouble linking subclasses to form a higher structure.

Between the Flamingo game and the tasks,

5. There was a significant correlation between the understanding of the game structure and the game performance.
6. There was an association between performances on the Flamingo game and the Number task, even after control for age.
7. There was an association between performances on the Flamingo game and the Tunnel task, after control for age.
8. There was an association between performances on the Number task and the Tunnel task, after control for age.

6.6. Discussions and conclusions

This study examined young children's operative logic across the Flamingo game and three tasks: the Flamingo game card arrangement, the Number task and the Tunnel task. The Flamingo game card arrangement task was used to investigate the children's understanding of the logical structure of the Flamingo game. The majority of participants was not able to arrange the game cards exactly as the game rules require. There were some indications that children partially understood the game structure. Most of the children arranged HOTs next to the goal, but not COLDs. 17 children who had trouble arranging the game cards still managed to complete the game with fewer moves than the criterion of success by chance. They inferred the goal place from each of the word clues one a time, understanding the game structure only partially. The word clues worked as a medium to make inferences for those who had not yet constructed the whole logical structure of the game. The significant correlation between the arrangement of the game cards and the game performance suggests an association between understanding the logical structure of the game and the inferential problem-solving performance. It would have been interesting to see how the practice of the game would alter the child's understanding of the logical structure of the games, if data had allowed the comparison of performances on the game card arrangement task after as well as before the practice of the games.

The Number task examined the child's understanding of the inclusion and exclusion rules: including multiple possibilities, while excluding impossibilities. The majority of children did not have any trouble inferring and holding multiple possibilities in numerical contexts. Receiving negative feedback for the first choice but remaining within the 'critical zone' for the second number(s) cannot be explained by the use of the

simple strategies of 'win-stay' and 'lose-shift'. Thus, the result that the majority of children inferred multiple possibility supports the cognitive approach, where children's choice behaviour is viewed in relation to the development of logical understanding.

In the Tunnel task, children's anticipatory actions were examined through their starting points and consequent moves across the three levels, and their responses after discovering the ribbon. The representation of the whole structure is crucial for the Tunnel task, because it consists of a hierarchical order of subclasses. It was found that the starting point was associated with subsequent choice actions. The majority of responses starting at level A went down to level B and then to level C without moving up and down across levels, showing the ability to classify groupings. The majority of responses starting at level C did not go to other levels, showing the inability to represent the whole connected structure with the subclasses. The finding that there was a significant relation between the Tunnel task and the Flamingo game seems to suggest that the coordination of inclusions and exclusions enables the child to perform across the Flamingo game and the Tunnel task with the same level of efficiency. Even though the majority of the children aged 6 years in this study knew how to include and exclude with numbers, they could not use the knowledge to synthesize subclasses to a higher form in the spatial context of the Tunnel task. It was beyond their ability to relate these subclasses to form a whole structure. These results are harmonious with Piaget's claim that the initial relations (inclusion and exclusion), first separately and then through combinations, serve to constitute fragments of structures that progressively become coordinated until "groupings" are formed, beginning at about 7-8 years (Piaget & Garcia, 1991). The significance of these findings is that, although the game is not a test, children still seem to use logical reasoning as observed in very different tasks (the Number and Tunnel tasks) in a social context (a computer game).

There were significant correlations between the Flamingo game, the game card arrangement task, the Number task and the Tunnel task even after controlling for age.

Thus, the child's performance of the Flamingo game could be predicted by the performance on the Flamingo game card arrangement task, the Tunnel task, or the Number task. Those who represented the logical structure of the game, those who inferred and held multiple possibilities, and those who coordinated inclusions and exclusions managed the game with high efficiency. I am confident in saying that there was an operative logic of inclusions and exclusions that enabled children to make inferences and use them across the game and the tasks. The findings from the Number and Tunnel tasks suggest that the children's performance on the Flamingo game developed from use of a clue (mostly HOT) to coordination of clues. At the advanced level, children represented more than one possible choice and chose the one that would result in the maximum gain. At this level, it could be said that children used gain maximizing strategies. They anticipated the results of their choices and planned their moves to find the target in the most efficient way.

Chapter 7. Conclusions and general discussion

The aim of the thesis was to develop ways to measure and analyze children's thinking and learning in computer game activities. An inferential problem-solving game called 'Find the Flamingo', and a simplified version, 'the Boat game', were used for the research, because most computer games have an element of problem solving that is, as Piaget and Garcia (1991) have argued, inferential in nature. The games consisted of sets of rules, given with affirmative and negative *if-then* statements. Four experiments were carried out, each with a specific design. In this chapter, I will summarize the findings of each of the experiments in relation to the research questions. Limitations of the study and educational implications for children's computer game play are also discussed.

7.1. The findings in relation to the research questions

There were four main research questions to be answered, related to children's thinking and learning in computer game contexts. They concerned the measurement of learning outcomes, decision-making in the context of games, children's construction of the meaning of the word clues, and ways of enhancing children's problem solving.

How are the learning outcomes measured?

One of the main research aims was to explore how psychological knowledge would help to develop ways to analyze children's thinking and learning in computer game contexts. Observation of the game performance and direct questions about how children made inferences from the game clues were the methods used for this. Firstly, the child's number of moves was observed over the practice of the games. In order to test whether the child used the game rules, the criterion of success at chance level was set. What number of moves would be needed to find the target by chance? This was calculated as approximately the mean number of the places in the game array. If the child used the game rules, she or he would not need as many moves as the mean. In order to decide what number of moves could be confidently judged as showing success beyond the chance level, the criterion of success at chance level was set by using a one-tail t-test with the mean, 13, (because there were 25 places), and the number of games for the degree of freedom. If the mean number of moves was below or the same as the criterion of 7.07, it could be confidently said there was a bias – that the game rules were being used.

Secondly, use of each clue given in the course of the game was examined. If the players made inferences from the game rules, their choices would be coherent with inferences from the information that they had. For example, faced with HOT in the Flamingo game, their choice would be one of the adjacent 'up', 'down', 'left' or 'right' places. Faced with WARM, they would move to one of the adjacent diagonal places. Faced with COLD, the next move would be to one of the non-adjacent places. If the players did not have any concept of the game rules, their choices would not correspond to how the game rules define the relations between the goal and the information. Thus, the focus of analyses of the game behaviour should be on choices (moves) in relation to information. Each move can be seen as a result of either a correct or an incorrect

inference from the previous information. To assess whether a move follows from a correct inference, two methods of analysis were applied. The first method was to assess 'single information inference', in which the immediately preceding information counts as the base of the inference. The other method was to assess 'double information inference', in which the two preceding items of information count as the base for inference. However, these methods did not include the cases when the children might have gone outside the zone of possible goal places, not because they did not infer from the rules, but because they tried to gather more information with more advanced strategies. Despite this weakness, the methods showed concerning how far the young children made links between the clues and the game rules.

In addition to observation of the game performance, questions on the game clues were also asked in a test context. Given one or two clues at a time, the child was asked to make inferences about the possible and impossible locations of the target. Scores for the tests would inform us of the children's ability to make inferences from the game rules, whereas the number of moves or the percentages of correct inferences would inform how the inferences were actually used to guide their choices in the process of the game. Would there be an association between what they said and what they did?

The results of Studies 1 and 2 showed that the mean number of moves in the games significantly correlated with the scores for the word clue tests. There was also high correlation of the mean numbers of moves between the even and odd sets of games. Thus, counting the number of moves until the completion of the game is a reliable measurement of the child's inferential problem solving in the game process. While the number of moves measured the child's inferential problem solving in the game, studies of single and double information inferences identified what concepts and operations children had learned, and how they integrated information. The study of single and double information inferences is supported by Piaget, who asserts that an action is evaluated in terms of effectiveness or usefulness in relation to a goal (Piaget & Garcia,

1991, p.4.). For analysis of children's goal-searching actions, their understanding of multiple possibilities and impossibilities and use of the understanding were examined through the methods of counting the number of moves and of single and double information inferences, and the separate tests on inference from clues.

How do children make decisions in computer game play?

From the very beginning of the research, two kinds of players were envisaged: those who saw the game as a task to tackle with logic, and those who perceived it as a random event, not being aware of the problem-solving nature of the game. The consequent decision-making processes were that those who saw the game as a task would try to use the game rules in their search for the target, whereas those who saw the game as a random event would make as many moves as the number of moves expected to lead to success from random choices. Each child's mean number of moves over the practice of the games was counted and compared to the criterion. Study 1 showed that the majority of 8-year-old children found the Flamingo with a smaller number of moves than the criterion of success at chance level. The correlation between the number of moves and the scores for the tests that measured the child's ability to make inferences from the game clues was highly significant. It was evident that these children did indeed represent the game as problem-solving and infer the goal place from the game rules. Thus, the results support the cognitive decision-making theory, according to which children represent the problem and calculate the outcomes of moves. The cognitive theory explains the children's game behaviour better than the probability matching theory, in which children merely reflect the likelihoods of happenings. There were too many places that the moves could be made to in the Flamingo game to explain the children's choices as a reflection on the likelihood of the target.

Developmental changes in inferential problem solving were shown in Study 2. Most of the 10-year-old children made fewer moves than the criterion for success at chance level, whereas on average the 7-year-old children did not do so. On average, it took as many moves as working by chance would need for the younger children to find the Flamingo. The results of the word clue tests also showed that the younger children usually made poorer judgements about the possibilities and impossibilities of the goal place than the older children. Within the same age groups, differences in performance were shown as consistent throughout the games. Backward learning curves, rather than forward learning curves, were used to trace the performances of the groups in order to find out whether those who performed well in the final games differed from the beginning of the games from those who performed poorly after the practice of the game. The numbers of moves for those whose number of moves of the last four games was above the criterion (the 'poor' performance group) and of those whose number of moves was below the criterion (the 'good' performance group) were traced from the first to the last games. It was found that the 'poor' performance group started the game poorly, while the 'good' performance group did better from the start.

Why was the 'good' performance group able to make and use the inferences whereas the 'poor' performance was not able to do so? It was assumed that there was an operative logic of inclusions and exclusions that enabled the more successful child to represent the game structure and to anticipate the results of the moves. Study 4 investigated children's inferential problem solving across the Flamingo game and two other tasks: the Number task and the Tunnel task. The Number task was used to investigate children's understanding of inclusion of multiple possibilities, that is, their understanding of exclusion of multiple impossibilities as well. The Tunnel task was used to assess the level of the coordination of inclusions and exclusions. The Flamingo game and the two tasks were similar in that they all required inclusions and exclusions. But they differed in the complexity of their requirements. The majority of 6-year-old

children could infer and hold the 'critical zone' of possibilities; their coordination of inclusions and exclusions could be judged as being only partially developed. There were significant correlations between the results for the Flamingo game, the Number task, and the Tunnel task after age had been controlled for. Those who could keep in mind multiple possibilities and those who coordinated inclusions and exclusions managed the game with high efficiency. The children's performance on the Flamingo game could be predicted by their performance of the Number and the Tunnel tasks. The results indicate that there are developmental stages of the operative logic of inclusions and exclusions. Children develop in their ability to understand logical structure from a simple form (i.e., the Number task) to a more complicated form (i.e. the Tunnel task) which requires coordination of inclusions and exclusions.

In conclusion, individual differences in decision-making in computer game contexts are clearly related to inferential ability. 8-year-old children, on average, could enact the game as a problem-solving task when the game is structured by rules, stated as affirmative and negative *if-then* sentences. The children younger than 8 years made random choices and showed poor judgement on the possible or impossible places for the target. The ability to understand inferential principles and to use this understanding in planning the search actions developed with age. These results are consistent with a variety of others that examined the understanding of logical relations in children of roughly the same age range (6 to 10 years) that I have examined in the experiments for this thesis (Bynes & Overton, 1983; Markovits, et al., 1989; Pieraut-LeBonniec, 1980; Scholnick & Wing, 1988). The children's game performance was successfully predicted by examining their performances on other inferential tasks that required inferring multiple possibilities or coordination of inclusions and exclusions. I can confidently say that there is an operative logic of inclusions and exclusions that enables these children to make and use inferences in their search. General skills of inclusion and exclusion were used to solve specific tasks that differed in their structures. The

general skills and their specific uses will be discussed in more detail in the next section concerning children's construction of the meanings of game clues.

How do children construct the new meanings of the words?

In the Flamingo game process, the children were faced with the clues: HOT, WARM, or COLD. The meanings of the clues in the game were no longer related to temperature. The efficiency of children's search for the target depended on the construction of new meanings for the words as they were defined in the game rules. The question of how the children interpreted the clues was tackled by an examination of choice making in relation to each of the game clues. Study 2, with 87 children aged 7 to 10 years, showed that, faced with HOT, about 90 percent of moves were to the horizontally or vertically adjacent places. Faced with WARM, there were only one out of each two moves to the diagonal places. Faced with COLD, about 65 percent of moves were to the non-adjacent places. HOT and WARM had the same meaning – the inclusion of possibilities. Why were HOT and WARM used so differently? In the Flamingo game, spatial information appears to confuse children about the two kinds of inclusion: the inclusion of horizontally or vertically adjacent places (HOT) and that of diagonally adjacent places (WARM). A directional difference (side vs. corner) may explain the greater successful use of HOT than WARM. Analysis of the moves showed that, faced with WARM, moves to the horizontally or vertically adjacent places were made nearly as often as to the diagonal places. But moves to the non-adjacent areas were quite rare. The children regarded WARM as the signal for moves to adjacent places. Were they confused by the directional cue for inclusion and exclusion? Were they incapable of keeping in mind simultaneously the two dimensions of the information: direction, and inclusion of multiple possibilities?

In order to investigate how the directional cues affected the children's inferential problem solving, a new game was designed to separate the inclusion information from the directional cues. Two rules and two kinds of information (inclusion and exclusion) structure the game called the 'Boat' game. The game has only one kind of inclusion information, without directional clues; that is, 'touching'. Study 3 showed that the children in the control group (without the planning intervention, N=32, mean age = 7.4 years) had a mean number of 10 moves in the first 8 games in the Boat game. The percentages of correct inferences were 81.1 for inclusion information (X) and 44.9 for exclusion information (O). Study 2 showed that the same age group (N=32, mean age = 7.4 years) had a mean number of 9.9 moves in the 8 games in the Flamingo game. The percentages of correct inferences were 88.5 for inclusion information about the horizontally or vertically adjacent places (HOT) and 43.8 for exclusion information (COLD). The elimination of directional cues did not affect inferential problem solving performance. This finding implies that inclusion and exclusion cues are more salient for drawing inferences than directional cues.

In conclusion, there were differences in the use of different kinds of information. The children found it easy to use HOT in the Flamingo game and X in the Boat game, taking them as signals of inclusion of possible locations of the goal. Exclusion information (COLD in the Flamingo game and O in the Boat game) was used much less correctly than inclusion information. Thus, the construction of the new meanings of the word clues in a goal search frame varies according to the demand for operations of inclusion and exclusion, and the complexity of the cues in the words. The directional cue disturbed the child's inference making when they were using inclusion and exclusion.

Can children's inferential problem solving be enhanced in computer games?

In order to see whether the children's inferential problem solving improved over the practice of games, the mean numbers of moves were compared between the blocks of games. Studies 2 and 3 found significant differences, in favour of the later games. Did children's understanding of possibilities and impossibilities and use of their understanding change? Or did the accumulated reinforcement lead the children to make choices in the way defined by the game rules? It was assumed that if children's understanding of possibilities and impossibilities and use of their understanding changed over the practice of games, intervention could be given to enhance inference-making from the game rules and its use in the game process. Study 3 investigated the effects of 'guided-planning' and 'pausing' in the Boat game, using a brief training session. In the guided-planning condition, the experimenter asked questions about the possibility and impossibility of the goal in relation to the clues; in the pause condition, 'pauses' were inserted into the game as a deliberate opportunity for children to think how to use inferences. During the training, the guided-planning group performed significantly better than the control group. The planning of use of inferences did make differences in inferential problem solving. Analyses of inferences from each type of information provided a precise picture of the effect of intervention in planning. In the case of inclusion inferences, there was no difference between the treatment groups. The percentages of inclusion inferences for all the participants increased from the pre- to the post-training games. In the case of exclusion inferences, the guided-planning group was significantly better at exclusion inference than the pausing group during the training; the pausing group was significantly better than the control group. But the difference disappeared in the post-training performance. The children aged 6 to 8 years gave interpretations to the information of O and COLD (assimilation), but the

knowledge was not truly appropriated because it had not been internally reconstructed (accommodation).

If the games provide children with plenty of opportunities to assimilate, children will learn through doing them. The selection of games, then, is better when the game exceeds a little the child's level of understanding of the concepts and logical structure of the game. The construction of the meaning of the game clues and the planning of the use of inferences will gradually improve over the practice of the game.

7.2. Limitations and suggestions for further studies

The 'Find the Flamingo' game (1985) used in the thesis is by now a rather simple and old-fashioned game. However, it requires the understanding of game rules stated as positive and negative, the two basic forms of statements. The inferential nature of the rules provided a research opportunity to investigate children's reasoning and problem solving in the context of games. The fact that the Flamingo game has an element of problem solving, like most computer games, including games in specific subject areas, suggests that findings from studies using this game can be extended to other games and children's thinking in general.

The finding that children learned concepts of inclusion and exclusion within the simple structure of the game gives an encouraging implication, that individual teachers might create simple games for teaching aids, for example, using hyper-text. Kelly and O'Kelly (1994) also suggested the use of teacher-designed instructional games in the classroom, emphasizing the need for simplicity in the games. Further studies should investigate how teachers might create their own games (involving children), what can be gained from these activities, and what training is needed to enable teachers to create their own games. In particular, research interest lies in the implicit and explicit knowledge gained in computer contexts. The comparative effects of implicit guidance (the pausing group) and explicit guidance (the guided planning group) in children's planning were shown in Study 3. Children's performance benefited from planning the use of inference from the game rules in relation to the current position in the game, by the experimenter's explicit questioning about the possible or impossible target places. How would children make explicit their implicit knowledge when designing new games? How would a teacher encourage them to do so?

7.3. Educational implications

Children's planning was shown as "planning in action" instead of "planning the action" in Studies 3 and 4. The intervention of an adult's questioning before action enhanced the use of inferences. This finding has implications for the design and use of educational game software. The first point to be made concerns its design. Deliberate questions should be posed in the course of games in order to encourage children to think and to plan. The approach which provides instructional control by CAI - Computer Assisted Instruction - favours provision of explanations or a guideline, for example, a help facility, in contrast to the approach which provides learner-controlled support, where more choices are left to the learner. The finding of the positive effect of guided planning supports the case for direct instructional control. In order to enable pupils to take greater responsibility for their own learning in computer activities (DES, 1990), software should provide and promote more opportunities to think and plan. The second point concerns the roles of the teachers in the use of educational software. Should teachers intervene in the course of children's computer activity, or is it sufficient for them to evaluate and choose a piece of software and leave the resources in the children's hands? Squires and McDougall (1994) pointed out that the vast majority of educational software packages is used in organized contexts in which a teacher has a formal responsibility to participate. Supporting Squires and his coworkers' claim that software evaluation should be made, using the paradigm of teacher-children interaction (McDougall & Squires, 1995; McDougall, Squires & Guss, 1996), this thesis draws the attention to the value of interaction between the pupil and the teacher in the context of computer games.

Individual differences and learning through problem solving in computer game contexts were demonstrated in the thesis. Two levels of analysis were involved. One concerned

the general problem-solving process for different age groups. The other level was the study of intrinsic individual differences, reflecting a fine-grained analysis of the psychological processes in sub-groups of children. Using the backward learning curves of two performance groups in each age group, Study 2 showed that some 7-year-old children played the game with higher efficiency than some 10-year-old children. Many younger children used the game rules only after the first four games, whereas most older children performed the game with high efficiency from the beginning of the games. The grouping offered a means of exploring the diversity of the problem-solving processes of children in a way that would not have been feasible with a reliance on a single mean for the whole group.

An impressive observation for the experimenter was that all the children were very eager to play the game on the computer. Regardless of whether or not a child managed to play the game properly, she or he reported high satisfaction and joy. Nonetheless, Study 1 showed that the apparatus used when the children played the game did not have a significant effect on the children's problem solving efficiency. There was no difference in this between the computer and the traditional game tools (the board and cards). Stronger motivation, or eagerness to play the game, did not lead to improved problem solving. This finding does not support some claims for the use of technology in education, that is, the positive effects of the strong motivation that computers offer (Cox, 1997; Silvern, 1986).

There have been studies showing the negative effects of computer use on children's learning. For example, Oyen and Bebko (1996) compared memory-enhancing strategies in computer games and lesson contexts. Children aged from 4 to 7 years took part in one of two computer games ("GrowWorm" and "ShipWrecks") and were taught in corresponding more formal lesson conditions (where they were simply given instructions to remember). During the stimulus presentation and delay periods, children were observed for strategy use. The observed rehearsal was much greater in game

contexts. At each age level the number of rehearsers nearly doubled. However, when covert use of rehearsal was also considered through the inclusion of children's reports of strategy use, there was no such effect. The game condition, while more enjoyable and interesting for the children, was more difficult for them to recall than the lesson. Oyen and Bebko speculated that the game contexts required the processing of multiple means and goals, whereas lesson tasks were less complex. Some of the elements of the games that were reported as interesting and fun by the children, such as the capability of dragging images across the screen or the movement of the stimulus items as they fell to the bottom of the display, may have distracted some of the children or become subgoals in themselves, supplementing the goal of the computer game. The results point to the need for careful consideration of specific aspects of the design of the software with reference to the purpose of its use. Along with the result reported here, the finding that the effects of computer activities did not come from the computer itself, but from the activity, Oyen and Bebko's work draws attention to the importance of care in the design and use of educational game software.

Another merit of the 'Find the Flamingo' game for children's learning was its gender-neutral nature (Littleton, et al., 1992; Littleton, Light, et al., 1998; Littleton, Ashman, et al., 1999). Girls and boys did not show any differences in their inferential game performance. This finding is harmonious with studies showing that training on spatial skills benefits both males and females (meta-analysis of Baeninger & Newcombe, 1989; a video action game study by Subrahmanyam & Greefield, 1994).

Overall, the thesis developed a frame for analyzing children's thinking and learning in computer game contexts. Four experiments investigated and discussed the measurement of learning, individual differences in inferential game performance, the construction of the meaning of the clues in the games, and the effects of intervention on planning. The approach is supported by the view that researchers must measure the promises of providers of high technology systems against our general knowledge of

how children think and learn (Littleton & Light, 1999; Squires & McDougall, 1994; Underwood & Brown, 1997; Wood, 1998). Educators should look carefully at games if they intend to make them a central part of the learning environment. More attention should be given to each piece of educational software, with careful consideration of why a specific package would be useful, what activities the child does with it and how, and what concepts are learned through it.

References

- Acredolo, C., & Horobin, K. (1987). Development and relational reasoning and avoidance of premature closure. *Developmental Psychology, 23*, 13-21.
- Anderson, D. R. (1972). The effects of prior training on the incidental discriminative learning of children. *Journal of Experimental Child Psychology, 14*, 416-426.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.
- Anooshian, L. J., Hartman, S. R., & Scharf, J. S. (1982). Determinants of young children's search strategies in a large-scale environment. *Developmental Psychology, 18*, 608-616.
- Bartlett, F. C. (1932). *Remembering*. Cambridge, Eng.: Cambridge University Press.
- Benzanilla, M. J., & Ogborn, J. (1992). Logical sentences and searches. *Journal of Computer Assisted Learning, 8*, 2, 37-48.
- Bjorklund, D. E. (1985). The role of conceptual knowledge in the development of organization in children's memory. In C. J. Brainerd & M. Pressley (eds.), *Basic processes in memory development: Progress in cognitive development research*. New York: Springer-Verlag.
- Blank, M. (1968). Experimental approaches to concept development in young children. In E. A. Lunzer & J. F. Morris (eds.), *Development in Human Learning*. Staples Press.
- Blaye, A., Light, P., Joiner, R., & Sheldon, S. (1991). Collaboration as a facilitator of planning and problem-solving on a computer based task. *British Journal of Developmental Psychology, 9*, 471-83.
- Bobko, P., Bobko, D. J., & Davis, M. A. (1984). A multidimensional scaling of video games. *Human Factors, 26*, 477-482.
- Braine M. D. S., & Rumin, B. (1983). Logical reasoning. In P. H. Mussen (Series Ed.), J. H. Flavell, & E. M. Markman (Vol. Eds.), *Handbook of Child*

- Psychology: Vol. 3. Cognitive Development (4th ed., pp.263-340)*. New York: Wiley.
- Bredemeier, M. I., & Greenblatt, C. S. (1981). The educational effectiveness of simulation games. *Simulation & Games, 12*, 307-332.
- Brosnan, M. J. (1998). The role of psychological gender in the computer-related attitudes and attainments of primary school children (aged 6-11). *Computers and Education, 30*, 3/4, 203-208.
- Bruner, J. S. (1996). *The culture of Education*. Cambridge: Harvard University Press.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Bryant, P. E. (1974). *Perception and understanding in young children*. London: Methuen.
- Bryant, P. E. (1977). In Hewson, S. (1977), Inferential problem solving in young children. Dphil thesis. Oxford University.
- Bryant, P. E., & Roazzi, (1992). Children's understanding of equivalencies. Paper presented in the *Biannual European Developmental Psychology Conference*.
- Byrnes, J. P., & Overton, W. F. (1986). Reasoning about certainty and uncertainty in concrete, causal, and propositional contexts. *Developmental Psychology, 22*, 6, 793-799.
- Byrnes, J. P., & Overton, W. F. (1988). Reasoning about Logical Connectives: A Developmental Analysis. *Journal of Experimental Psychology, 46*, 194-218.
- Carter, J. (1984). Instructional learner feedback: A literature review with implications for software development. *The Computing teacher, October*, 53-55.
- Casey, M. B. (1976). Blocking of incidental information in young children. *Journal of Experimental Child Psychology, 22*, 292-301.
- Ceci, S. (1990). *On intelligence - more or less: a bio-ecological theories of intellectual development*. New York: Prentice Hall.
- Colman, A. (1982). *Game theory and experimental games*. Pergamon Press.

- Cox, B. C., Ornstein, P. A., Naus, M. J., Maxfield, D., & Zimler, J. (1989). Children's current use of rehearsal and organizational strategies. *Developmental Psychology, 25*, 619-27.
- Cox, M. J. (1997). *The effect of Information Technology on Students' Motivation. Final Report.* King's College London.
- Cunningham, H. (1995). Motal kombat & computer game girls. In Bazalgett & Buckingham (Eds.), *In front of the Children Screen Entertainment & Young Audiences.* British Film Institute.
- Davis, M. D. (1970). *Game theory: A nontechnical introduction.* New York: Basic Books.
- De Corte, E. (1996). Changing views of computer-supported learning environments for the acquisition of knowledge and thinking skills. In S. Vosniadou, E. De Corte, R. Glaser, & H. Mandi (Eds.), *International perspective on the design of technology-supported learning environments.* Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Dempsey, J., Lucassen, B., Gilley, W., & Rasmussen, K. (1993-1994). Since malone's theory of intrinsically motivating instruction: what's the score in the gaming literature? *Journal of Educational technology sytems, 22, 2*, 173-183.
- DES (1990). *Technology in the National Curriculum.* HMSO, London.
- Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* Hillsdale, NJ: Erlbaum
- Evans, J. St.B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction.* Lawrence Erlbaum Associates Ltd.
- Fabricius, W. V., Sophian, C., & Wellman, H. M. (1987). Young children's sensitivity to logical necessity in their inferential search behavior. *Child Development, 58*, 409-423.
- Ferguson, R. P., & Bray, N. W. (1976). Component processes of an overt rehearsal strategy in young children. *Journal of Experimental Child Psychology, 21*, 490-506.

- Flavell, J. H. (1977). *Cognitive Development*. Englewood Cliffs London: Prentice-Hall.
- French, L. A. & Nelson, K. (1985). *Young children's knowledge of relational terms: Some ifs, ors and buts*. New York: Springer-Verlag.
- Gauvin, M., & Rogoff, B. (1989). Collaborative problem solving and children's planning skills. *Developmental Psychology*, 25, 131-51.
- Gellatly, A. R. H. (1987). The Acquisition of a concept of logical necessity. *Human Development*, 30, 1, 32-47.
- Giacquinta, J., Bauer, J., & Levin, J. (1993). *Beyond technology's promise*. Cambridge, Cambridge University Press.
- Goodnow, J. J. (1955). Determinants of choice Distribution in two-choice situation. *American Journal of Psychology*, 68, 106-116.
- Greenfield, P. M. (1983). Video games and cognitive skill. In *Video games and human development: Research agenda for the '80s* (pp. 19-24). Cambridge, MA: Monroe C. Gutman Library. Graduate school of Education.
- Greenfield, P. M. (1984). *Mind and media: The effect of televeision, computer and video games*. Fontana.
- Greenfield, P. M. (1993). Representation competence in shared symbol systems: Electronic media from radio to video games. In R. R. Cocking & K. A. Renninger (Eds.), *The development and meaning of psychological distance* (pp. 161-183.). Hillside, NJ: Erlbaum.
- Greenfield, P. M., Brannon, C., & Lohr, D. (1994). Two dimensional representation of movement through three-dimensional space: The role of video game experience. *Journal of Applied Developmental Psychology*, 15, 87-103.
- Greenfield, P. M., Camaioni, L., Ercolani, P., Weiss, L., Lauber, B. A., & Perucchini, P. (1994). Cognitive socialization by computer games in two cultures: Inductive discovery or mastery of an iconic code? *Journal of Applied Developmental Psychology*, 15, 59-85.

- Griffiths, M. (1996). Computer game playing in childhood and adolescence: A review of the literature. In Gill, T. (Ed.) *Electronic Children: How children are responding to the information revolution*. London: National Children's Bureau.
- Haake, R. J., Somerville, S. C., & Wellman, H. M. (1980). Logical ability of young children in searching a large-scale environment. *Child Development*, 51, 199-1302.
- Halpern, D. E. (1986). *Sex differences in cognitive abilities*. Hillsdale, NJ: Erlbaum.
- Hampton, J. A. (1989). Negating noun concepts. Paper presented at the *Edinburgh Round-Table on the Mental Lexicon*, University of Edinburgh, Scotland.
- Hawkrige, D. (1990). Computers in third world schools: the example of China. *British Journal of Educational Technology*, 21 (1), 4-20.
- Hay, D. F., & Lockwood, R. (1989). Girls' and boys' success and strategies on a computer-generated hunting task. *British Journal of Developmental Psychology*, 7, 17-27.
- Herron, R. E., & Sutton-Smith, B. (1971). *Child's Play*. New York: Wiley.
- Hewson, S. (1978). Inferential problem solving in young children. *Developmental Psychology*, 14, 1, 93-98.
- Horobin, K., & Acredolo, C. (1989). The impact of Probability Judgements on Reasoning about Multiple Possibilities. *Child Development*, 60, 183-200.
- Hoyles, C. (1988). *Girls and computers*. London: Institute of Education, Bedford Way Papers, 34.
- Hoyles, C., & Sutheland, R. (1989). *Logo mathematics in the classroom*. London: Routledge and Kegan Paul.
- Hubbard, P. (1991). Evaluating computer games for language learning. *Simulation and Gaming*, 22, 2, 220-223.
- Hughes, M. (1990). Children's computation. In Grieve, R. & Hughes, M. (Eds.) *Understanding children*, 121-139. Oxford: Basil Blackwell.

- Hughes, M., Brackenridge, A., & MacLeod, H. (1987). Children's ideas about computers. In J. Rutkowska, & C. Crook (Eds.), *Computers, cognition and development*. Chichester: John Wiley.
- Hughes, M., Brackenridge, A., Bibby, A., & Greenhough, P. (1988). Girls, boys and turtles: Gender effects in young children learning with Logo. In S. Hoyles (Ed.), *Girls and Computers* (pp. 31 -39). London: University of London, Institute of Education.
- Hughes, M., & Greenhough, P. (1995). Feedback, adult intervention, and peer collaboration in initial logo learning. *Cognition and Instruction, 13*, 4, 525-539.
- Huttenlocher, J. (1968). Constructing spatial images: A strategy in reasoning. *Psychological Review, 75*, 286-298.
- Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.
- Ioerger, T. R. (1994). The manipulation of images to handle indeterminacy in spatial reasoning. *Cognitive Science, 18*, 551-593.
- Isaacson, R. L., Hutt, M. L., & Blum, M. L. (1965). *Psychology: The science of behavior*. Harper & Row.
- James, W. (1890). *Psychology (Briefer Course)*. Republished by Cleveland: World, 1948.
- Janssen Reiner, I., & Plomp, T. (1997). Information Technology and gender equality: a contradiction in terms? *Computers & Education, 28*, 65-78.
- Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge, Mass.: Harvard University Press.
- Johnson-Laird, P. N. (1987). The mental representation of the meaning of words. *Cognition, 25*, 189-211.
- Jolicoeur, K. & Berger, D. E. (1988). Implementing educational software and evaluating its academic effectiveness: Part 2. *Educational Technology, 28*, 10, 13-19.

- Keeney, T., Cannizzo, S. R., & Flavell, J. H. (1967). Spontaneous and induced verbal rehearsal in a recall task. *Child Development*, 953-966.
- Kelly, A. E., & O'Kelly, J. B. (1994). Extending a tradition: teacher designed computer-based game. *Journal of Computing in Childhood Education*, 5, 2, 153-166.
- Kendler, H. H., & Kendler, T. S. (1962a). Vertical and horizontal processes in problem solving. *Psychological Review*, 69, 1-16.
- Kendler, T. S. (1995). *Level of Cognitive Development*. Lawrence Erlbaum Associates Ltd.
- Kendler, T. S., & Kendler, H. H. (1959). Reversal and nonreversal shifts in kindergarten children. *Journal of Experimental Psychology*, 58, 1, 56-60.
- Kendler, T. S., & Kendler, H. H. (1960). Reversal and nonreversal shifts in nursery school children. *Journal of Comparative and Psysiological Psychology*, 53, 1, 83-88.
- Kendler, T. S., & Kendler, H. H. (1962b). Inferential behavior in children as a function of age and subgoal constancy. *Journal of Experimental Psychology*, 4, 460-466.
- Kendler, T. S., & Kendler, H. H. (1967). Experimental analysis of inferential behavior in children. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior*, Vol. 3. New York: Academic Press, pp. 157-190.
- Kendler, T. S., & Kendler, H. H. (1970). An ontogeny of optional shift behavior. *Child Development*, 41, 1-27.
- Kendler, T. S., Kendler, H. H., & Carrick, M. A. (1966). Verbal labels and inferential problem solution of children. *Child Development*, 37, 749-763.
- Kendler, T. S., Kendler, H. H., & Learnard, B. (1962). Mediated responses to size and brightness as a function of age. *American Journal of Psychology*, 75, 571-586.

- Kendler, T. S., Kendler, H. H., & Wells, S. D. (1960). Reversal and nonreversal shifts in nursery school children. *Journal of Comparative Physiology and Psychology*, *53*, 83-88.
- Kendler, T. S., & Ward, J. W. (1972). Optional reversal probability is a linear function of the log of age. *Developmental Psychology*, *7*, 337-348.
- Kerns, K. A., & Berenbaum, S. A. (1991). Sex differences in spatial ability in children. *Behaviour Genetics*, *21*, 383-396.
- Kingsley, P. R., & Hagen, J.W. (1969). Induced versus spontaneous rehearsal in short-term memory in nursery school children. *Developmental Psychology*, *1*, 40-46.
- Kingston, A., & Klein, R. (1991). Combining shape and position expectancies: Hierarchical processing and selection inhibition. *Journal of Experimental Psychology*, *17*, 2, 512-519.
- Klein, J. D., & Freitag, E. (1991). Effects of using an instructional game on motivation and performance. *Journal of Educational Research*, *84*, 5, 303-308.
- Kozma, R. B. (1991). Learning with media. *Review of Educational Research*, *61*, 179-211.
- Langley, P. (1987). A general theory of discrimination learning. In D. Klahr, P. Langley & R. Neches (Eds.), *Productive system models of learning and development*. The MIT Press.
- Levinson, B., & Reese, H. W. (1967). Patterns of discrimination learning set in preschool children, fifth-graders, college freshmen, and the aged. *Monographs of the Society for Research in Child Development*, *32* (7, Serial No. 115).
- Light, P. (1997). Computers for learning: Psychological perspectives. *Journal of Child Psychology and Psychiatry*, *38*, 5, 497, 504.
- Light, P., Blaye, A., Gilly, M., & Girotto, V. (1989). Pragmatic schemas and logical reasoning in 6- to 8-year-old children. *Cognitive Development*, *4* (1), 49-64.
- Light, P., Girotto, V., & Legrenzi, P. (1990). Children's reasoning on conditional promises and permissions. *Cognitive Development*, *5* (4), 369-383.

- Light, P. & Littleton, K. (1999) *Social processes in children's learning*. Cambridge Press.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Development*, 56, 1479-1498.
- Linn, S., & Lepper, M. (1987). Correlates of children's usage of video games and computers. *Journal of Applied Social Psychology*, 17, 72-93.
- Littleton, K. (1996). Children and computer. In D. Bancroft & R. Carr (Eds.), *Influencing Children's Development*. The open university. Blackwell.
- Littleton, K., Ashman, H., Light, P., Artis, J., R., Roberts, T., & Oosterwegel, A. (1999). Gender, task contexts, and children's performance on a computer-based task. *European Journal of Psychology of Education*, 16, 129-39.
- Littleton, K., & Light, P. (1999). *Learning with computers: Analysing productive interaction*. Rutledge. London and New York.
- Littleton, K., Light, P., Joiner, R., Messer, D., & Barnes, P. (1992). Pairing and gender effects on children's computer-based learning. *European Journal of Psychology of Education*, 7, 311-24.
- Littleton, K., Light, P., Joiner, R., Messer, D., & Barnes, P. (1998). Gender, task scenarios and children's computer-based problem solving. *Educational Psychology*, 18, 3, 327-40.
- Lowery, B. R., & Knirk, F. G. (1982-1983). Micor-computer video games and spatial visualization acquisition. *Journal of Educational Technology Systems*, 11, 155-166.
- Lycsak, R., & Tighe, T. (1975). Stimulus control in children under a blocking paradigm. *Child Development*, 46, 115-122.
- Maier, N. R. F. (1932). The effect of cerebral destruction on reasoning and learning in rats. *Journal of Comparative Neurology*, 54, 45-75.
- Maier, N. R. F. (1936). Reasoning in children. *Journal of Comparative Psychology*, 21, 357-366.

- Malone, T. W. (1980). What makes things fun to learn? A study of intrinsically motivating computer games. *Cognitive and Instructional Science series. CIS-7* (SSL-80-11). XEROX Palo Alto research center.
- Manktelow, K. I., & Over, D. E. (1990). *Inference and Understanding*. Routledge.
- Markovits, H., Fleury, M., Quinn, S., & Venet, M. (1998). The development of conditional reasoning and the structure of semantic memory. *Child Development, 69*, 3, 72-755.
- Markovits, H., Schleifer, M., & Fortier, L. (1989). Development of elementary deductive reasoning in young children. *Developmental Psychology, 25*, 5, 787-793.
- Marsall, S. P. (1995). *Schemas in problem solving*. New York: Cambridge University Press.
- Mazur, J., & Hastie, R. (1978). Learning as accumulation: A reexamination of the learning curve. *Psychological Bulletin, 85*, 1256-1274.
- McClurg, P. A., & Chaille, C. (1987). Computer games: environments for developing spatial cognition? *Journal of Educational Computing Research, 3*, 1, 95-111.
- McDougall, A., & Squires, D. (1996). An empirical study of a new paradigm for choosing educational software. *Computers and Education, 25*, 3, 93-103.
- McDougall, A., Squires, D., & Guss, S. (1996). Emphasising use over attributes in selection of educational software. *Education and Information Technologies, 1*, 151-164.
- McKinsey, J. C. C. (1952). *Introduction to the Theory of Games*. New York.: McGraw-Hill.
- Murray, F. B. (1987). Necessity: The developmental component in school mathematics. In L. S. Liben (Ed.), *Development and Learning: Conflict or Congruence*. Hillsdale, NJ: Erlbaum.
- Naus, M. J., Ornstein, P. A., Aviano, S. (1977). Developmental changes in memory: the effects of processing time and rehearsal instructions. *Journal of Experimental Child Psychology, 23*, 237-51.

- Newell, A. & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: gender effects in late adolescence. *Journal of Applied Developmental Psychology, 15*, 33-58.
- Ornstein, P. A., Naus, M. J., & Stone, B. P. (1977). Rehearsal training and developmental differences in memory. *Developmental Psychology, 13*, 15-24.
- Oyen, A. & Bebeko, J. M. (1996). The effects of computer games and lesson contexts on children's mnemonic strategies. *Journal of Experimental Child Psychology, 62*, 173-189.
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful ideas*. Harvester Wheatsheaf.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behaviour and Human Performance, 16*, 366-387.
- Payne, J. W., Betterman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge University Press.
- Pelgrum, W. J., & Plomp, T. (1991). *The Use of Computers in Education Worldwide*. Oxford: Pergamon Press.
- Pellegrini, A. D. (1995). *The Future of Play Theory*. State University of New York Press.
- Piaget, J. (1932). *The Moral Judgement of the Child*. Lonon: Routledge Kegan Paul.
- Piaget, J. (1951). *Play, Dreams, and Imitation in Childhood*. Lonon: Routledge Kegan Paul.
- Piaget, J., & Garcia, R. (1991). *Towards a Logic of Meaning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Piaget, J., & Inhelder, B. (1975). *The Origin of the Idea of Chance in Children*. New York: Norton (Original work published, 1951).

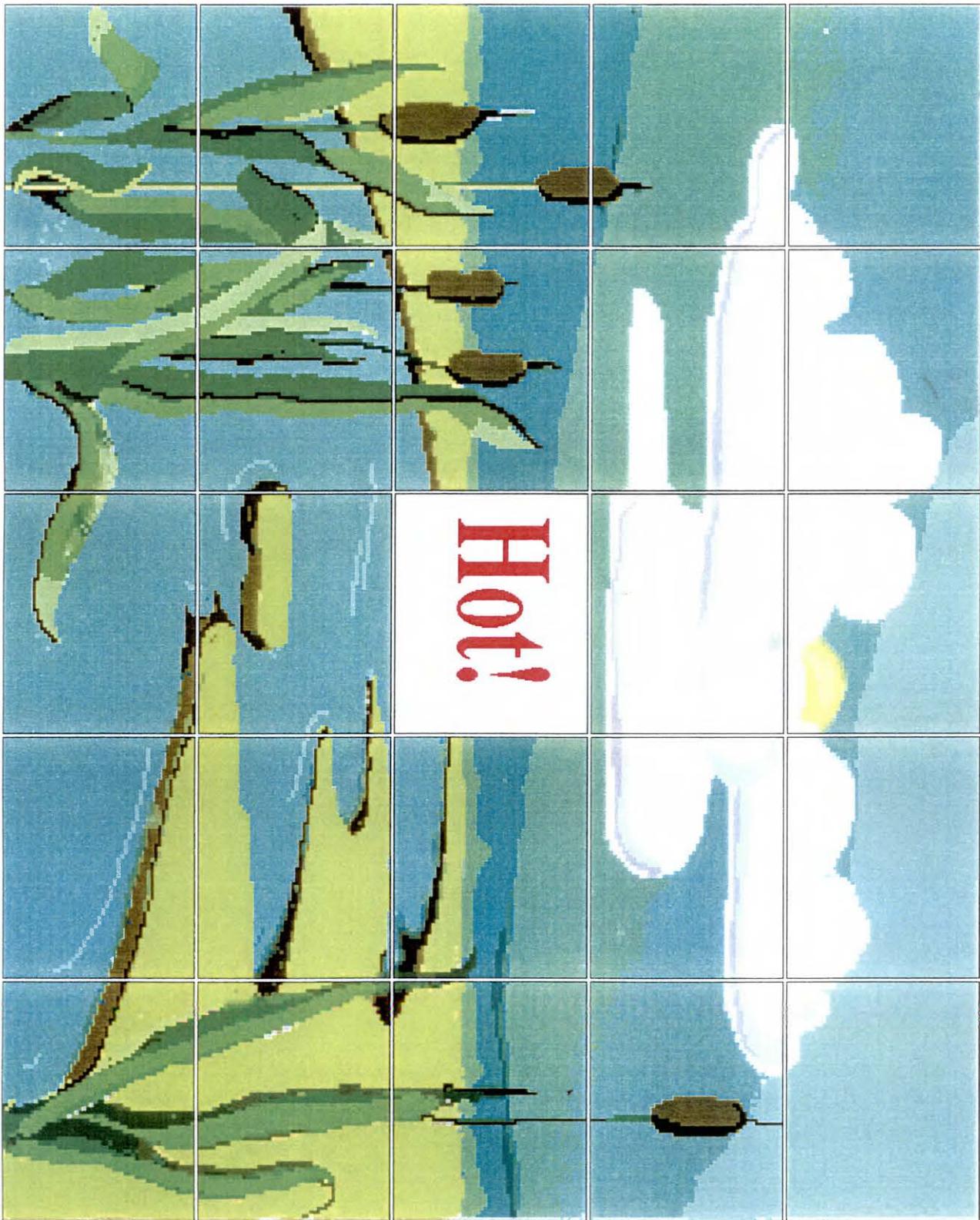
- Pieraut-LeBonniec, G. (1980). *The development of modal reasoning: Genesis of necessity and possibility notions*. New York: Academic Press.
- Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: critique, synthesis, and suggestions for the future. *Annual Review of Psychology*, 28, 368-92.
- Quinn, C. N. (1995). *Designing Educational Computer Games*. IFIP. Amsterdam: Elsevier.
- Radziszewska, B., & Rogoff, B. (1988). Influence of adult and peer collaboration on children's planning skills. *Developmental Psychology*, 24, 840-48.
- Randel, J. M., Morris, B. A., Wetzel, C. D. & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: A review of recent research. *Simulation and Gaming*, 23, 3, 261-276.
- Rapoport, A. (1990). Psychological dimensions of war. In B. Glad (Ed.), *Violence, Cooperation, Peace*. Sage publications, CA, US.
- Reese, H. W. (1962). Verbal mediation as a function of age level. *Psychological Bulletin*, 59, 502-9.
- Resnick, L. B. & Neches, R. (1984). Factors affecting individual differences in learning ability. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence Vol. 2*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Sanger, J., Wilson, J., Davies, B. & Whitaker, R. (1997). *Young children, videos & Computer games*. The Falmer Press.
- Scholnick, E. K. (1999). Representing logic. In I. E. Sigel, *Development of Mental Representation: theories and applications*. Mahwah, NJ: L. Erlbaum Associates.
- Scholnick, E. K., & Wing, C. S. (1992). Speaking deductively: Using conversation to trace origins of conditional thought in children. *Merrill-Palmer Quarterly*, 38, 1-20.
- Scholnick, E. K., & Wing, C. S. (1995). Logic in conversation: Comparative studies of deduction in children and adults. *Cognitive development*, 10, 319-345.

- Searle, J. R. (1969). *Speech Acts: An essay in the philosophy of language*. Cambridge. Cambridge University Press.
- Searle, J. R. (1995). *The Construction of Social Reality*. London: Penguin Press.
- Sedighian, K., & Klawe, M. M. (1996). Super Tangrams: A child-centred approach to designing a computer supported mathematics learning environment. *Annual report of University of British Columbia*.
- Sedighian, K., & Sedighian, A. (1997). Can educational computer games help educators learn about the psychology of learning mathematics in children? *Annual report of University of British Columbia*.
- Shepard, R., & Metzler, J. (1971). Mental Rotation of three-Dimensional objects. *Science, 191*, 952-954.
- Siegel, M. A., & Misselt, A. L. (1984). A adaptive feedback and review paradigm for computer-based drills. *Journal of Educational Psychology, 76:2*, 310-317.
- Siegel, S., & Goldstein, A. (1959). Decision-making behavior in a two-choice uncertain outcome situation. *Journal of Experimental Psychology, 57*, 37-42.
- Siegler, R. S. (1988). Individual differences in strategy choices: good students, not-so-good students, and perfectionists. *Child Development, 59*, 833-851.
- Siegler, R. S. (1989). How domain-general and domain-specific knowledge interact to produce strategy choices. *Merrill-Palmer Quarterly, 35*, 1-26.
- Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.
- Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: how do children know what to do? In C. Sphian, (Ed.), *Origins of Cognitive Skills*. Lawrence Erlbaum.
- Siegler, R.S., & Jenkins, E. (1989). *How Children Discover New Strategies*. Hillsdale, N.J.: Erlbaum.
- Silvern, S. B. (1986). Classroom use of video games. *Education Research Quarterly, 10*, 10-16.

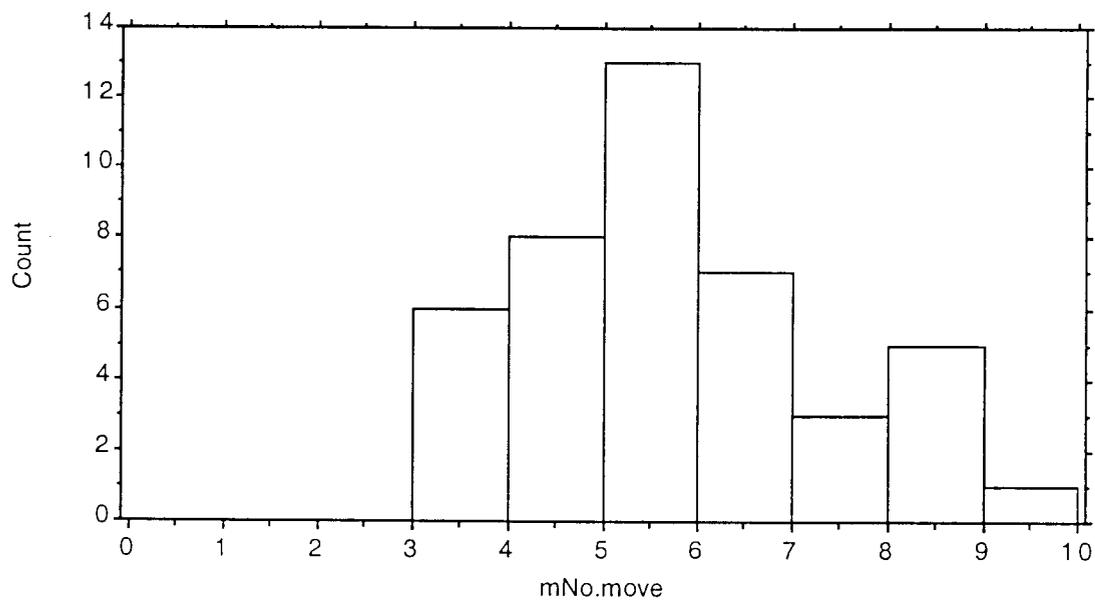
- Simon, H. A. (1990). Invariants of human behaviour. *Annual Review of Psychology, 41*, 1-19.
- Skinner, B. F. (1938). *The behaviour of Ogranisms*. New York: Appleton-Centry-Crofts.
- Smith, E. E., Langston, C., & Nisbett, R. E. (1992). The case for rules in reasoning. *Cognitive Science, 16*, 1-40.
- Somerville, S. C., Hadkinson, B. A., & Greenberg, C. (1979). Two levles of inferential behaviour in young children. *Child Development, 50*, 119-131.
- Spiker, C. C., Cantor, J. H. & Klouda, G. V. (1985). The effect of pretraining and feedback on the reasoning of young children. *Journal of Experimental Child Psychology, 39*, 381-395.
- Squires, D., & McDougall, A. (1994). *Choosing and Using Educational Software: A Teacher's Guide*. The Falmer Press.
- Squires, D., & Preece, J. (1996). Usability and learning: evaluating the potential of educational software. *Computers and Education, 27*, 1, 15-22.
- Sternberg, R. J., & Davidson, J. E. (1995). *The Nature of Insight*. Cambridge, MA: MIT Press.
- Subrahmanyam, K., & Greenfield, P. M. (1994). Effects of video game practice on spatial skills in girls and boys. *Journal of Applied Developmental Psychology, 15*, 13-32.
- Taylor, M. (1990). Simulations and adventure games in CALL. *Simulation and Gaming, 21*, 4, 461-466.
- Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science, 25*, 219-225.
- Thornton, S. (1995). *Children Solving Problems*. Harvard University Press.
- Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review, 45*:1-41.
- Toppino, T. C. (1980). Kindergartner's reasoning in simple problems. *Journal of Experimental Child Psychology, 30*, 496-512.

- Trabasso, T. R., & Bower, G. H. (1968). *Attention in Learning*. New York: Wiley.
- Underwood, J. & Brown, J. (1994). *Intergrated Learning Systems*. Oxford: Heinemann.
- Walford, R. (1995). A quarter-century of games and simulations in geography. *Simulation and Gaming*, 26, June, 236-248.
- Walker, J. T. (1996). *The Psychology of Learning: Principles and Processes*. Prentice-Hall Inc.
- Wason, P. C. & Johnson-Laird, P. L. D. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge M: Harvard University Press.
- Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, 21, 92-107.
- Wellman, H. M. (1985). *Children's searching: the development of search skill and spatial representation*. Lawrence Erlbaum Associates.
- Wellman, H. M., Somerville, S. C., & Haake, R. J. (1979). Development of search procedures in real-life spatial environments. *Developmental Psychology*, 15, 530-542.
- White, P. A. (1993). *Psychological Metaphysics*. London: Routledge.
- Whitley, B. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. *Computers in Human Behavior*, 13, 1, 1-22.
- Winograd, T., & Flores, F. (1987). *Understanding Computers and Cognition*. Addison-Wesley Publishing Company, Inc.
- Wood, D. (1998). *How children think and learn: The social contexts of cognitive development (2nd ed.)*. Blackwell Publishers Ltd.
- Wretsch, J. V. (1996). The role of abstract rationality in Vygotsky's image of mind. In A. Tryphon, & J. Veneche (Eds.), *Piaget-Vygotsky: The social genesis of thought*. Psychology Press.
- Yelland, N. (1998). Making sense of gender issues in mathematics and technology. In Yelland, N (Ed.), *Gender in Early childhood*. Routledge.

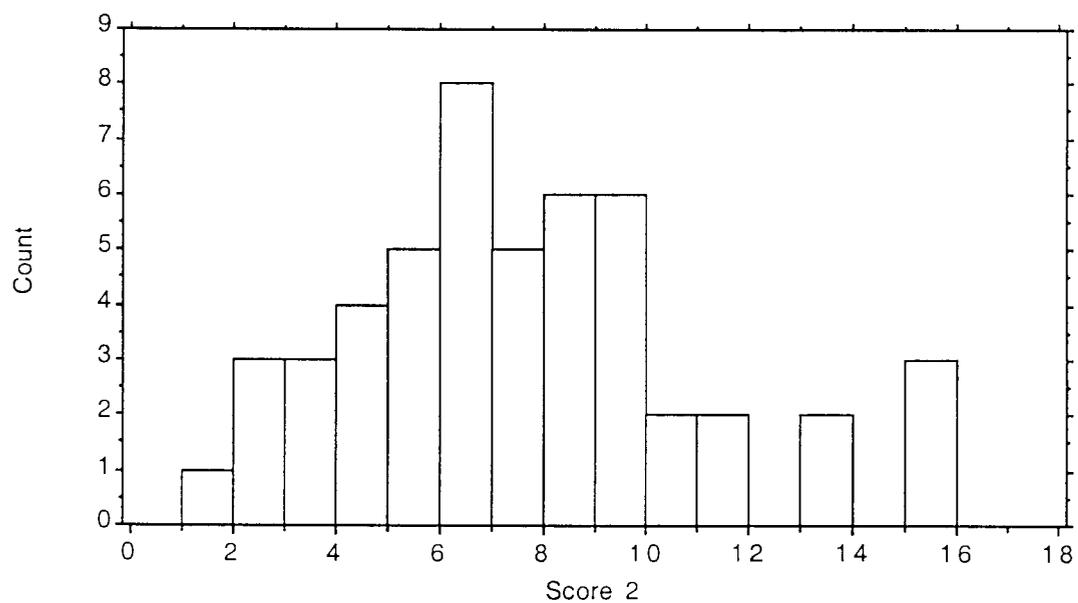
Zeaman, D., & House, B. J. (1963). An attention theory of retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of Mental Deficiency: Psychological Theory and Research*. New York: McGraw-Hill.



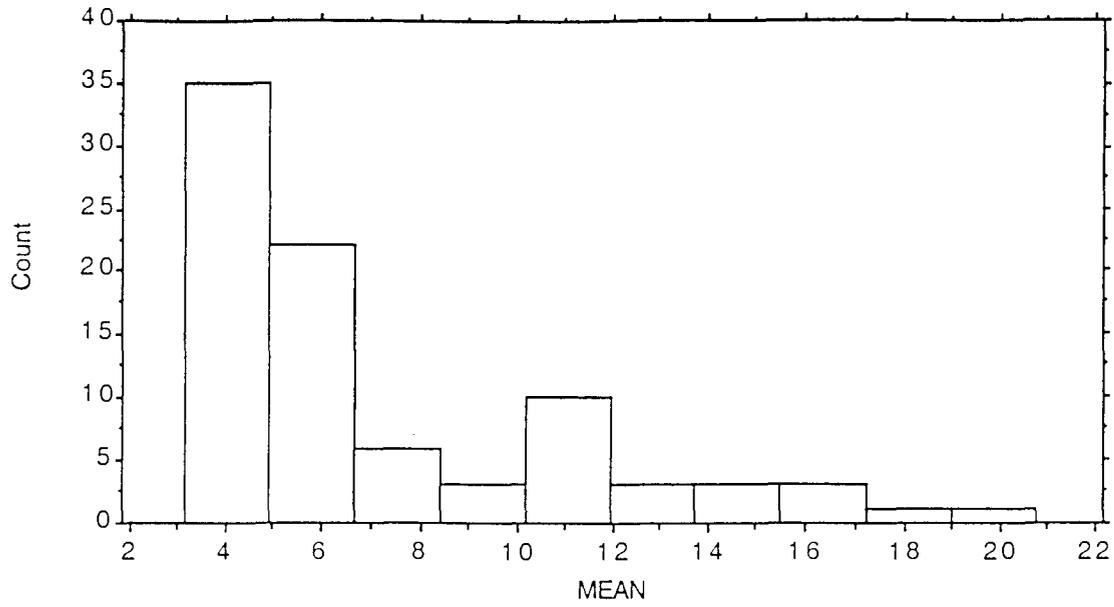
Appendix 1. A word clue test sheet



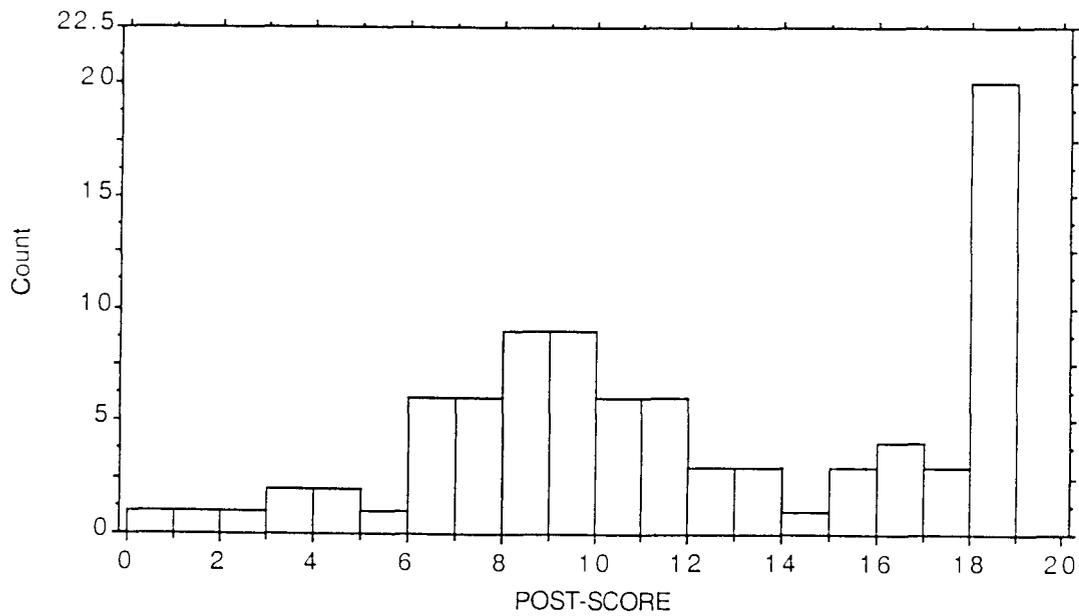
Appendix 2. Distribution of the mean numbers of moves in Study 1



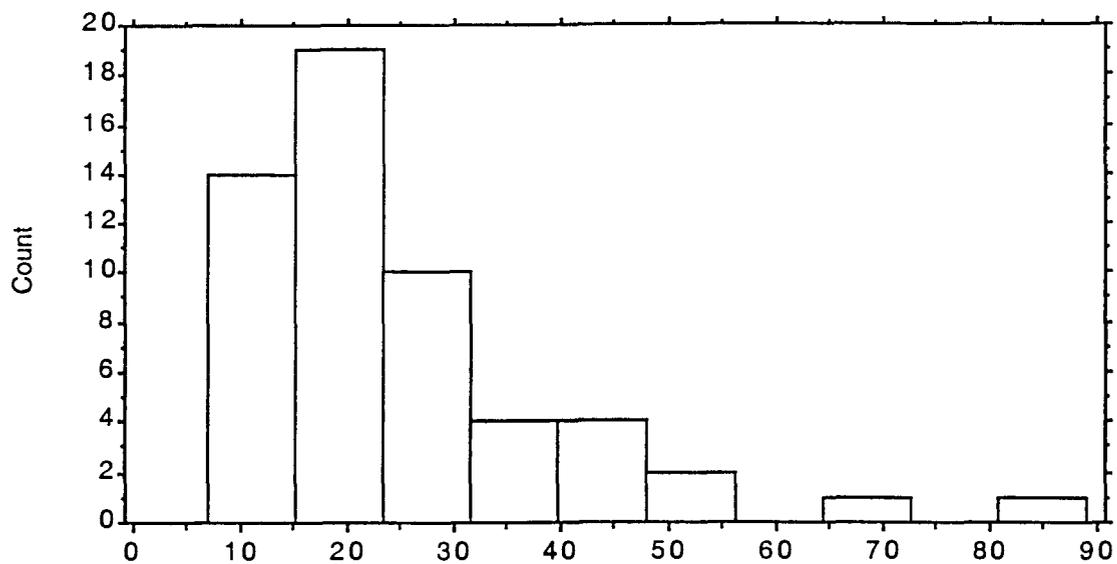
Appendix 3. Distribution of the word clue test total scores in Study 1



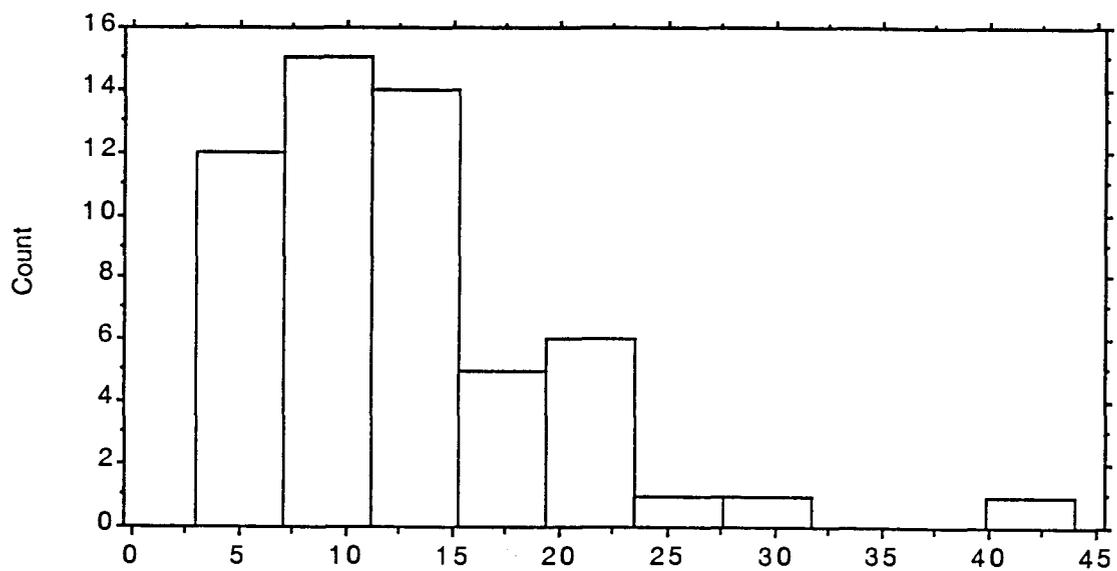
Appendix 4. Distribution of the mean numbers of moves in Study 2



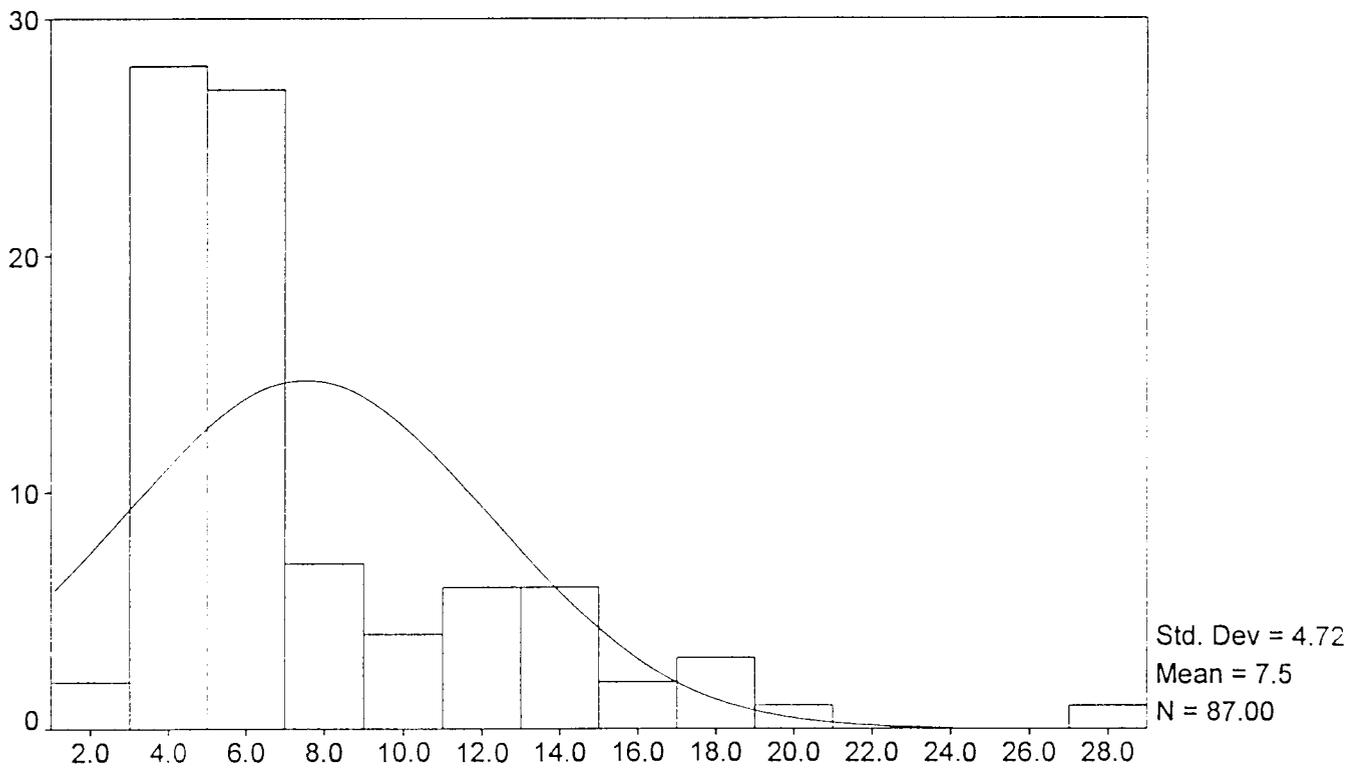
Appendix 5. Distribution of the word clue test total scores in Study 2



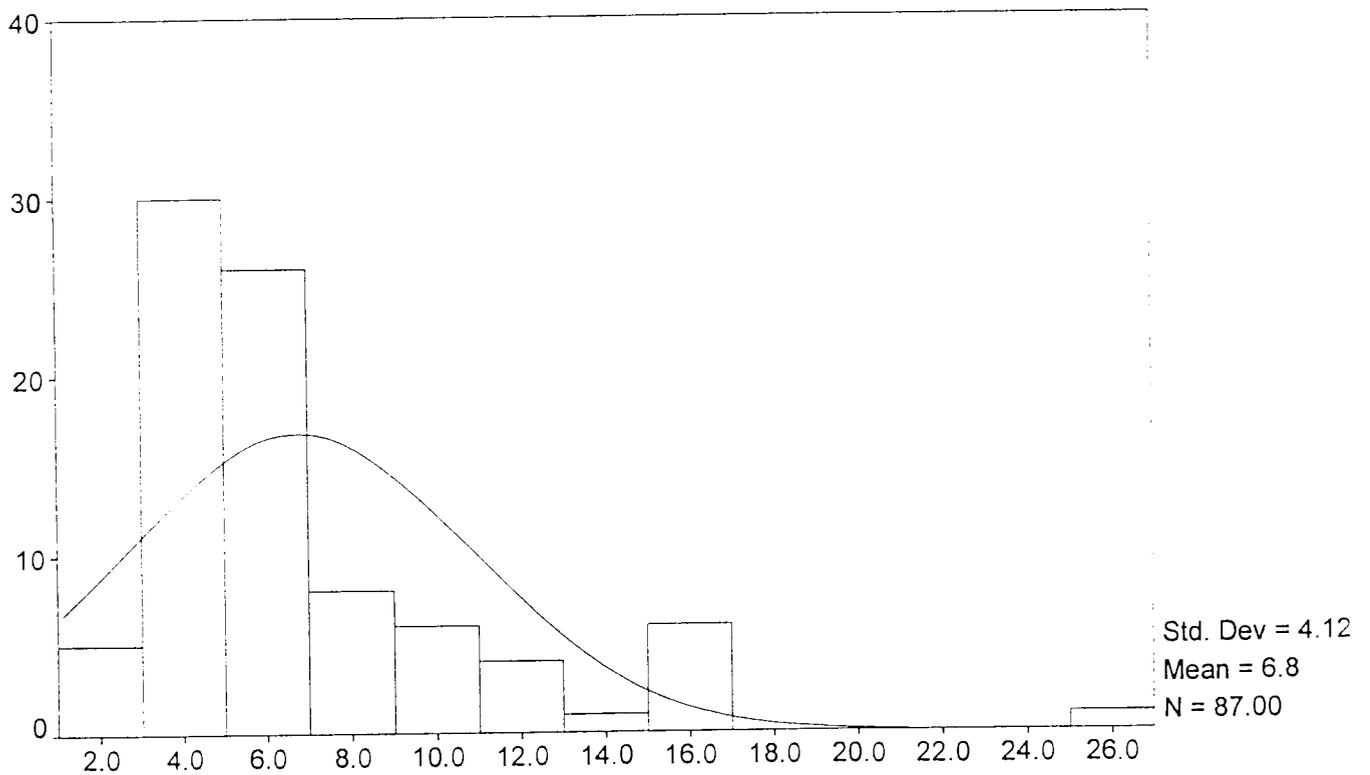
Appendix 6. Distribution of the mean time taken in the first 4 Flamingo games in Study 2



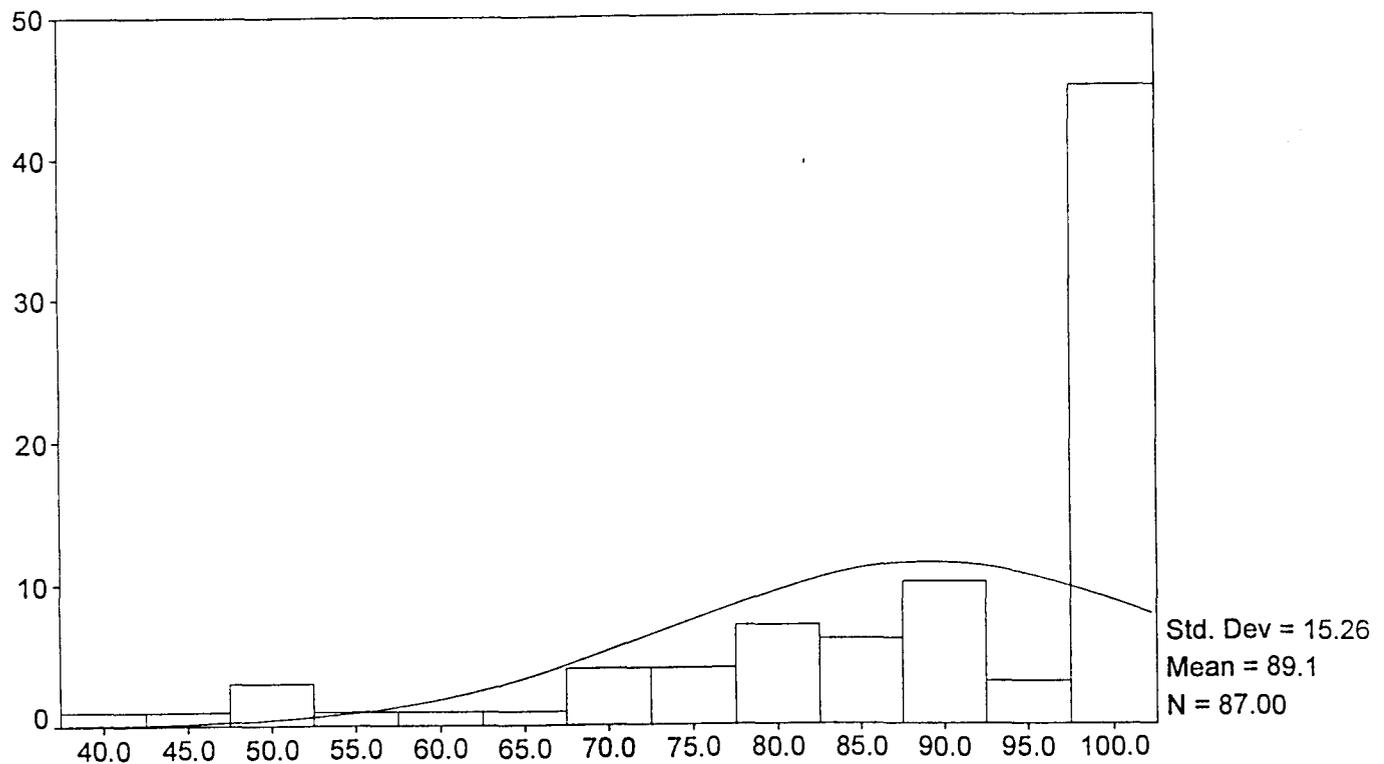
Appendix 7. Distribution of the mean time taken in the last 4 Flamingo games in Study 2



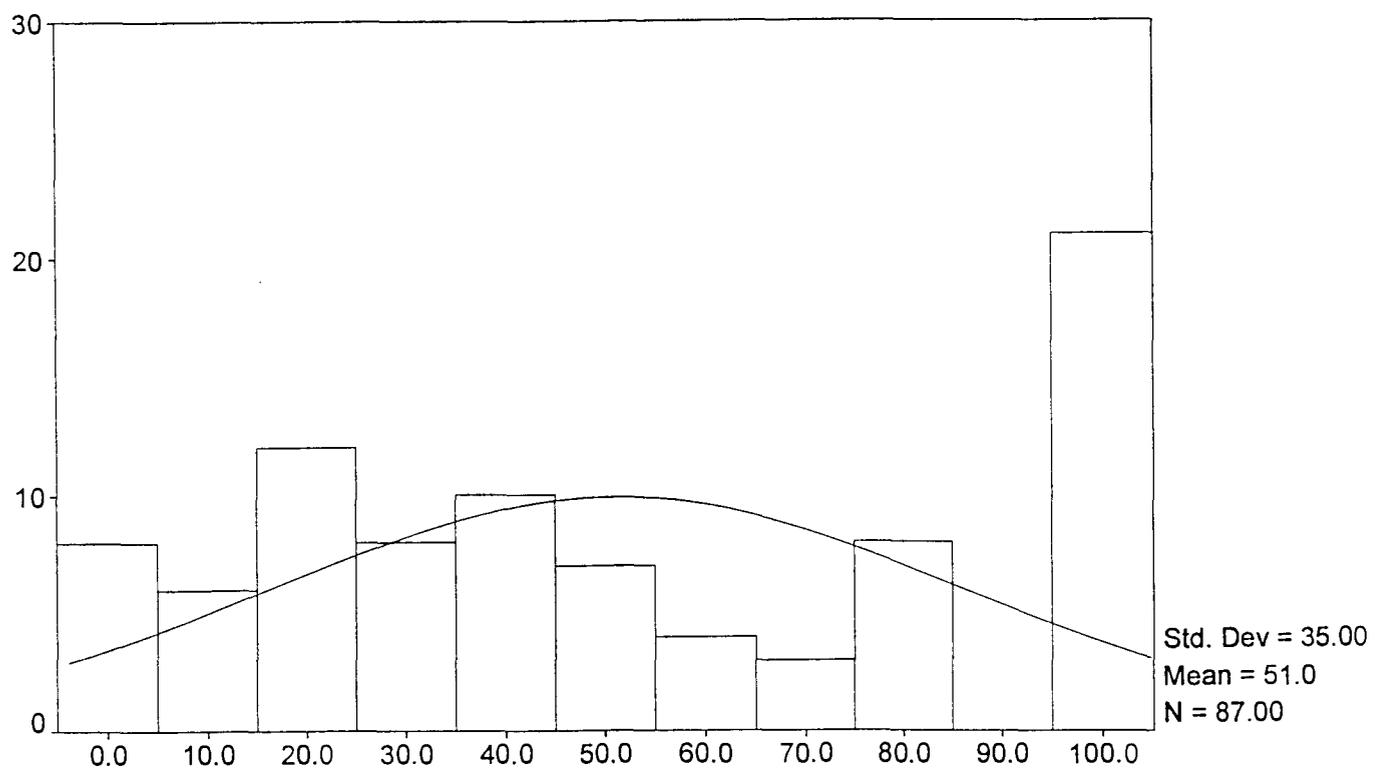
Appendix 8. Distribution of the mean number of moves of the first 4 Flamingo games in Study 2



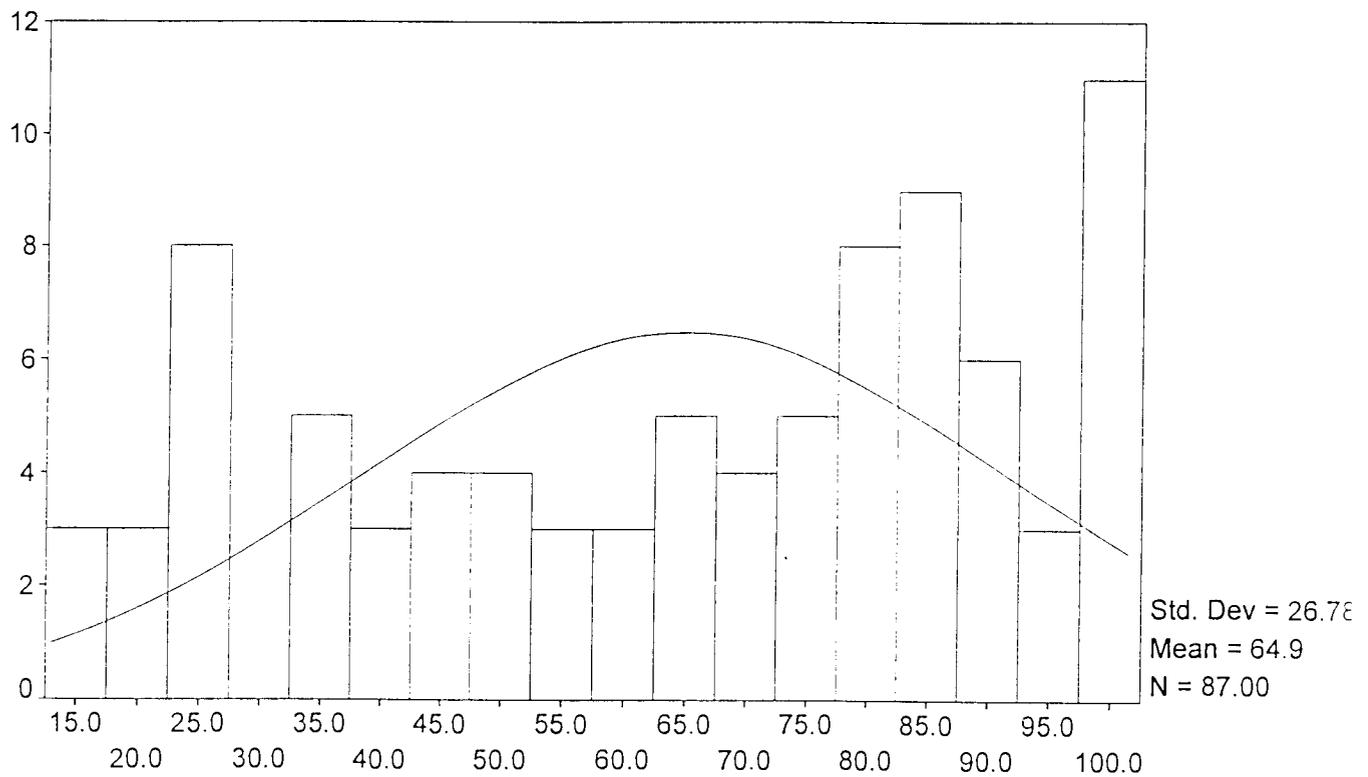
Appendix 9. Distribution of the mean number of moves of the last 4 Flamingo games in Study 2



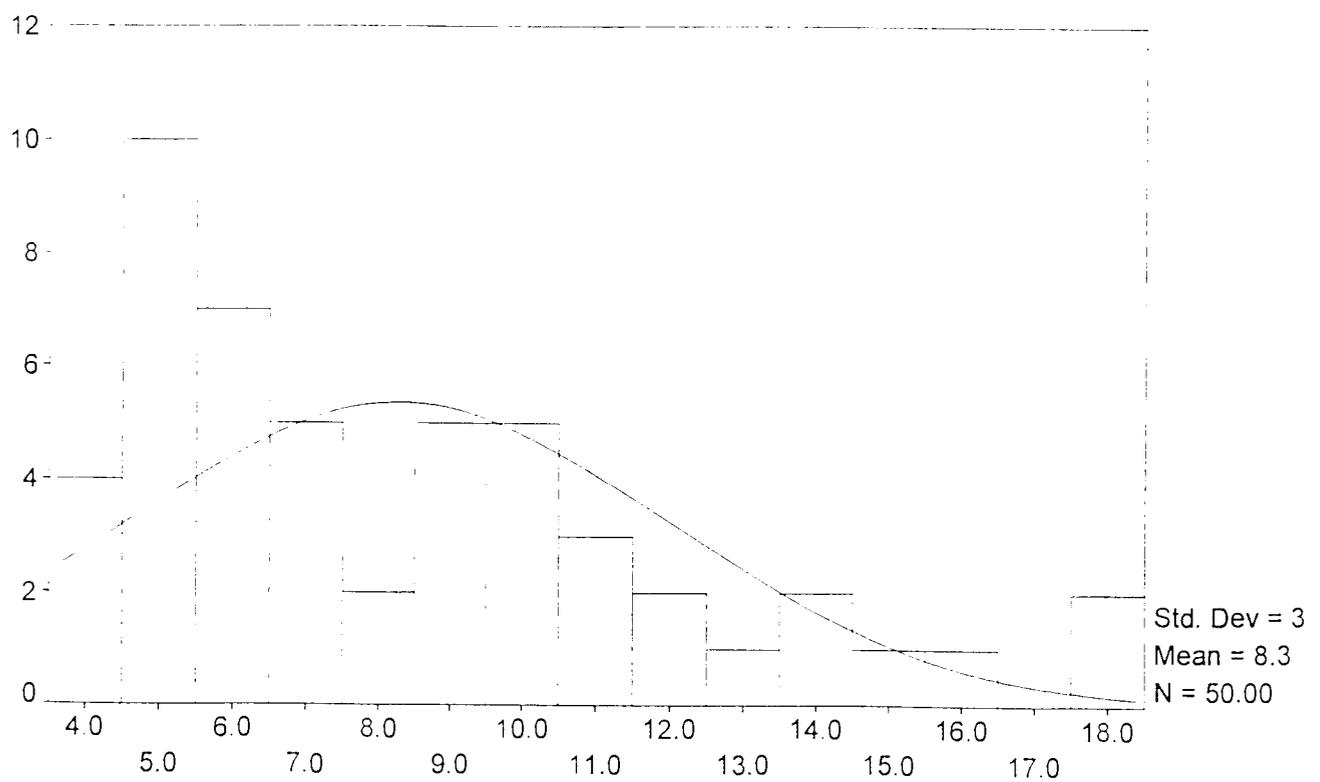
Appendix 10. Distribution of the percentages of correct inferences from HOT in Study 2



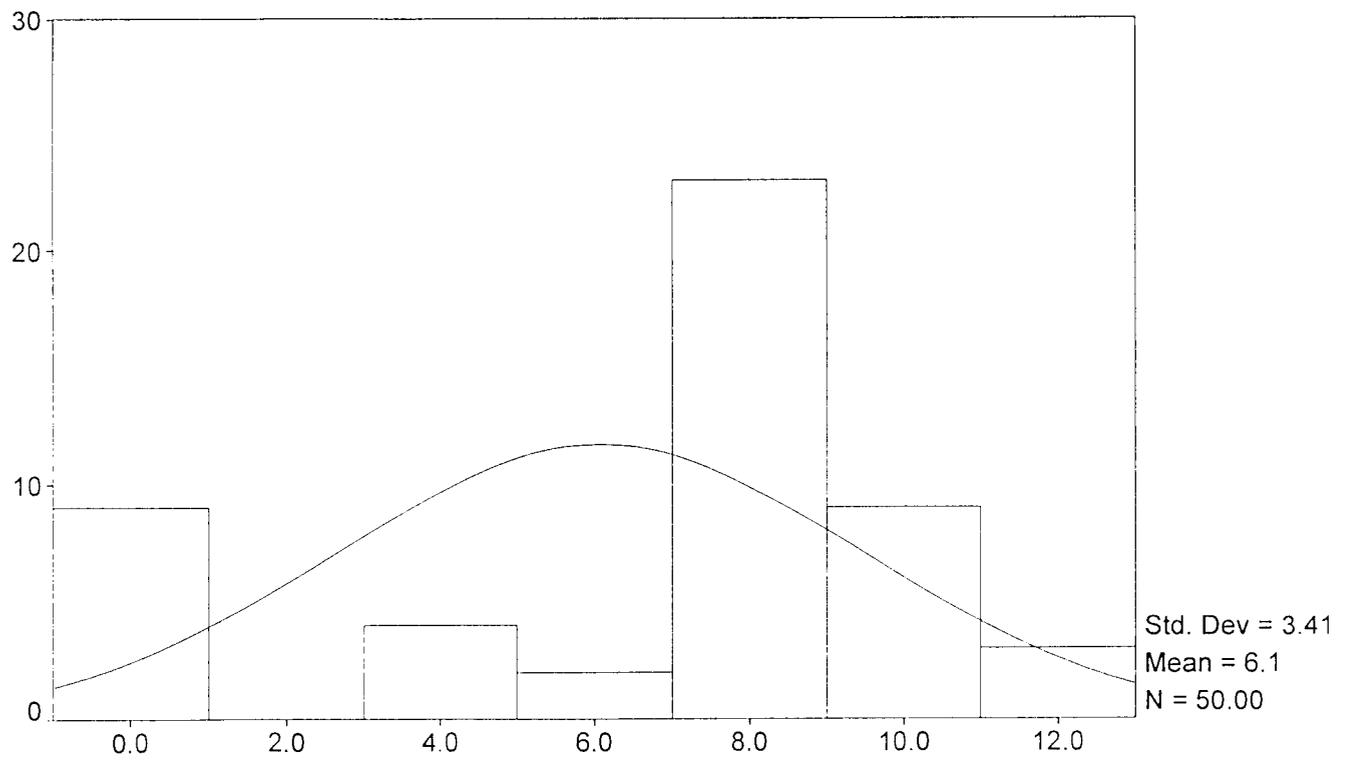
Appendix 11. Distribution of the percentages of correct inferences from WARM in Study 2



Appendix 12. Distribution of the percentages of correct inferences from COLD in Study 2



Appendix 13. Distribution of the mean number of moves of the last 4 Flamingo games in Study 4



Appendix 14. Distribution of the error scores of the Flamingo game card arrangement in Study 4