

1987

## The balance beam :: rule induction and transfer.

Leslie K. Arriola  
*University of Massachusetts Amherst*

Follow this and additional works at: <https://scholarworks.umass.edu/theses>

---

Arriola, Leslie K., "The balance beam :: rule induction and transfer." (1987). *Masters Theses 1911 - February 2014*. 2119.  
<https://doi.org/10.7275/7675783>

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

UMASS/AMHERST



312066014648782



THE BALANCE BEAM: RULE INDUCTION AND TRANSFER

A Thesis Presented

by

LESLIE K. ARRIOLA

Submitted to the Graduate School of the  
University of Massachusetts in partial fulfillment  
of the requirements for the degree of

MASTER OF SCIENCE

May 1987

Psychology

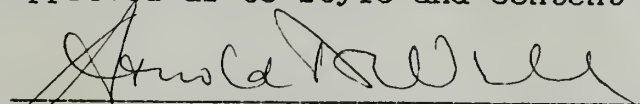
THE BALANCE BEAM: RULE INDUCTION AND TRANSFER

A Thesis Presented

by

LESLIE K. ARRIOLA

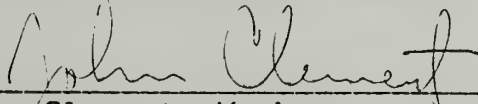
Approved as to style and content by:



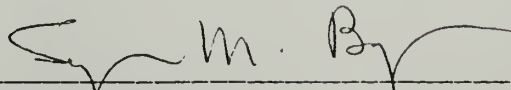
Arnold Well, Chairperson of Committee



Alexander Pollatsek, Member



John Clement, Member



Seymour M. Berger, Department Head  
Psychology

## ACKNOWLEDGEMENTS

I wish to thank the members of my Committee for their help in forming and refining the ideas in this thesis. To Arnle Well go particular thanks for the many long hours spent helping to define the research questions, design the methodology, interpret the results and gather it all into a coherent whole.

Sandy Pollatsek's comments and suggestions raised many necessary and interesting questions and John Clement provided invaluable links to other problem solving issues.

I also want to thank Jim Chumbley for his patient help in programming and debugging the balance beam graphics.

Most special thanks and appreciation go to my daughter, Ramona, for her unflagging moral support and smiling encouragement.

## ABSTRACT

One focus of this research centered on rule induction in the balance beam task; the second was concerned with the "far" transfer of balance knowledge to understanding of means and weighted means.

One major question was whether structuring the sequence of problems would facilitate rule induction. Forty subjects, pretested for lack of product-moment rule knowledge, were shown computer graphics representations of balance beam problems. They were given one of four sequencing conditions and asked to "think aloud" while making and explaining their predictions for each configuration on the beam. Error patterns and verbal protocols were analyzed.

The data indicated that solving was not dependent on whether the problems were highly sequenced or mixed, and the coded verbal protocols revealed a far more extensive taxonomy of reasoning rules and strategies than has been found in previous studies. The coded verbal protocols also provided a means for differentiating between the reasoning patterns of two types of solvers and three types of non-solvers. The reasoning patterns of solvers and non-solvers

show marked differences in general problem solving approaches and the inference drawn was that solvers were those subjects who had previously acquired problem solving skills that enabled them to make better use of the balance training.

Previous research suggested that a deep understanding of balancing would have a beneficial effect on understanding of means and weighted means (Hardiman, 1983). In the weighted mean transfer task of this study, performance of 40 balance beam trained subjects was compared to that of 40 controls. Each group was further divided into those who received an extra judgment task (EJT), which was thought to provide a possible bridge between balancing and weighted means, and those who did not.

Subjects, in "think aloud" interviews, were asked to solve two weighted mean word problems, a foil (simple mean) word problem and a choice word problem (simple mean vs. weighted mean solutions), plus two graphed frequency distribution problems.

The findings were that neither balance training alone nor the EJT had an effect on performance, but that those who had solved in the balance beam task performed significantly better than the non-solvers on the two weighted mean word problems.

The conclusion drawn was that, while it is possible that inducing the product-moment rule in the balance beam task provided subjects with knowledge applicable to understanding weighted means, it is more likely that the solvers's success resulted from a combination of being better problem solvers to begin with and being more motivated by their success in the balance beam task to try harder in the transfer session.



## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
ABSTRACT.....	iv
LIST OF TABLES.....	x
LIST OF FIGURES.....	xi
 SECTION I - The Balance Beam: Rule Induction.....	 1
Chapter	
I.    INTRODUCTION. ....	1
II.   PREVIOUS RESEARCH .....	4
Overview.....	4
Developmental Studies.....	7
Adult Studies.....	21
Major Findings on Adult Reasoning.....	28
III.  PROPOSED MODELS OF BALANCE BEAM REASONING....	34
Rationale.....	34
Assumptions for the Proposed Models.....	37
Proposed Model: Rule-Focused Subject....	38
Proposed Model: Prediction Subject.....	42
IV.   THE PRESENT STUDY.....	44
Goals and Research Questions.....	44
The Problem Set: Considerations and Rationale.....	46
Is sequencing important?.....	46
Composition of the problem set.....	47
Order of problems.....	49
Manner of presentation.....	52
Problem Sets: The Current Study.....	54
Concrete Beam vs. Computer Graphics Representation.....	55
V.    METHODS.....	57
Pretest.....	57
Training.....	60

VI. RESULTS AND DISCUSSION.....	67
Methodological Issues.....	68
Task Conditions and Problem Types.....	70
Sequencing.....	70
Problem Types: Are There Key Problems?.....	71
Reasoning, Rules and Transform Strategies.....	80
Qualitative Reasoning.....	83
Proportional Reasoning.....	92
Combining Rules.....	96
Transform Strategies.....	102
Patterns of Use: The Verbal Protocols...	105
Coding.....	109
Non-solvers.....	110
Solvers.....	113
Atypical Solvers.....	119
VII. GENERAL DISCUSSION.....	121
Task Cues vs. General Problem Solving Skills.....	121
Fit to the Proposed Models.....	130
VIII. CONCLUSIONS AND FUTURE DIRECTIONS.....	134

## SECTION II - TRANSFER

### Chapter

I. INTRODUCTION.....	141
II. PREVIOUS RESEARCH.....	143
Hardiman: The Study.....	143
Hardiman: What Transferred?.....	147
III. THE PRESENT TRANSFER STUDY.....	152
Methods.....	152
Pretest.....	152
Transfer phase.....	153

IV. RESULTS.....	157
Analysis: Pretest and Transfer Data.....	157
Pretest.....	158
Transfer Session.....	158
Problems 1-4.....	160
Discussion: Problems 1-4.....	167
Problems 5 & 6.....	173
Discussion: Problems 5 & 6.....	173
Conclusions.....	176
.....	
APPENDIX	
A. Sequences for Conditions I - IV.....	177
REFERENCES.....	179

## LIST OF TABLES

### SECTION I - BALANCE BEAM: RULE INDUCTION

B1.	Inhelder and Piaget Classification.....	8
B2.	Balance Beam Problem States.....	11
B3.	Siegler's Rule Models.....	12
B4.	Siegler's Predictions for Percentage of Correct and Error Responses for Rules I - IV.....	13
B5.	Production System Representations for Models I - IV.....	16
B6.	Balance Beam Pretest.....	58
B7.	Number of Solvers, Non-solvers and Atypical Solvers.....	70
B8.	Average Percentage of Errors on Simple Conflict Problems.....	72
B9.	Percent (and Number) of Subjects Correct on First Conflict Problem.....	78
B10.	References to Previous Problems.....	90
B11.	Non-solvers: Coded Verbal Protocols.....	106
B12.	Solvers: Coded Verbal Protocols.....	107

### SECTION II - TRANSFER

T1.	Hardiman's Transfer Test Problems.....	146
T2.	Pretest Weighted Mean Problems.....	153
T3.	Transfer Problems: #5 - Histobeam and #6 - Histogram.....	156
T4.	Percent Correct: Trained and Control Groups.....	160
T5.	Percent Correct: EJT and No EJT Groups.....	161
T6.	Percent Correct: Solvers and Non-solvers.....	162
T7.	Mean Correct Responses per Subject for Arriola Pretest Problems and Arriola and Hardiman Transfer.....	163



## List of Figures

1. Proposed Model: Rule-Focused.....	39
--------------------------------------	----

S E C T I O N    I  
THE BALANCE BEAM: RULE INDUCTION

C H A P T E R    I  
INTRODUCTION

This research is an investigation into the types and patterns of reasoning adult subjects use in an induction task and the conditions that affect performance. The specific concern is with how people develop an understanding of physical concepts in which 1) there are two, physically separable, relevant variables, 2) the variables can be individually measured or quantified, and 3) these measurements can be combined, using an algebraic rule, in order to predict what will happen for any combination of the two variables. Examples of such physical concepts include density, the size of shadows, and the balance beam.

This research is concerned with the induction of the product-moment (p-m) rule for the balance beam, given experience with various configurations of weights and distances on the beam. The p-m rule is an integrative algebraic computation for predicting the action of a balance beam using a simple calculation of torque (i.e., the sum of the products of each weight and its distance from the

fulcrum) for each side of the beam. The two torques are then compared and the side with the greater torque is the side that will tip down. If the torques are equal, the beam will balance.

Balancing is an interesting domain in that while nearly all subjects have had experiences with balancing (seesaws, scales, etc.) and most subjects recognize weight and distance as the relevant variables to use in predicting the action of the beam, few are able to specify the general computational rule that would allow them to predict correctly the effects of all configurations of weights and distances on the beam.

Even more interesting is the fact that subjects who have been provided with various types of learning experiences with a concrete beam still have difficulty in inducing the rule (Siegler, 1976; Siegler & Klahr, 1982).

Performance of untrained subjects in the Siegler studies suggested that subjects who did not know the  $p-m$  rule were using a limited set of qualitative, non-computational rules that were adequate to predict the outcome of simple configurations of weights and distances on the beam, but were inadequate for more complex configurations in which weight and distance cues conflicted. Even after training with these more complex problems, subjects were seldom able to make the transition from the simpler rules to the type of

computational reasoning that would allow them to induce the p-m rule, despite the ease with which they understood and used the p-m rule once they knew it. In one study (Hardiman, 1983), however, the majority of subjects did learn to induce the p-m rule.

This gives rise to the two aspects of rule induction that the current research investigated. The first issue is the types and patterns of reasoning rules people use as they progress through the balance beam induction tasks, and the second is the difficulty they have in making the transition from simple reasoning rules to the algebraic p-m rule.



## C H A P T E R   I I

### PREVIOUS RESEARCH

#### Overview

With regard to the first issue, there have been two major theories about the types and phases of reasoning used in understanding the balance beam: Inhelder and Piaget's (1958) stage theory, and Siegler's (1976) hierarchical decision tree models. These theories are discussed briefly below. In general, these studies have found evidence that subjects proceed through a sequence of increasingly systematic stages of reasoning. However, these theories are based primarily on studies of the reasoning processes of children and the applicability to adult reasoning, which is the focus of this research, has yet to be determined.

When one looks at the second issue - the difficulty children and adults alike have in inducing the p-m rule, even after learning experiences with the balance beam - it is not at all clear from the previous research (Siegler, 1976; Klahr & Siegler, 1978) that the difficulty can be attributed to the same factors for both age groups.

The question of the relevance of previous research to the study of adult reasoning requires a more explicit

comparison of problem solvers by age, by performance or by a combination of both.

By comparing age groups, it becomes possible to determine whether the Inhelder and Piaget, or Siegler models of the developmental advances in children's reasoning abilities are useful in characterizing the path of reasoning through which adults progress as they attempt to understand the interactive effects of the balancing variables. Is adult learning about balancing "developmental" in the same sense as it is with children or is the sequence of rule acquisition observed by Inhelder and Piaget and by Siegler solely a reflection of children's developing reasoning?

Comparing the performance of different age groups also enables us to explore the possibility that children and adults adopt the same beginning sequence of reasoning rules when presented with the balance task, but, at some critical point, their reasoning patterns diverge according to level of ability or amount of previous experience with similar tasks.

If we compare subjects by performance (i.e., solvers vs. non-solvers), modeling successful and unsuccessful subjects separately has two benefits. One, it prevents the premature assumption that the poorer solver is just a good solver gone astray. This may, in fact, turn out to be the case, but it may also be the case that the poor solver's reasoning has

very little in common with that of the good solver. Developing separate models should clarify the issue. The second benefit would be to gain a much clearer picture of the way in which the unsuccessful subject's reasoning processes hinder the induction process and the type of instruction that might remedy the situation.

The framework for the following discussion is one in which performance is considered relative to age. In this way, commonalities and differences in reasoning across age groups can be highlighted while differences in performance within age groups become more clearly defined.

Therefore, the following discussion of previous research has been divided into those studies that have looked at developmental questions about children's stages or progressions of reasoning (Inhelder & Piaget, 1958; Slegler, 1976), and those that have looked at the reasoning of adult (high school and college-age) subjects (Slegler, 1976; Slegler & Klahr, 1982 ; Hardiman, 1983). This should define the issues more precisely and clarify the previous conclusions.

## Developmental Studies

### Inhelder and Piaget

Inhelder and Piaget (1958) used two types of balancing tasks for subjects ages three to fourteen. In the first, they presented subjects with a beam that had twenty-eight holes equally spaced on both sides of the fulcrum, and various sized weights that could be hung anywhere along the beam. In the second task, the beam was solid and the distance along the beam unmarked. The weights were dolls that were placed in a basket on either side of the fulcrum and the basket could be slid along the beam. Subjects were told to experiment with the beam to discover how it worked.

Inhelder and Piaget characterized the children's reasoning as a progression through increasingly more sophisticated stages of reasoning. The stages, listed in Table B1, suggest that children advance developmentally to more systematic and rule-governed reasoning and that there are qualitative changes in conceptual understanding with age across domains. However, since subjects in the first task never hung weights from more than one hole on a side, the reasoning studied with these children is limited to reasoning about proportionality in that when there is only one stack of weights on a side, the ratio rule ( $w_1/w_2 =$



TABLE B1  
Inhelder and Piaget Classification\*

Stage IA: Subjects fail to distinguish their own actions from external processes (e.g., the subject will push the beam so that it is level and expect it to remain that way).

Stage IB: Subjects realize that weight is needed on both sides of the fulcrum to achieve balance but there is as yet no systematic correspondence between weight and distance.

Stage IIA: Subjects achieve balance by making weight and distance both symmetrical. Subjects discover by trial-and-error that there is equilibrium between a smaller weight at a large distance from the fulcrum and a greater weight at a small distance but do not draw out general consequences.

Stage IIB: Subjects develop qualitative understanding of the relationship between weight and distance.

Stage IIIA: Subjects start to discover the quantitative law for balancing. It takes the form of the proposition  $W/W' = L'/L$ , where  $W$  and  $W'$  are two unequal weights and  $L$  and  $L'$  are the distances from the fulcrum at which they are placed.

Stage IIIB: Subjects search for a causal explanation.

\* From Hardiman, Pollatsek, & Well, 1985

d2/d1) will suffice to predict accurately whether or not the beam will balance. Consequently, subjects who progressed to more sophisticated and numerical proportional reasoning induced the ratio rule which appeared general but was, in fact, limited to two-stack situations (one on either side) and to the determination of balance or imbalance, but not the direction of imbalance. Thus, the stages are limited to proportionality reasoning, and the broader (multiplicative) understanding of the balance beam that would result in induction of the p-m rule is missing. Also lacking from the stage models as described in Table B1 are the processes by which a person progresses from one stage to another.

An interesting question for the current research on adults is whether a stage approach, which reflects the development over time of children's cognitive abilities, can be extended to explain the changes in reasoning that occur when a person is presented with a problem for which he or she has the cognitive capacity to understand the solution, but for which the solution still needs to be worked out. In such a situation, the adult solver might pass through two or more different stages of reasoning in a single problem-solving session.

Siegler

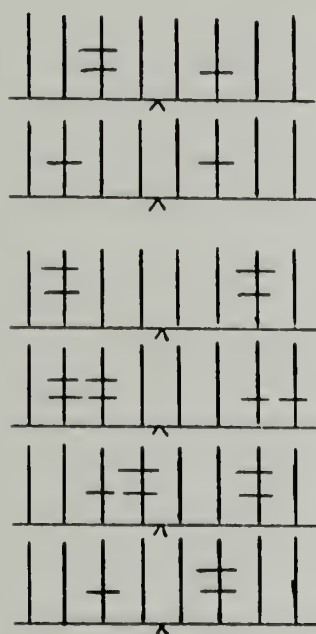
Siegler (1976) used a balance beam with pegs evenly spaced on either side of the fulcrum. Equal-size weights with holes in the middle could be placed in various configurations on the pegs. He assessed K - 12th graders' balance knowledge following experience in one of three conditions. In the a priori condition, subjects were not given any experience with the balance beam before their knowledge was tested. In the experimentation condition, subjects were told to experiment with the weights on the pegs in order to discover the rules by which the balance beam worked. In the observation condition, subjects were also told to try to figure out the rules to predict the action of the beam, but in this condition the experimenter decided how to put the weights on the pegs and the subjects were only allowed to observe the resulting effects.

For each condition, a subject's understanding of balancing was assessed in a posttest consisting of 30 balance problems. On each problem the subject began with an empty beam, the arms of which were supported in a horizontal position by wooden blocks. The weights were placed on the pegs on both sides of the fulcrum and subjects were asked to predict whether the beam would tip left, tip right or balance for each of the problems. Problems were divided into the six types described in Table B2. There were four each

of the weight, distance and balance problems and six each of the conflict-weight, conflict-distance and conflict-balance problems.

Table B2

## Balance Beam Problem States



Weight - unequal amounts of weight  
equidistance from the fulcrum

Distance - equal amounts of weight at  
unequal distances from the  
fulcrum

Balance - equal amounts of weight at  
equal distances from the  
fulcrum

Conflict-weight - the side with the  
weight will drop

Conflict-distance - the side with the  
greater distance will drop

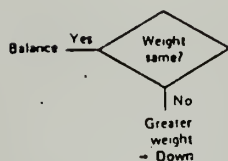
Conflict-balance - the beam will  
balance

Siegler hypothesized that children's knowledge of how balance scales operate could be represented by the set of four decision rule models shown in Table B3, and that their knowledge differed only in their consideration of weight and distance factors. Children using Rule I consider only weight; children using Rule II consider distance from the fulcrum, but only if the weights are equal; and children using Rule III consider weight and distance as in Rule II,

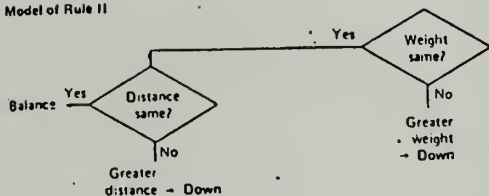


Table B3  
Siegler's Rule Models

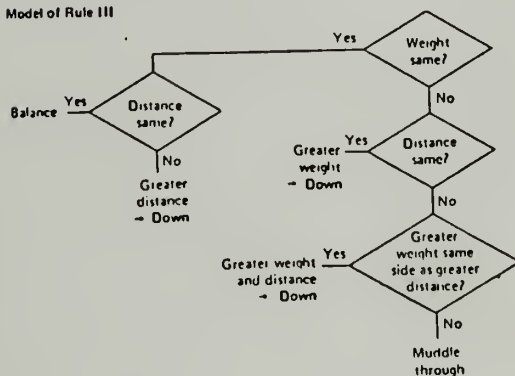
Model of Rule I



Model of Rule II



Model of Rule III



Model of Rule IV

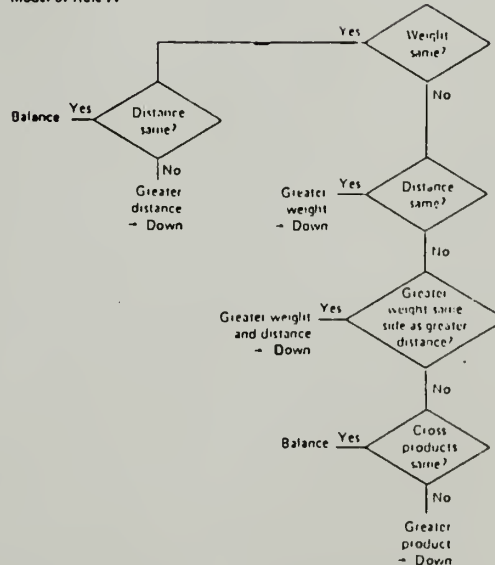


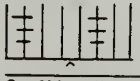
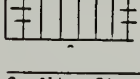
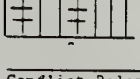
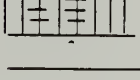


FIG. 3.3. Decision Tree Models of Rules for Performing Balance Scale Task. From "Three aspects of cognitive development" by R. S. Siegler, *Cognitive Psychology*, 1976, 8, 481-520. Copyright 1976 by Academic Press. Reprinted by permission.

Table B4

# Siegler's Predictions for Percentage of Correct and Error Responses for Rules I - IV

Problem Type	Rule			
	I	II	III	IV
Balance	100	100	100	100
				
Weight	100	100	100	100
				
Distance	0	100	100	100
				
			(Should say "Balance")	
Conflict-Weight	100	100	33	100
				
			(Chance Responding)	
Conflict-Distance	0	0	33	100
				
	(Should say "Right Down")	(Should say "Right Down")	(Chance Responding)	
Conflict-Balance	0	0	33	100
				
	(Should say "Right Down")	(Should say "Right Down")	(Chance Responding)	


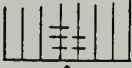




## Siegler's Developmental Trends Observed and Predicted on Different Problem Types (Experiment 1)\*

Problem Type	Age				Predicted Developmental Trend
	5-6	9-10	13-14	16-17	
Balance	94	99	99	100	No change-all at high level
Weight	88	98	98	98	No change-all at high level
Distance	9	78	81	95	Dramatic Improvement with age
Conflict-weight	86	74	53	51	Decline with age-possible upturn for oldest
Conflict-distance	11	32	48	50	Improve with age
Conflict-balance	7	17	26	40	Improve with age

\*Percentage of problems predicted correctly.

Table B4

## Siegler's Predictions for Percentage of Correct and Error Responses for Rules I - IV

Problem Type	Rule			
	I	II	III	IV
Balance	100	100	100	100
				
Weight	100	100	100	100
				
Distance	0	100	100	100
				
	(Should say "Balance")			
Conflict-Weight	100	100	33	100
				
			(Chance Responding)	
Conflict-Distance	0	0	33	100
				
	(Should say "Right Down")	(Should say "Right Down")	(Chance Responding)	
Conflict-Balance	0	0	33	100
				
	(Should say "Right Down")	(Should say "Right Down")	(Chance Responding)	

## Siegler's Developmental Trends Observed and Predicted on Different Problem Types (Experiment 1)\*

Problem Type	Age				Predicted Developmental Trend
	5-6	9-10	13-14	16-17	
Balance	94	99	99	100	No change-all at high level
Weight	88	98	98	98	No change-all at high level
Distance	9	78	81	95	Dramatic Improvement with age
Conflict-weight	86	74	53	51	Decline with age-possible upturn for oldest
Conflict-distance	11	32	48	50	Improve with age
Conflict-balance	7	17	26	40	Improve with age

\*Percentage of problems predicted correctly.

except when weight and distance cues conflict (i.e., there is more weight on one side and more distance on the other), in which case they "muddle through" or guess. Rule IV represents the computation of the p-m rule. Siegler predicted that a child using a particular decision rule would show a characteristic pattern of responses, consistent within, as well as across, problem types, and that there would be developmental changes in rule use with age. The predicted and observed patterns for each decision rule model are shown in Table B4.

Across several experiments, the decision rules accurately described the performance of more than 80% of the children. That is, the pattern of a child's correct and error predictions allowed Siegler to classify a significant number of children as using one of the decision rules as their basis for making predictions about the balance beam.

The models indicate an invariant sequence of rule development but, as in the Inhelder and Piaget stages, it is not clear how a person's reasoning progresses from one rule model to the next -- what triggers a change in the representation of the problem, and what is the effect of the change on the representation?

Klahr and Siegler: Production systems

In an attempt to address the limitations of the decision tree representations of children's balancing knowledge, Klahr and Siegler (1978) restated the four models in Table B3 in terms of production systems. Production system representations for Rule Models I-IV are shown in Table B5. In their production system, a set of rules - called productions - are written in the form of condition-action pairs which operate via a recognize-act cycle. When the conditions of one production are satisfied (i.e., match the current contents of the subject's activated portion of long term memory), the associated action 'fires.' If more than one production is satisfied at a given moment, a conflict resolution principle is invoked - in this system, special cases have priority over general cases. (See Klahr & Siegler, 1978, for a more extensive description of conflict resolution.)

The production system representations are an advance over the decision tree rule models in that they are more explicit models of the conditions and actions involved in balancing discriminations. Further, the system makes clearer the appropriate underlying operator knowledge needed to add a production, and hence gives a clearer picture of which elements of the operators are important for the task,



as well as pointing to the issue of individual differences in

Table B5

Production System (P) Representations for Models I-IV  
(D = distance; W = weight)

## Model I

P1: ((Same W ----> (Say "balance"))  
P2: ((Side X more W ----> (Say "X down"))

## Model II

P1: ((Same W ----> (Say "balance"))  
P2: ((Side X more W ----> (Say "X down"))  
P3: ((Same W) (Side X more D) ----> (Say "X down"))

## Model III

P1: ((Same W ----> (Say "balance"))  
P2: ((Side X more W ----> (Say "X down"))  
P3: ((Same W) (Side X more D) ----> (Say "X down"))  
P4: ((Side X more W) (Side X less D) ----> muddle through)  
P5: ((Side X more W) (Side X more D) ----> (Say "X down"))

## Model IV

P1: ((Same W ----> (Say "balance"))  
P2: ((Side X more W ----> (Say "X down"))  
P3: ((Same W) (Side X more D) ----> (Say "X down"))  
P4: ((Side X more W) (Side X less D) ----> muddle through)  
P5: ((Side X more W) (Side X more D) ----> (Say "X down"))  
P6: ((Same Torque) ----> (Say "balance"))  
P7: ((Side X more Torque) ----> (say "X down"))

Transitional requirements  
Productions Operators

I -> II      add P3      add distance encoding and comparison

II -> III    add P4, P5

III -> IV    modify P4;      add torque computation and  
              add P6, P7      comparison

this underlying knowledge. This poses an interesting conflict in thinking about adult subjects who are assumed to already possess all the underlying knowledge needed to perform the balance task.

However, Klahr and Slegler's production systems do not seem to be self-modifying. Unlike Lewis and Anderson's (1985) geometry production systems (ACT) which attempt to account for modification of a production as a result of finer and finer discriminations among conditions whenever a condition-action pair is found to have inappropriately fired, the Klahr and Slegler productions do not account for learning in this sense.

### Methodological Issues

There are two further problems with Siegler's rule assessments, both a result of the methodology. The first is that the task of predicting left, right or balance is a forced-choice task. The determination that a subject is using a particular rule to predict all problems is based on the number of correct and incorrect predictions the subject makes for each type of problem. So, for example, if a subject correctly predicts the outcome of most balance (e.g., 0003/3000 -- Note: notation indicates the number of weights located 1,2,3 or 4 distance units from the fulcrum. In this example, there are three weights located one distance unit from the center on the left and three weights also at one unit from the center on the right), weight (e.g., 0003/2000) and conflict-weight (e.g., 0300/0002) problems, but incorrectly predicts that all distance problems (e.g., 0300/0300) will balance, all conflict-distance problems (e.g., 0200/4000) will tip right and all conflict-balance problems (e.g., 0020/4000) will also tip right, that subject would be classified as using Rule I, i.e., relying solely on weight cues.

However, by limiting subjects' responses to these three choices, one limits the set of rules that can be inferred from their responses. Certainly Rules I and II are commonly found reasoning rules for this task. But Rule III in this

classification is not a decision rule at all, but simply an indication that the solver recognizes that there are problems with conflicting weight and distance cues that cannot be solved by Rules I and II. Verbal protocols in a study by Hardiman (1983), which will be discussed in detail later, showed clear evidence that subjects were not just "muddling through", as Siegler suggests, but were in fact generating and testing rules of limited generality, such as ratio and addition rules, as well as referring to previously encountered problems as a basis for making current predictions. This strongly suggests the need for a more complex model of the balance beam reasoning processes than has been heretofore put forth.

A further criticism of the number-correct criteria is that it is entirely possible for subjects to exhibit a pattern of correct responses that would appear to indicate use of the p-m rule without having any notion about the exact rule itself. A pilot study for the proposed research was conducted using adult subjects. Some subjects who reached the Siegler number-correct criterion for Rule IV use had verbal protocols that clearly indicated they did not know, and were not using, the p-m rule to make predictions.

The second, related problem is that when responses are limited to three choices, there is no way to determine what problem-specific strategies and general problem-solving

skills a person brings to bear on the task. For instance, in the pilot study mentioned above, many subjects used a variety of strategies to transform the current problem into an equivalent, previously learned configuration.

For example, many subjects attempted to redistribute the weights for a particular problem, increasing or decreasing weights to compensate changes in the distance of each weight, such that the new arrangement was somehow equivalent to the original problem but was now in some familiar configuration for which the subject already had a reliable rule. A second heuristic was to cancel equalities, side to side, and predict on the basis of the leftovers, much like people simplify and reduce fractions. These strategies are missing in the Decision Tree analysis, as are the decisions that send the reasoning process off in the wrong direction.

Furthermore, while Rule II may logically follow Rule I, since it includes Rule I knowledge, and Rule III invariably follows Rule II, the invariance of the sequence of rule acquisition may stop at Rule III. In essence, Rules I and II patterns of responses indicate that the solver has identified and is considering the relevant variables, and a Rule III pattern indicates the solver has become aware that there are 'tricky' (conflict) problems for which he has, as yet, no existing rule. But the type of reasoning that



occurs after this point is not necessarily categorically and logically the same as that which was used before. Thus, the various types of rules a person will hypothesize to deal with these 'tricky' conflict problems will not necessarily be generated in any particular order. That is, defining the relevant variables and identifying the conflict situations is likely a very different problem-solving process than figuring out how to use those variables to operate on the conflict problems.

Siegler also investigated why older (8-year-old) children benefit more from training with the balance beam than younger (5-year-old) children and found that the younger children did not encode distance. After training in encoding distance on the balance beam, the younger subjects, who could then be equated for rule level and distance encoding with older subjects, still showed no change in their balance beam prediction rules. However, both groups now benefitted equally from training experience with the next higher level of balance problem.

In sum, Siegler showed that children's reasoning is systematic and rule-governed and that encoding distance may be the key to a child's ability to advance in his reasoning about balancing. However, the forced-choice task and the resulting decision rule models are inadequate to represent

the variety and complexity of rules and strategies adults use in this task.

### Adult Studies

When looking at the performance of adult subjects in this induction task, there are certain assumptions that can be made about their existing knowledge and prior experience that can not be made about younger subjects: 1) they know how to encode the numerical values of weight and distance; 2) they have prior knowledge of the math operations (add, subtract, multiply and divide) necessary to compute the p-m rule; and 3) they know the difference between qualitative and quantitative rules. However, even with the advantage of prior knowledge of and experience with the individual elements of the rule, older subjects still have tremendous difficulty in inducing the rule.

The issue when looking at older subjects, then, is not about what they know, but about how and when previous knowledge is accessed and applied. What types of rules and strategies do older subjects use? Are there patterns of reasoning that differentiate good solvers from poor solvers? What conditions affect adult performance?

Siegler and Klahr

Siegler and Klahr (1982) hypothesized two points at which difficulty in inducing the p-m might occur. First, when subjects were instructed to try to find a general rule to predict balance situations, they did not realize they were looking for a mathematical rule. That is, they failed to recognize the quantitative nature of the task and, therefore, failed to encode distance numerically. Second, choosing the correct algebraic equation from the many possible ones is too imposing a task without some type of external memory aid for disconfirming large numbers of incorrect hypotheses.

In this study, 13- and 17-year-olds, equated for lack of Rule IV (p-m) knowledge, were tested in four balance beam training conditions. In all conditions, subjects were presented with a sequence of 18 balance-scale feedback problems and asked to make a prediction for each, one at a time. In the external memory aid condition, subjects were provided with a sheet of paper with schematic representations of each problem and its outcome. In the quantified encoding condition, rather than simply asking subjects on each trial "What do you think the beam will do," the experimenter asked "Three weights on the third peg versus two weights on the fourth peg; what do you think will happen?" Subjects in the third condition received both

external memory aids and quantified encoding, and those in the fourth condition received neither aids nor encoding training.

The findings were that the 13-year-olds needed both distance and encoding aids to advance to Rule IV, while, in contrast, the 17-year-olds benefitted from either aid. This suggests that there are age differences in the number, rather than the kind, of aids needed for a scientific induction task. The important point for the current research is that, with older subjects, accessing all the necessary elements for inducing the p-m rule may be triggered by giving subjects a hint or aid for any one of the elements.

### Hardiman

Hardiman (1983) gave college undergraduates who did not exhibit Rule IV behavior in a pretest, balance training using an experimental situation very similar to Siegler's. However, in contrast to the Siegler and Klahr studies, in which few older subjects advanced unaided to Rule IV behavior, the subjects in the Hardiman study all met the criterion for Rule IV use during the training session.

Why were Hardiman's subjects successful while Siegler's were not? In the following section, the Hardiman study is compared to the Siegler paradigm.

Hardiman-Siegler Similarities. In both the Siegler and Hardiman studies, subjects were run individually. Training was administered using similar concrete balance beams (see Figure 1) and weights of equal size and shape, with the only difference being that Siegler used metal rings on pegs evenly spaced from the fulcrum and Hardiman used wooden blocks placed on markings evenly spaced along the horizontal surface of the beam.

In both studies, before the training began, subjects were told that they were to predict the action of the beam for each problem and that they were to try to determine a rule that would allow them to make accurate predictions for all problems. The beam was supported in a balanced state while the experimenter placed the weights for a problem on the beam. After the subject made a prediction, the beam was released and the subject was allowed to observe the beam's action. Then the next problem was presented.

In both studies, easier two-stack problems were presented first, moving from weight, distance and balance situations to the more difficult conflict problems. Changes in problems were made by one unit progressive changes in either weight or distance, so that subjects could see the effects on the balance state of these incremental changes. These problems were followed by a group of conflict, multi-stack problems, none of which had any direct



relationship to the immediately preceding problem, and each of which apparently required a partial or complete displacement of the weights from the previous problem. That is, presenting new problems in this more difficult set required, in essence, erasing one problem in order to present the next. Hence, there was no obvious continuity between these problems and each became an individual and isolated learning situation.

Hardiman-Siegler Differences. A closer inspection of these studies brings to light some important differences in sequencing and manner of presentation of the problems that may well account for the success of Hardiman's subjects in inducing the rule.

While Hardiman began with the same simple two-stack problems as Siegler, transforming one problem into the next via one-unit changes in weight or distance, when she presented subjects with Siegler's more complex and difficult set of multi-stack problems, if subjects became confused or were unable to develop integrative rules, she then provided them with additional related problems. This was in contrast to the apparently arbitrary order of multi-stack complex problems used extensively by Siegler.

It is not possible to specify here in detail what orderings Hardiman used to help her subjects, since the experimenter added a different set of problems in response



to each subject's particular confusion. But the fact that complex problems were presented in some non-arbitrary relationship to each other in this study, and not in Siegler's, suggests that ordering problems, i.e., sequencing them so that the more difficult problems, like the simpler ones, also relate to immediately preceding problems, has an effect on induction.

An additional concern with regard to the problems is that Hardiman's subjects had nearly twice as many (up to sixty and, in some cases, as many as eighty) examples to learn from as did Siegler's. Perhaps the larger learning set alone would account for her subjects' success. To look at this, in a pilot study (using a computer graphics representation of the balance beam), we presented subjects first with Siegler's sequence of easy problems, then with double the number of complex, conflict problems. This did not seem to help subjects make the transition to a more computational approach, much less induce the torque rule. On the contrary, subjects seemed to become more and more overwhelmed and confused by the sheer amount of confirming and disconfirming evidence from such a large number of problems for which they had no systematic means to form and test hypotheses. Therefore, while number of learning examples may well contribute to ease of learning, the organization or sequencing of the examples seems a more

Important determining factor. The hypothesis, then, that was tested in this research was whether the sequence in which balance problems are presented would have an effect on subjects' ability to make the transition from non-computational to computational rules, and subsequently to the p-m rule.

There are some further differences between the two studies that deserve consideration. First, Siegler's subjects ranged in grade from K through twelfth grade, while Hardiman used only college freshmen. However, if we look just at Siegler's older subjects, it seems reasonable to assume that these eleventh and twelfth graders were cognitively on a par with, or very close to, Hardiman's college freshmen. Hence, it is fair to compare the two subject populations' performances on this task.

A second, more important issue concerns the role active hypothesizing plays in learning. Siegler's subjects merely observed the problems and their outcomes, while Hardiman's subjects were asked to "think aloud" and to explain their reasoning for each prediction before observing the beam's resulting action. There is reason to believe that verbalizing predictions and the rationale for each prediction may help solvers to be more systematic and precise in how they think about the balance problems and to generate more and better hypotheses. Consequently, active

hypothesizing and verbalization may lead to faster learning. In the current study, subjects were asked to think aloud, as they were in the Hardiman study, so that a more direct comparison with Hardiman could be made. In future research, it will be interesting to compare thinking aloud subjects with non-verbalizing subjects.

### Major Findings on Adult Reasoning

The proposed research is concerned with two questions about balance beam rule induction: 1) what are the reasoning rules and strategies people use as they attempt to induce the general rule for the beam; and 2) what conditions make the task easier for subjects to perform? With regard to the first question, Hardiman, Pollatsek and Well (1986) point to three major findings from Hardiman (1983): 1) most of her subjects seemed to develop and use a quantitative rule of limited generality -- the ratio rule, in particular -- prior to using the p-m rule; 2) many subjects during the first part of the learning session gave evidence of using rules that did not involve encoding distance numerically; and 3) most subjects seemed to employ specific information about previously experienced configurations in making decisions about some of the balance problems. Hardiman, et al. conclude that Siegler's hierarchical decision rule models

are inadequate to describe a subject's reasoning processes in this task.

Rules of limited generality and distance encoding. At least two-thirds of Hardiman's subjects verbalized some type of quantitative rule prior to indicating use of the p-m rule. Most hypothesized rules had extremely limited applications, with the exception of the ratio rule, which will predict whether the beam will balance or not for all two-stack problems. However, subjects were not always consistent in their use of the ratio rule. For instance, many subjects used ratio reasoning for 2:1 ratios (e.g., 0004/0200 or 0002/0100), but did not use it on other ratio configurations (e.g., 3:1). One interpretation for this inconsistency is that subjects did not initially learn the general form of the ratio rule. Other data suggest that subjects did not always encode distance accurately enough to predict correctly (Hardiman, Pollatsek & Well, 1986). In fact, many subjects early in the session indicated that they were using distance only to make ordinal (more, less, equal) or rough perceptual ('it looks about the same.') judgments. On the average, Hardiman's subjects did not encode distance numerically (e.g., '2 blocks at the 3rd space ...') until the 28th trial.

Comparisons with previous problems - Instance-based reasoning. On 15% of the problems, Hardiman's subjects

made clear statements that they were basing their predictions on a comparison with a previous problem or class of problems. The comparisons were of three types: comparisons to previous problems that differed by a one-unit change in weight or distance; comparisons with previous problems that differed by more than a one-unit change; and comparisons with a known ratio configuration. Subjects used the comparisons to determine how a critical difference would affect the action of the beam or, in some cases, subjects judged a previous problem and a similar, but not identical, current problem to have the same outcome -- a qualitative, rather than quantitative, judgment.

Instance-based reasoning helps or hinders? What we want is for subjects to induce a general rule from experience with individual problems. References to previous problems that are related by one unit differences in weight or distance can help subjects to see the effects on the balance state of these single transformations, to realize that distance values are important numerically (vs. ordinally) because the weight and distance values must be combined numerically, and to illustrate a pattern of weight-to-distance relationships. In other words, previous problems are helpful when they make the algebraic nature of the task more salient and the systematic effects of changes in weights and distances easier to keep track of.



References to previous problems hinder the rule Induction process when information from previous problems is used only to help predict the current problem. References to previous problems may sidetrack the solver from the real task of finding a general rule for the whole system to the more immediate task of isolated predictions.

In sum, it is clear that adult subjects hypothesize and test a variety of qualitative and quantitative rules and strategies prior to inducing the p-m rule. Further support for this conclusion was found in a pilot study for the proposed research in which many subjects verbalized their use of several special-purpose strategies for transforming complex conflict problems into configurations for which they already had established a reliable rule. One particularly common (and easy to document) strategy was the use of comparisons to previous problems. In the current research, subjects' verbalized reasoning for each problem provided explicit evidence of this and other strategies that were used as alternatives to an algebraic rule. It is also clear that encoding of distance numerically is an essential key to success in this task.

The Siegler and Klahr data suggest that the conditions in which subjects perform better are those in which the numerical values of distance are made salient and the memory



load is reduced by the use of external memory aids. In the Hardiman study, the insertion by the experimenter of helpful problems relating to a subject's incorrect hypothesis or confusion may have sequenced the more difficult problems such that the multiplicative effects of one-unit changes in weight and distance became more understandable, resulting in induction of the multiplicative p-m rule. On the other hand, sequences of closely related problems could help subjects to make predictions about only a current problem, while diverting them from looking for one general rule. Active hypothesizing and "thinking aloud" may also be a key factor in helping subjects to be more precise and systematic in their reasoning. The role that instance-based reasoning plays in the induction task has yet to be determined.

# CHAPTER III

## PROPOSED MODELS OF BALANCE BEAM REASONING

### Rationale

#### Subject perception of task goal

The balance beam is a dynamic physical system involving weight and distance variables. The goal in the balance beam task is to make accurate predictions about the behavior of the system, no matter how the weights and distances are configured. The key to the goal is knowing the general rule (the p-m rule) about the combined effect of the variables on the system.

One claim of the current research is that performance in the balance beam task is influenced by a subject's perception of the goal of the task. In the Hardiman study (and as was the case in this study), naive subjects were presented with balance problems and instructed to make predictions, after which they could observe the action of the beam. They were also told that they were to try to find the rule that would allow them to predict accurately for any situation on the beam.

From these instructions, subjects might focus on very different task goals -- that is, on very different criteria for success in the task. A subject who focuses on

point is that the rule-finding solver knows that the one general rule can replace any set of more limited rules.

### Assumptions for the Proposed Models

Given the findings from previous research and from the pilot study for this research, the following models of adult reasoning in the balance beam rule induction task were hypothesized. The models are based on four assumptions about the adult solvers in this study:

- 1) adult solvers know how to encode distance numerically;
- 2) adult solvers know the difference between a qualitative rule and a quantitative, arithmetic rule;
- 3) adult solvers know the basic math operations (add, subtract, multiply, divide) that make up the set of possible algebraic combining operations that can be used to generate quantitative rules; and
- 4) some subjects will bring to the task better (more systematic and practiced) general problem-solving skills than others.

The models also take into consideration a solver's potential vacillation between two general problem-solving approaches -- hypothesize and test, and prediction by

prediction alone will ignore the notion of a general rule and, instead, will devise rules of limited generality -- a set of configuration-outcome rules for various subsets of balance problems. While it is hypothetically possible for a subject to develop a complete set of special-purpose rules for the system, it does not seem likely, at least in the case of balance beam configurations, that without some external means of organizing and classifying the numerous types of configurations, a person could be sure if or when he had a complete set of rules.

If, however, a subject focuses on finding the general rule for the system and knows that the rule is the means to the goal of one-hundred percent accurate predictions, he will have a very different approach to the information from example problems. A subject looking for a general rule knows that it must be one rule to cover all possible situations.

One particularly interesting finding from subjects' verbalizations in the pilot study was that it was possible to distinguish between those problems in which subjects were focused on correctly predicting the current problem as an end unto itself, in which case they used previously learned rules or strategies or invented new strategies to make a prediction about the current problem, and those problems for which subjects were clearly focused on finding the one

general rule. The distinction is between two general problem-solving approaches: a hypothesize-and-test approach used to generate and test candidates for the one general rule, and a classification approach used systematically to type problems by their configurations of weights and distances and to store the learned outcome with each class.

It appears, then, that reasoning at different points in the balance beam learning experience is guided by either of these two general problem-solving approaches and that a subject may follow one approach all the way through the training or he may vacillate back and forth between the two. This will prove to be a distinction important to the stated ultimate goal of modeling the changes in reasoning of successful and unsuccessful solvers as they proceed through the task.

It is important to state here the assumption that knowing the general computational rule does not preclude use of simpler rules in simpler balancing situations. Siegler and Klahr (1982) tested adults who knew the p-m rule for use of both qualitative and quantitative rules. They found that these subjects used qualitative comparisons of weight and distance values first and computed torques only when the qualitative comparisons did not yield a clear answer. "Why compute when you can compare?" seems a reasonable and adaptive approach on the part of the solver. The important



classification - each of which reflects a different perception of the goal of the task. That is, the instructions both to make predictions and to find the one rule that will allow 100% accurate prediction will cause subjects to adopt one or the other of these approaches for each problem as it is presented.

Because success in this study is defined as the ability to explicitly state and use the p-m rule, it is hypothesized that the subjects who will be most likely to succeed by this definition will be those who focus predominantly on a hypothesize and test approach for each problem throughout the training.

#### Proposed Model: Rule-Focused Solver

When a solver focuses on a rule-finding approach, the model shown below predicts that, for a particular configuration on the balance beam, he will first differentiate between simple weight/distance/balance situations that can be predicted using qualitative (ordinal) rules (e.g., if there is greater weight and greater distance on the left, it will tip left) and those problems in which a conflict situation exists (i.e., the greater weight is associated with the lesser distance). If the problem is a



conflict, a quantitative approach will be invoked. This switch by default from qualitative to quantitative reasoning implies a previously learned awareness that, at the top level of reasoning, qualitative and quantitative reasoning are the two possible alternatives to explore.

Once in a quantitative mode (i.e., having isolated the problems for which a qualitative approach is inadequate), the model predicts a rapid identification and quantification of the relevant variables, in this case weight and distance, followed by a simultaneous attempt to understand the physical and numerical relationship of the weight and distance variables. By combining physical and numerical information about the problem set, the successful solver's task then becomes a somewhat mechanical matter of hypothesizing and testing the list of possible math combining operations (add, subtract, multiply or divide) until the multiplicative p-m rule is found.

## THE MODEL

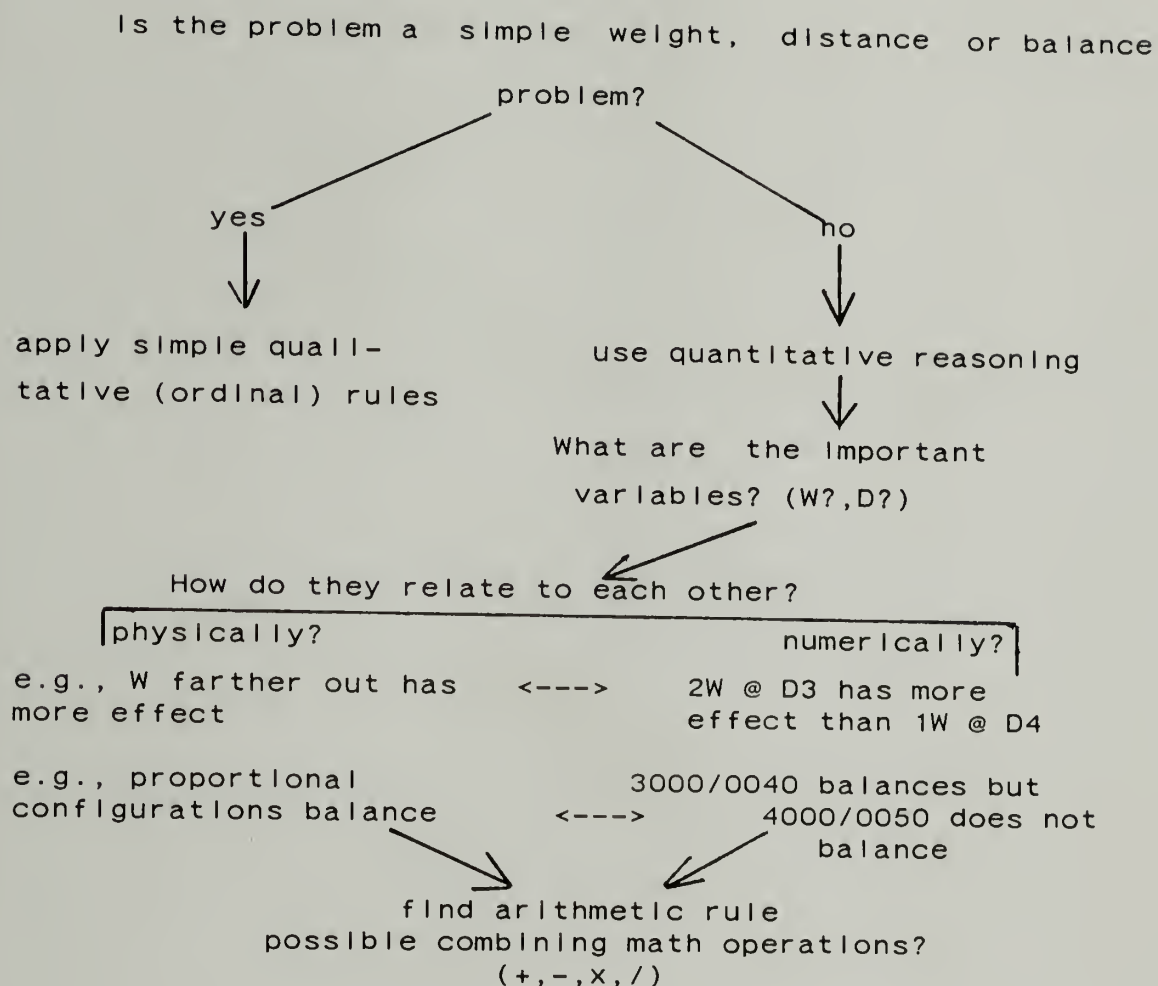


Fig. 1  
Proposed Model: Rule-focused

### Predictions for the Successful Rule-Focused Solver

The predictions for a subject taking this approach are as follows:

1. early numerical encoding of both W and D variables

2. early statement of a ratio rule
3. explicit statements indicating the subject knows he is looking for a "mathematical rule" or a rule to combine W and D
4. few retests of combining rules that have been disconfirmed
5. rapid changes, after disconfirmation of the use of one math operation, to test of another of the possible math operations until the correct rule is found (i.e. predictions are consistently correct).

The successful solver using this approach will systematically identify and quantify the variables, and proceed to test each possible math operation until he finds the solution. The unsuccessful solver using this approach will fail if he does not quantify either weight or distance, fails to recognize numerical ratios, fixates on refining or adjusting a disconfirmed rule, or if he fails to test enough math operations to discover the correct one. The unsuccessful solver will also be one who vacillates between a rule-finding approach and a prediction approach.

## Proposed Model: Prediction Subject

### Prediction as a categorization process

An induction task is a task in which one identifies the important variables or dimensions of a set of exemplars and then attempts to find a rule about the variables or dimensions that will describe the entire set. In contrast, if a subject in the balance beam task takes an approach whereby he groups problems according to certain characteristics of their weight and distance configurations into subsets with known balance or tip outcomes, and then makes a prediction for a new problem on the basis of its similarity to problems in one of the subsets, then one can view the approach as a process of categorization.

The process involved in categorization of objects, for instance, is one of abstracting a rule about the features and their dimensions. In a well-defined category, the rule will be the same for all members. In a "fuzzy" or ill-defined category, no one rule will describe all members (Mervis and Rosch, 1981).

Abstracting the rule for a physical system such as the balance beam is somewhat different. Examples of an object category that are classified by features and dimensions are static; the rule describes only their appearance. Exemplars of a physical system, on the other hand, are instances of

the dynamics of the system; the rule for the system describes the interactive effects of the variables.

If a subject in the balance beam task perceives the system to be too complex to determine a single rule for prediction, he will resort to an approach whereby he classifies and predicts new problems in the manner described above, forming in the process many classes of problems and, hence, a growing set of prediction rules.

#### Predictions for the Prediction Solver

1. use of many special purpose rules or categories
2. many references to previous problems differing only by one or two critical changes in weight or distance--particular in the highly sequenced conditions.
3. many retests of disconfirmed rules

## CHAPTER IV

### THE PRESENT STUDY

#### Goals and Research Questions

Previous research has shed some light on developmental issues, identified some of the variables that affect performance, and documented some of the intermediary rules and strategies subjects use in this induction task. The goal of the present research was to further investigate the reasoning processes of adult problem solvers in a specific induction task and to characterize their general and domain-specific problem-solving skills and strategies as they proceeded through the task. The approach in the past has been to look at successful and unsuccessful solvers collectively, and across age groups. The approach for the current research was to look only at adult subjects and to differentiate between high-ability subjects (solvers) and low-ability subjects (non-solvers), based on the hypothesis that a model of the successful solver will reveal general problem-solving skills and strategies brought to bear on the specific task that will not be evident in a model of the unsuccessful solver.



The research challenge for the balance beam task is to discover the process by which people learn from examples to abstract or induce the rule that defines the system. But the bigger question is how do successful solvers know how to learn - what makes them good scientists, diagnosticians or detectives - and how are their learning strategies different from those of the unsuccessful solver? The applied question, of course, is one of instruction. How can the low-ability solver be taught better problem-solving skills?

Therefore, this research addressed the general issues stated above by answering the following questions about acquisition of balancing knowledge:

1. How do subjects perceive the goal of the balance beam task?
2. What rules and strategies do people use while learning to induce the product-moment rule?
3. Is one pattern of reasoning more likely to lead to success than any other?
4. What task conditions influence the performance of solvers?

Information and data pertaining to the first three questions came primarily from subjects' verbalizations as they made and explained predictions for each balance

problem. The rationale and discussion of these questions were presented in depth in the previous section.

Data for the fourth question occurred as a result of comparing subject performance across four problem-sequencing conditions and two different manners of presentation. A discussion of why sequencing of problems might be an important variable in the balance task and the rationale for the sequencing and presentation conditions is presented in the next section, followed by a brief discussion of the use of computer graphics in this study.

### The Problem Set: Considerations and Rationale

#### Is sequencing important?

Siegler's subjects did not solve the balance beam task; Hardiman's did. One interpretation of the data is that the effective factor in the Hardiman study was that, when she interjected helpful problems to confused subjects, she created a sequence of related problems for subjects to observe. If sequencing is the key, there should be a difference in performance between subjects who are presented with sequences of highly related (by one or two unit changes in weight or distance) and those who are presented with the same set of problems, but randomly mixed.

Comparing the highly sequenced condition with the mixed condition also allows us to look at the difference between having the easier, non-conflicts (which can be predicted with non-computational rules) grouped together at the beginning of the set, and interspersing them throughout.

The current research, therefore, tested the effects of sequencing in order to determine whether the relationship of the current problem to the immediately preceding problems is an important factor in the rule induction process? At the same time, through verbal protocols of subjects' active hypothesizing and verbalized reasoning, the research attempted to more fully identify and classify subjects' reasoning rules and strategies.

Ordering the set of balance problems has three aspects: 1) composition of the problem set, 2) the order in which they are presented, and 3) the manner in which they are visually presented.

#### Composition of the problem set

The goal in the balance beam task was for subjects to induce a general rule for prediction the action of the beam. That rule, the p-m rule, is a quantitative, algebraic computation. When presented with the balance beam problems, subjects in past studies generally made their initial

predictions on the basis of non-computational (rough perceptual or ordinal) judgments about the weights and distances on the beam. In order to eventually arrive at the p-m rule, subjects had to learn that the qualitative rules have limited application, as do most of their ensuing hypothesized quantitative rules. The question with regard to the kinds of problems to be included in the learning set for this study was whether or not the set should include the simple problems that can be predicted correctly on the basis of non-computational (ordinal) judgments - all the non-conflict problems - and, if so, in what proportion?

Do these easy-to-predict non-conflicts help people to notice that there are two kinds of problems, non-conflicts and conflicts, and thus to realize there are two solution approaches, non-computational and computational? Is this a distinction that must be made salient to the solver early in the learning process? Or do problems that can be solved with a non-computational rule garden-path subjects into fixating onto a non-computational mode of rule induction, to the exclusion of any computational considerations?

The assumption for this study was that the non-conflicts provide an essential contrast to the conflict situations, that they help subjects to differentiate the types of problems for which a computational rule is needed. They were therefore included in the learning sets. The

question then arose as to how best to include them so as to provide the contrast while, at the same time, not mislead subjects into thinking the qualitative rules that will solve them are the typed of ruled that will solve all problems, given a little refining. This is a question of sequencing and is discussed in the following section.

#### Order of Problems - What sequence should we test?

Easy to hard? To ask what sequence we should test is really to ask what makes a "good" sequence of balance problems. What reason do we have for picking one sequence over another as better for learning? The most common view is that problems should be presented in an easy-to-hard ordering (Lesgold, 1983; Karmiloff-Smith, 1984). This is based on the theory that a clear understanding of the usefulness and limitations of simple, qualitative rules for simpler problems is a necessary conceptual foundation for moving on to more complex rules for more complicated problems. It is also a view that mirrors the reasoning process observed in both children and adults. That is, most people begin the problem-solving process by hypothesizing qualitative, non-computational rules (often a sufficient as well as efficient strategy) and will switch to a quantitative, computational approach only when they become convinced of the limited generality of their non-



computational rules (Karmiloff-Smith, 1984 ; Welsberg & Alba, 1981; and as observed in Siegler, 1976, and Hardiman, 1983). Thus, in this view, in order for the solver to become aware that he needs a new approach, he must have sufficient opportunity to exhaust the list of inadequate rules generated by the old approach.

For the balance beam task, easy-to-hard sequencing has two overlapping dimensions: 1) the overall transition from simple to complex (two-stack to multi-stack) discussed above, and 2) the transition from non-conflict weight, distance and balance problems to conflict situations within first the simple problems and then again in the later complex problems.

The second aspect of sequencing reflects the order of decision rule acquisition observed by Siegler (1976). He found that subjects typically first test the qualitative, non-computational Rules I, II and III. Those using Rule I rely solely on weight cues for prediction; those advancing to Rule II consider distance, but only when the weights are equal (including simple balance problems such as 0300/0030); and those advancing to Rule III consider both weight and distance, and can accurately predict all weight, distance and non-conflict balance problems. In addition, they recognize a conflict situation, but soon find their existing rules are inadequate for these situations in general. For



consistent correct prediction of conflicts they need the p-m rule. From this view, degree of learning difficulty proceeds from weight problems to distance problems (including non-conflict balance problems) to conflict situations.

But it is the transition from non-computational to computational rule use that occurs during experience with conflict problems that is of most interest (which is not to say that this transition never occurs with other types of problems, but that it most often occurs with conflicts). Is there a 'good' sequence for conflicts that would facilitate the transition to a computational approach and ultimately lead to induction of the p-m rule? Since prior studies have not used any principled order for conflict problems and since there are thousands of sequences of conflicts that could be tested, this research broke ground not by defining a particular sequence, but by imposing in one condition-- the highly sequenced condition -- the general constraint that each conflict problem, simple and complex, be presented as a member of a group or progression of problems, each one related to the next by one or two unit changes in weight and/or distance, rather than as isolated situations.

Mixed? On the other hand, beginning with a large group of easy problems may hinder the reasoning process (Sweller,

Mawer and Ward, 1983) by encouraging solvers to believe that the approach that was successful with earlier problems is the correct approach for all problems. It might also misdirect attention to analyzing individual situations in isolation from the others rather than to abstracting an overall analysis of all of the problems (Sweller and Levine, 1982). In other words, a sequence of easy problems at the beginning of the learning session may lull subjects into a false sense of adequacy for their qualitative rules with the result that they attempt to confirm them with each problem, but do not use the problems as a whole to collect more general information. It was in the midst of the later, more difficult problems that several subjects in the pilot study for this research became aware of the need to switch approach, but by then the problems were so complex that subjects indicated that a systematic analysis at that point was overwhelming. This led to the alternative hypothesis that if easier problems were mixed with the harder ones, subjects would be alerted from the start to the need for a more general, quantitative approach.

#### Manner of Presentation

If sequencing of problems is important (i.e., problems that relate to each other progressively by one or two unit changes in the weight or distance variables are grouped

together during learning, then how might the salience of the relatedness within a group or progression be increased? One possibility is for subjects to see within a progression the transition from one problem state to the next actually take place - for them to witness the metamorphosis as it occurs, rather than to see each problem presented as an entirely new configuration. In effect, this is the difference between completely clearing the weights from the beam between each problem in a progression, and starting with an originating configuration and graphically adding, subtracting or moving the weights to form, step by step, new configurations, without clearing the beam.

Intuitively, it does not seem possible that the discontinuous, clear-the-beam presentation would be better for learning than the continuous metamorphosis. At the least, the method of presentation might have no effect at all. What seems more likely is that actually viewing the changes would make the numerical relationships between the weights and distances more salient to the solver by a) providing a visual trace of changes and effects, and b) reducing memory load for previous problems.

In order to test this hypothesis, there were two identical highly-sequenced conditions with the exception that in one the progressive changes could be watched as they occurred and in the other the weights were removed

(graphically erased) from the beam after each and every problem.

### Problem Sets: The Current Study

#### Sequences: composition and order

In this research the sequences to be compared were defined as follows:

Condition I - Highly sequenced/transformed: sequencing in which problems move from simple to complex at the same time that they progress from weight/ distance/balance situations to conflict situations, with the further stipulation that all problems will be presented as members of a group or progression of problems, each related to its adjacent problems by one or two unit changes in weight or distance. Within a progression, changes in weights or distance are made as transformations of the existing configuration. The beam is cleared only between progressions.

Condition II - Highly sequenced/clear beam: Identical to Condition I, but the beam will be cleared between each problem.

Condition III - Mixed: a mixed-order presentation of the same problems as in I and II.

Condition IV - Hardiman: (a replication of Hardiman's (1983) sequencing of problems) a sequence beginning with two-stack, simple weight, distance and balance problems, moving to two-stack conflicts and then to multi-stack conflicts. The two-stack problems were sequenced such that each was related to the next by a one-unit change in weight or distance, but the multi-stack conflicts each had no obvious relationship to the immediately preceding problems. In order to have a total of 80 problems in the set, the smaller set of multi-stack conflict problems used by Hardiman was randomly recycled to make up the difference in the size of the problem set.

The complete sequence for each condition is shown in Appendix I.

#### Concrete Beam vs. Computer Graphics Representation

It would be technically advantageous if the research on rule induction could make use of the uniformity of presentation and data collection abilities afforded by computer presentations. The question arises as to whether, for a physical concept such as balancing, subjects need an actual physical balancing apparatus in order to 'get a feel' for the effects of weight and distance changes and interactions. This might be true if we only wanted subjects



to intuit the effects of these changes in the variables. However, since it is a precise algebraic rule that they are to induce, an abstract formalization of observed changes in weight and distance values, it would seem that a less concrete, somewhat more abstract graphic representation such as the computer produces might trigger or cue a more abstract, quantitative approach to the task.

Furthermore, if subjects perform the same or better in this task using the computer graphics (in comparison to using the concrete beam), then the flexibility of the computer suggests ways to more easily extend the research in new directions, while maintaining a high level of experimental control.

And, finally, it is of importance to find out to what extent the computer can be used as an effective and dynamic teaching tool.

Therefore, subjects in this study were shown a computer graphics representations of the balance beam problems, a description of which follows in the Methods section.



## CHAPTER V

### METHODS

The rule Induction study consisted of two phases conducted in two sessions: a paper-and-pencil pretest phase and a computer graphics representation learning phase.

\*NOTE: The pretest for the Transfer Task (see Section II) was given at the same time as the balance beam pretest. Only subjects who were identified in the pretest to be both non-balancers (i.e. did not indicate knowledge of the product-moment rule) and non-calculators (i.e., were unable to solve weighted mean problems, as described in Section II) were used in either the Rule Induction Study or the Transfer Task. Only the Rule Induction Methods are discussed here.

#### Pretest

##### Subjects

Students enrolled in psychology classes at the University of Massachusetts were offered bonus credit for participation in the pretest.

##### Problems

The pretest (Table B6) consisted of 12 schematic line drawings of a balance beam, each with a different configuration of weights and distances. There were three

simple weight/ distance/ balance problems (Problems 11, 3, and 6, respectively), and nine multi-stack conflict situations: three conflict-weight (Problems 1, 5, and 8), three conflict-balance (Problems 2, 7 and 10), and three conflict-distance (Problems 4, 9, and 12). Conflict situations are those in which the greater weight on one side is associated with the lesser distance from the fulcrum. Conflict-weights tip to the side with the greater weight, conflict-distances tip to the side with the greater distance, and conflict-balance situations balance. The problems were all on one page with "L" "R" "B" under each one. Written instructions asked subjects to decide for each problem whether the beam would tip right, left or balance, and to circle their answer.

### Procedure

The pretest was administered to groups of 10 to 20 subjects. Subjects were given a booklet containing instructions, a demographic questionnaire, the balance beam pretest, and the weighted mean transfer task pretest (see Section II) and allowed to proceed at their own pace. No feedback was provided.

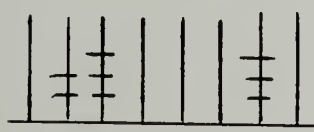
### Analysis

Subjects were classified as non-balancers if they made

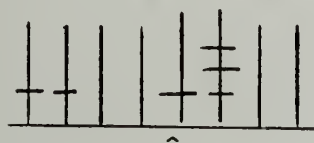
Table B6

## Balance Beam Pretest

The pictures below are illustrations of balance beams with different sets of weights on each beam. Each "-" mark represents an equal weight. You are to determine whether each set of weights would balance if it were placed on an actual balance beam. For each picture, mark B if you think it will balance, L if it will tip to the left, and R if it will tip to the right.



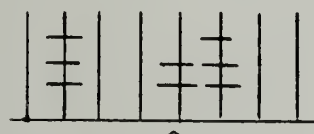
1)  
L      B      R



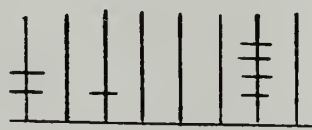
2)  
L      B      R



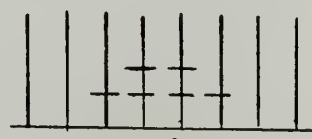
3)  
L      B      R



4)  
L      B      R



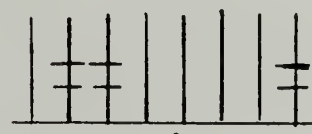
5)  
L      B      R



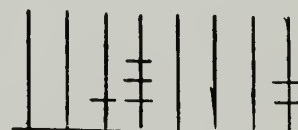
6)  
L      B      R



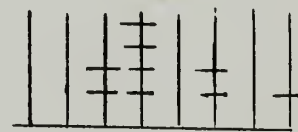
7)  
L      B      R



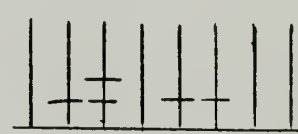
8)  
L      B      R



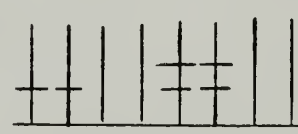
9)  
L      B      R



10)  
L      B      R



11)  
L      B      R



12)  
L      B      R

Incorrect predictions on any of the three simple problems or if they correctly predicted the simple balance states but made errors on three or more of the conflict problems.

### Training

#### Subjects

Only subjects who are classified as non-balancers and non-calculators (see Section II: Methods for definition of non-calculators) were used in the study. These subjects were given bonus course credit and \$5 for participation in the balance training and transfer phases of the experiment.

#### Materials

A color-graphics representation of a balance beam was shown on a Zenith color monitor. The picture of the beam was a rectangle measuring 6" x 1/4" with a small triangle for a fulcrum placed under the midpoint. Distance from the fulcrum was denoted by four unit marks drawn at evenly spaced intervals on each side of the fulcrum. The weights were represented by 1/8" squares which appeared centered over the distance marks in stacks of up to six weights. For each of the four sequencing conditions, the configurations, sequences and manner of presentation were entirely under the control of the computer program. A four-key board, placed in front of the monitor, was used by subjects to record

their responses: one key each for left, right, and balance predictions, and one key to advance to the next problem. Feedback as to the correctness of the subject's prediction for each problem appeared on the screen in the form of a printed statement (e.g. "Yes (No), the beam will tip left.") along with the appearance of arrows pointing down under the side that would tip down or no arrows if the beam would balance. A computer printout (in another room) recorded each problem configuration, the correct response, and the subject's keyboard response.

#### Problems (Appendix I)

There were four sequencing conditions of eighty problems each. In Condition 1 (Highly sequenced and transformed), problems increased in complexity from a series of 32 two-stack problems to a series of 48 three- and four-stack configurations. Within the two-stack division, problems increased in prediction difficulty, beginning with simple weight, distance and balance situations, which can be predicted on the basis of ordinal weight and/or distance cues (i.e., weight or distance or both are equal on opposite sides of the fulcrum, or there is both greater weight and greater distance on one side) and proceeding to the three types of conflict problems, which require the use of the p-m rule for consistent accurate predictions. Within the multi-

stack problems, the majority were complex conflict situations such as 0024/0204 and 0300/2300.

Each problem in Condition I was a member of a subset or progression of three to six problems, with each problem in the progression related to the next by a one or two unit change in weight or distance. Problems within a particular progression were related in such a way as to illustrate the progressive effects of the various one or two unit changes.

In addition, weight and distance changes within a progression in Condition I were made by adding or erasing only the relevant squares, leaving the unchanging squares in place. The existing configuration was thus transformed into the next configuration, with the changes completed on one side before they were begun on the other. The beam was cleared of all the weights only between progressions.

Condition II (Highly sequenced-clear beam) was identical to Condition I except that the beam was cleared of all weights between each problem.

In Condition III (Mixed), the problems in Condition I were presented in a mixed ordering that was the same for all subjects in this condition. No problem had any direct relationship to the one before it and the beam was cleared of all boxes between problems.

In Condition IV (Hardiman), the sequencing was modeled after that used in Siegler's (1976) observation condition



and in the Hardiman (1983) study. The first 24 two-stack problems related progressively to each other and were presented without clearing the beam between them. They began with simple weight, distance and balance situations, and were followed by three sequences, each of which will begin with two equal stacks of weights at unequal distances from the fulcrum. Blocks were added one at a time to the stack closer to the fulcrum until the beam balanced and then tipped in the opposite direction, after which the beam was returned to the balance configuration. The remaining 56 problems were a mixed ordering of the multi-stack conflicts used by Hardiman (taken from Siegler, 1976), recycled twice to total 80 problems. The beam was cleared between each of these later problems. As in Condition III, all subjects in this condition saw the same ordering of problems.

### Procedure

Before the session began, each subject was asked to consent to the session being audio tape-recorded. Subjects were run individually. They were seated in front of the monitor with the keyboard on the table in front of them, the microphone to one side, and the experimenter seated behind. In order to minimize subject-perceived feedback from the experimenter, subjects were asked to remain facing the microphone in order to insure good voice pickup. This

prevented them from turning to check the experimenter's reaction to predictions and reasonings.

Subjects were told that the computer representation of the balance beam would behave exactly like a real balance beam. They were told that they would see a series of balance beam problems and, for each, to make a prediction about whether the beam would tip left, tip right, or balance. They were also told that there is a general rule that will result in correct predictions for all of the problems and that they should try to determine what this rule might be. Subjects were asked to think aloud during the entire session, to state and explain their prediction for each problem and to tell the experimenter whenever they had an idea about what the rule might be.

The session began with a picture of the beam on the monitor. Subjects saw the first configuration of boxes drawn one box at a time on the beam. When the drawing was completed, the question "What will the balance beam do?" appeared above the beam. Subjects stated their predictions and were encouraged to verbalize why they made the prediction. After subjects entered a response on the keyboard (L, R, or B), arrows appeared pointing down under the side that would tip down, or no arrows appeared if it would balance. At the same time, feedback text appeared next to the arrows indicating "Yes (or no), the beam will

(tip left/tip right/balance)." followed by the text "Press the 'continue' key when you are ready to go on." The next problem was either a transformation of the preceding configuration or the beam was cleared and the new configuration drawn, according to the constraints of the four sequencing conditions detailed above. The session, thus, was self-paced and the only role of the experimenter was to encourage subjects to verbalize as precisely as possible the reasoning behind their predictions and any hypothesized rules. The sessions ended whenever the subject stated the product-moment rule as it applies to multi-stack problems or at the end of the 80 problems in a particular condition.

### Analysis

Three types of data were of interest. First, from the protocols across conditions, the types of reasoning rules used for predictions were identified, as well as the strategies used to transform complex problems into simplified situations on which previously learned rules can operate. Second, a comparison of performance in the three different sequences and the two manners of presentation was made. And third, between conditions there was a comparison of four rule induction behaviors: a) number of problems to encode distance numerically, b) number of problems to first

Indication of computational use of weight and distance values; i.e. subject combines (adds, subtracts, multiplies or divides) weight and distance numbers (as opposed to merely comparing magnitudes), including documentation of the use of ratio rules, c) number of problems to induce the p-m rule for multi-stack conflict problems, and d) number and types of references to previous problems.

## CHAPTER VI

### RESULTS AND DISCUSSION

There were two basic questions with regard to rule induction and the balance beam for this research: the first was whether success in inducing the product-moment rule was dependent on task conditions and the second was whether there were characteristic types or patterns of reasoning that differentiate solvers from non-solvers.

The findings will show that task conditions -- sequence and manner of presentation of learning examples -- did not have an overall effect on success. Rather, the indications are that success was dependent on the type of problem solving approach a subject took. Non-solvers tended to rely either on a nonanalytic, categorization approach or to generate only a limited set of rules for which they had poor testing skills. Solvers also tended to categorize problems by type and outcome, but only initially, and moved quickly to a systematic, analytic approach to hypothesizing and testing possible rules. The inference drawn by the author is that these differences in performance in this task are attributable to non-task related, previously learned general problem solving skills.

This study has also revealed, through verbal protocols of subjects' unaided active hypothesizing, a more extensive

set of balance beam reasoning rules and strategies than was found in previous studies and has given some further insight into the key role ratio reasoning plays in cuing subjects to the multiplicative nature of the product-moment rule.

Finally, the findings will clearly show the methodological limitations of presenting and analyzing the balance beam task in a forced choice paradigm.

This section is organized as follows: first, a brief discussion of the methodological issues; second, results and discussion of a) the findings from the task conditions manipulations, and b) performance on the different types of balance problems; third, a taxonomy of observed reasoning rules and strategies; and, fourth, a qualitative analysis of subjects' patterns of reasoning.

The section concludes with a discussion of the proposed models of solvers and non-solvers.

### Methodological Issues

There were two methodological issues in question. The first concerns the criterion for success in this task. In Hardiman's (1983) study, a subject was considered a solver when he or she had made five correct predictions in a row. In the current study, the criterion for success was a statement of the product-moment rule and correct application of it to three multi-stack problems. If, in this study,



success had been determined solely by the five-in-a-row correct criteria, fifteen out of eighteen solvers (83%) would have been considered successful prior to actually inducing the rule, and nineteen out of twenty-two non-solvers (86%) would have been considered solvers without ever inducing the rule. In Condition IV alone, the condition most like Hardiman's, four non-solvers and one solver met this criterion. Clearly, subjects given these sequences could meet such a criterion without solving.

The second issue concerns the limitations of Siegler's rule-assessment approach. His determination of the reasoning rule a child was using was based on a quantitative analysis of responses in a forced-choice paradigm. That is, subjects' patterns of predictions (tip left, tip right or balance) for the different types of problems were taken as evidence that they were using one of four rules with which to reason. In the current study, subjects were asked to explain their reasoning for each prediction. The findings of this study will show that when subjects are actively hypothesizing in this manner, their verbalizations reveal the use of a far more extensive set of reasoning rules than has heretofore been shown. The advantage of the more qualitative approach used in this study will be supported by the extensive list of rules and strategies, outlined and discussed below, taken from the verbal protocols.

# Task Conditions and Problem Types

## Sequencing

The motivation for presenting subjects with different sequences of problems was the author's hypothesis that a major factor in the success of Hardiman's subjects was the inherent sequencing of problems that resulted from the experimenter interjecting helpful, related problems when subjects became confused. Also in question was whether there would be a difference in performance between subjects who actually saw the one-unit progressive changes to problems in a particular progression and those who did not—Condition I vs. Condition II.

Note: The criterion for success in this study was that a subject be able to state and apply the product-moment rule.

Table B7  
Number of Solvers, Non-solvers and Atypical Solvers

n	Cond.	#Solvers	Pre #40 Early	After #50 Late	#No Solution	#Atypical
10	I	4	2	2	6	0
10	II	5	4	1	5	0
10	III	5	4	1	4	1
10	IV	4	4	0	5	1
		<hr/> 18	<hr/> 14	<hr/> 4	<hr/> 20	<hr/> 2

Overall, 18 (45%) of the 40 subjects induced the p-m rule. Of these, 14 (78%) solved early (before the 40th problem) and 4 (22%) solved late (after the 50th problem). Twenty-two (55%) of the subjects did not induce the p-m rule. Of these, 2 (5%) devised atypical systems that resulted in consistently correct predictions without cognizance or use of the p-m rule calculations. These atypical solvers form a distinct class of subjects and will be discussed as such.

Excluding the two atypical solvers, the different sequencing conditions had virtually no effect on rule induction success. Subjects performed no better in Conditions I and II, the identically highly sequenced conditions, than in the mixed Condition III. Nor was there any advantage in the Condition IV sequence that employed Hardiman's basic series of problems. Clearly, in the present study, there was no evidence that sequencing had any effect on whether or not a subject induced the p-m rule. It is also clear that whether or not the one-unit changes in weight or distance (Condition I vs II) were visually salient did not effect overall performance.

#### Problem Types: Are There Key Problems?

Since success in this study could not be attributed to sequencing or manner of presentation, perhaps there were key

problems or types of problems which, once understood by the subject, triggered product-moment rule calculations. The findings here suggest that there is a correlation between performance on the simple conflict-balance problems (SCBs) and eventual success in inducing the rule.

Two-stack configurations can be separated into three classes: simple weight and distance, simple conflict-balance (SCB) and simple conflict non-balance (SCNB).

Simple Weight and Distance Problems. All subjects-solvers and non-solvers alike - in all conditions were able to predict with nearly 100% accuracy all of the simple weight and distance (i.e. those 2-stack configurations in which the greater weight is associated with the greater distance on a side, or those in which the weight and distance on one side equals the weight and distance on the other side).

Simple Conflict-balance and Conflict Non-balance Problems. Simple conflict-balance problems (SCB) are two-stack conflict situations in which the ratio of the weights is inversely proportional to the ratio of the distances and thus balance (e.g. 0200/0300, 0200/6000); simple conflict non-balance problems (SCNB) are two-stack conflicts that do not balance (e.g. 0004/0020, 5000/0060).

Table B8

## Average Percentage of Errors on Simple Conflict Problems

## NON-SOLVERS

<u>% ERRORS</u>	<u>COND</u>	I	II	III	IV
ALL		57.0	58.0	49.0	25.0
SCBs		45.0	54.0	50.0	17.8
SCNBs		58.5	56.0	50.0	33.0

## SOLVERS\*\*

<u>% ERRORS</u>	<u>COND</u>	I	II	III	IV
ALL		45.8	51.2	21.1	28.4
SCBs		25.8	37.4	27.5	27.8
SCNBs		59.8	59.2	6.8	27.8

\*\* % errors pre-product moment rule induction

Table B8 shows the mean percent of incorrect predictions of solvers and non-solvers for all of the simple conflicts in general, and broken down into SCBs and SCNBs. Note that, for solvers, only those simple conflicts predicted prior to stating the p-m rule for 2-stack problems were included in the analysis.

To clarify, the first 37 problems in Conditions I and II, and the first 25 problems in Condition IV were 2-stacks



(simple weight and distance and simple conflicts); the remaining problems in each of these conditions are multi-stack. Consequently, subjects could, and often did, induce the p-m rule early in the 2-stack portion of the sequence. However, a subject had to also be able to apply the p-m rule to multi-stack situations in order to be considered a solver. Thus, a subject who induced the rule early in the 2-stack problems had to then routinely predict the remaining 2-stack problems until the multi-stacks came up in the sequence. Since interest is primarily in reasoning leading up to rule induction, only problems predicted pre-rule induction in either the two- or multi-stacks are included in the analysis.

In Condition III, where the two and multi-stack problems are mixed, the percent of incorrect predictions are based on the number of total conflicts, SCBs and SCNBs a subject predicted prior to inducing the product-moment rule.

Errors: All two-stack conflicts and two-stack conflict non-balance (SCNB). Table B8 shows that solvers and non-solvers within each of Conditions I, II and IV (the sequenced conditions) had about the same percentage of errors on the simple conflicts taken as a whole, and on the SCNBs alone. The overall fewer errors in Condition IV can



be attributed to the frequent presentation of incremental sequences with repeating configurations such as:

0020/2000  
0020/3000  
0020/4000  
0020/5000  
0020/4000

Having found that the first 0020/4000 balanced, it was easy for subjects to predict the next problem on the basis of "There's an added weight on the right, so it has to tip right." It was also easy for subjects to predict the second 0020/4000 by reference to the first occurrence. In the other conditions, there was not this repetition and close proximity of identical problems.

Condition III (mixed) presents another picture. Here non-solvers performed overall on the simple conflicts on a par with non-solvers in the other conditions (49% of the total simple conflicts wrong and 50% of the SCNBs wrong), but the solvers in this condition performed much better than either solvers or non-solvers in the other conditions, with only 21.1% incorrect predictions on the simple conflicts in general, and only 6.8% incorrect on the SCNBs.

This may simply be a reflection of the fact that the two-stack (simple) conflicts in this condition were interspersed with the multi-stacks and solvers just did not encounter as many two-stacks prior to solving as solvers did in the other conditions. This of course tells little about

how it was that Condition III solvers managed to solve in the first place.

Condition III was the only condition in which there was a within-condition significant difference between solvers and non-solvers in performance on the simple conflicts in general ( $t(8) = 3.203$ ,  $p < .01$ , 1 tailed) and the SCNBs in particular ( $t(8) = 4.7082$ ,  $p < .001$ , 1 tailed). This is, in part, consistent with the author's speculation that the mixed condition might keep subjects from becoming garden-pathed into focusing on the set of simple rules of limited generality that were frequently evoked by the concentrated array of simpler problems presented first in Conditions I, II and IV. The unanticipated finding here is that the mixed condition was helpful to solvers only, which suggests that certain subjects were better able than others to take advantage of being immediately confronted with the most complex types of problems.

Errors: simple conflict-balance (SCB). What is of particular interest in Table 13 is the consistent difference between solvers and non-solvers in Conditions I-III on the SCBs (e.g. 0200/0300, 0100/3000, 0020/4000). An analysis of variance did not reveal a significant effect of solving on SCB errors ( $F(1,16)=3.09$ ,  $p < .098$ ), but the trend evident in the data of Table B8, taken in conjunction with the analyses

of verbalizations in the Verbal Protocols section, lends credence to the argument that a subject's reasoning on the SCBs played an important role in determining success in this task.

The highly significant ( $t(8)=77.88$ ,  $p<.001$ ) reverse non-solver/solver difference in performance on SCBs in Condition IV, 17.8/27.8, is puzzling. The high  $t$  value can be attributed to the small amount of variability in the group. That non-solvers were correct more often than solvers on these problems can possibly be explained by the fact that in this condition, each SCB is presented twice, with only one problem intervening. Non-solvers tend to rely heavily on references to previous problems and, in this condition, the close proximity of the two presentations of the same SCB configuration made it easy for them to make an accurate prediction on the second occurrence of a problem by simply remembering the outcome of its first occurrence.

First simple conflict: Is performance on it a predictor?

The first simple conflict in I-III was an SCB -- 0200/0300 appeared on problems 5, 5, and 3, respectively. In IV, the first simple conflict was the SCNB 0002/0010, on problem #7; the first SCB, 0003/0010, followed on problem #8.

On the first conflict problem, solvers and non-solvers performed identically in I, II & IV (Table 14). In III, solvers made fewer errors than non-solvers, but both groups were still only performing at or below chance level.

Table B9

Percent (and Number) of Subjects Correct on First Conflict Problem

<u>COND</u>	<u>PROB</u>	<u>#</u>		<u>S</u>	<u>NS</u>
I	SC	5	(0200/0300)	50% (2)	50% (3)
II	SC	5	(0200/0300)	80% (4)	80% (4)
III	SC	3	(0200/0300)	33% (2)	0 (0)
IV	SC	7	(0002/0010)	40% (2)	40% (2)
	SCB	8	(0003/0010)	40% (2)	100% (5)

If one compares performance on just the first SCB in each condition (the second conflict in Condition IV), the perfect performance of non-solvers in Condition IV can most likely be attributed to the small number of subjects and random variability. The parallel difference across subjects between Conditions I and II is also difficult to explain, as the only difference between the two conditions was that in II the beam was visually cleared between problems, a

difference that was found to have no effect on overall performance.

Given these conflicting results, it does not seem prudent to rely on performance on the first conflict, the first SCB, or, for that matter, any one problem as a predictor of success.

These results suggest that even though solvers and non-solvers had similar initial difficulty with simple conflicts. In general, solvers seem to benefit from their mistakes on the simple conflicts that balanced (SCBs) and non-solvers did not. That is, solvers' fewer errors on the SCBs are an indication that, as the session progressed, they were able to predict correctly SCBs they had never seen before, presumably by use of a ratio rule or, at the least, by a growing sense of proportionality.

As will be shown later, non-solvers tended to memorize specific conflict balance problems they encountered while solvers focused on the numerical (multiplicative) relationship between the weights and distances in balancing conflicts. Solvers indicated an early awareness of the inverse proportionality of these problems which not only enabled them to generalize to new SCBs, but led them on SCBs such as 0030/6000 to restate the proportionality as a comparison of products for each side. For example, for



0030/6000, solvers first explained why it balanced with a ratio --  $3:6::1:2$  -- just as they explained 0100/3000 as  $1:3::1:3$ . However, for the former problem, they were also cued by the "doubled" numbers to represent the balance in the form of equal products of weight and distance for each side. It is not at all clear how this second, and crucial, transition takes place, but the analyses of verbal protocols in the next section give some insight into the patterns of reasoning that characterize solvers and non-solvers and lend strong support for delegating a key role to ratio reasoning as a necessary precursor to product-moment rule induction.

### Reasoning, Rules and Transform Strategies

The quantitative analyses of performance (percent wrong) on the different problem types and on success (percent of subjects who solved) in the four sequencing conditions and two manners of presentation provide little information about why some subjects solve and some do not. Task conditions (sequencing and manner of presentation) did not facilitate rule induction and initial performance on the conflict problems was not predictive of success. Coupled with performance on the pretest and with the impression of the experimenter as subjects verbalized, the indication is that, as they began the task, solvers and non-solvers were equally naive about the balance beam and about how to predict the



conflict situations. The only quantitative finding that sheds any light on differences in success was the difference in performance on the simple conflict-balance problems, problems that can be accurately predicted on the basis of ratio reasoning.

Other studies (Siegler, 1976, Hardiman, 1983) have also found that ratio reasoning plays a significant role in inducing the product-moment rule for the beam. But what types of proportional or ratio reasoning are used by subjects and which, if any, lead to the product-moment rule calculations cannot be determined by solely on the basis of the error data. A more qualitative method of looking at the way in which proportional reasoning weaves itself into different subjects' learning experiences is needed in order to see the process whereby proportional reasoning becomes a bridge to the product-moment rule. The verbal protocols of subjects as they make and explain each prediction performs this function.

Also of interest are the other types of reasoning rules and strategies subjects generate for making predictions and the differences in patterns of rule use, proportional and otherwise, between solvers and non-solvers. This information is also well provided by the verbal protocols, making it possible to compile an extensive list of rules and strategies and to pattern their use.

In the first part of this section, a taxonomy of rules and strategies used by subjects in this study is presented. In the second part, the patterns of reasoning rules and strategies of solvers are compared to the patterns of non-solvers, and some conclusions drawn.

### The taxonomy

A major goal of this study was not only to identify the reasoning, rules and strategies observed in the balance beam task, but to classify them in such a way that they would provide a more general analysis of their use by subjects. Each subject's session was coded in an attempt to infer which types of reasoning were being used, and their pattern and frequency of use. It was hoped that by comparing solvers' patterns to those of non-solvers, some insight would be gained into rule-induction reasoning processes.

This taxonomy classifies observed types of reasoning into three basic categories: 1) qualitative reasoning and rules, 2) proportional reasoning and rules, and 3) combining rules. A further distinction is made between rules and transformational strategies.

The rules are a means for evaluating the relative effects of the weights and/or the distances on the state of the beam. Transform strategies, as used here, are operations the subject performs on the given problem to

transform it into an equal (in the sense of retaining the total torque on each side) but now familiar configuration, one for which the subject thinks he has a rule for prediction. In this study there were three predominant strategies observed: cancelling, combining stacks and compensating changes.

### Qualitative Reasoning

Qualitative is used here to mean all reasoning about the beam that does not involve calculation of ratios nor combining of weight and distance values in an algebraic computation of torque for each side of the beam.

A subject was considered to be using qualitative reasoning whenever a prediction about the beam was based on 1) perceptual judgment; 2) ordinal comparison of either the weights or distances, but not a mathematical combining of the two values; 3) ordinal comparison of weight relative to distance; 4) a mixed comparison of ordinal values for one variable (weight or distance) with values that are multiples of the other; 5) comparison of relative numerical differences in weights and distances; and 6) reference to previous problems.

#### 1. Perceptual judgments

Subjects who relied on perceptual judgments for making predictions, they made statements such as "It just looks like it will tip left." and "A weight farther out weighs more than when it's closer to the center." and "Seems like all those weights on the right ought to offset how far out the ones on the left are." Subjects in this mode of thought might count the weights but not the distances. That is, they might explicitly count the number of blocks on each side in order to verify which is greater, especially in the more complex multi-stack problems, but the greater distance is determined visually -- without number values.

At least two subjects in each condition never explicitly counted distance. Instead, they made statements such as "Blocks piled on each other are heavier than if they are spread out" and "The total weights are equal on each side and there's somewhat the same distance...", indicating throughout the run that distance was never encoded numerically.

## 2. Weight or distance only judgments

Subjects explicitly state they are using only one variable -- weight or distance -- for predictions; e.g. "I'm only comparing the weights," or "Maybe only the distances are important."

Nearly every non-solver in each condition used weight alone as a basis for prediction at least once. Some tested it as a rule at the beginning of the run and quickly abandoned it; five non-solvers tested it late in the run (after problem #50), when they were confused and could not think of any other rules to test. A few briefly tested a distance-only rule. No solvers ever gave evidence that they were relying on only one variable to formulate a rule.

### 3. Ordinal comparison of weight relative to distance

"There's more weight on the right, but there's more distance on the left...so it should balance," or "There are two more blocks on the right but I don't think there are enough to compensate for the extra unit away from the fulcrum that the blocks on the other side are." The number of blocks and distance units are counted, but only to determine whether the sides have "more", "less" or the "same" amounts of weights and distances.

This was the most common qualitative reasoning and was used by fifteen non-solvers and three solvers.

### 4. Ordinal(Incremental) differences in distance compared to ratios of weights

Subjects explicitly count both weights and distances, but combine two different types of comparisons into one



rule: Increments of distance are compared with multiples of weight. For example, for 0010/2000, a subject would reason that the beam should balance because there are two times as many weights on the right and the distance on the left is one more unit from the center than that on the right. This gives a correct prediction for a limited set of problems, such as 0020/4000 and 0030/6000, but not for 0020/0010, 0200/0400 or 0030/0060.

It is certainly possible for subjects to devise a rule using "triple" the weight or different increments than one in distance, but, of the seven subjects in this study who used this type of reasoning, the predominant form observed was of the "twice" and "one more" form.

##### 5. Relative numerical differences in weights and distances

Subjects compare the difference in weights to the opposite difference in distances and predict on the basis of whichever is bigger. For example, for 0400/0003,  $W1 - W2 (4-3) = D2 - D1 (4-3)$ ; the differences are the same, therefore, predict balance, which is correct. For 0050/0030,  $W1 - W2 (5-3) > D2 - D1 (3-2)$ ; therefore, predict tip left, which is incorrect.

Four non-solvers and six solvers used this means of reasoning to make predictions. Eight subjects in Conditions I and II used this reasoning (and were wrong) on problem



#13, 0300/0400, after finding that problem #12, 3000/0040, balanced. It is not clear on which problem the ninth subject, in condition IV, developed this reasoning, but the application was to 0200/0400 which he correctly, but with the wrong reasoning, predicted would tip right. A similar application was made by a solver in condition II to predict the imbalance of a problem. Other than these two subjects, all other uses of this reasoning were for the prediction of balance.

#### 6. References to previous problems: valid and invalid

A statement was considered to be a reference to a previous problem if the subject explicitly stated that the current problem had some relationship to a previous problem and its outcome or that they were using a previous problem as a basis for making the current prediction. For example, many subjects made statements such as "This is like the one two problems ago," or "This has one more weight than the one before that balanced..." (Only observed types of references are listed.)

Valid references include:

##### Single transformation

A) same problem

B) critical difference from a known balance state (e.g. 0030/5000 tips left because

0030/6000 balances, 0011/4000 tips right because previous problem, 0011/3000, balances C) critical difference from a specified balance ratio -like 2) except subject explicitly refers to a previous problem as a ratio that balances

#### Multiple transformations

D) when distances are equal, if equal weight is added to each stack on each side, the outcome is the same

E. Sum of two previously learned balanced ratios

Invalid references include:

#### Single transformation

F) problems differing by a single transformation have the same outcome

G) any single change in weight or distance to the lighter side of a tipped beam will make the beam tip to that side

#### Multiple transformations

H) when either distance or weight is unequal, if add or subtract equal weight or distance to or from each stack, same outcome

1) adding or subtracting a weight to a stack on one side compensates moving the stack on the other side one distance unit (e.g. 0020/4000 balances, therefore 0200/5000 balances)

#### Uncodable references

Subject refers to another problem but doesn't say why, or It was not clear which problem he was referring to.

Table B10 shows that the forty subjects made a total of 266 explicit references to previous problems, averaging 6.65 references per subject. On the average, the early solvers made 2.1 references each (8% of all early solvers' responses), the late solvers made 10.8 references each (17.9% of all late solvers' responses), and the non-solvers made 8.35 each (12.8% of all non-solvers' responses). Of all non-solvers' reference-based predictions, 12% were valid, 33% were invalid and 55% were uncodable. For solvers, 20% of their references were valid, 41% were invalid but only 38% were uncodable.

These results indicate that early solvers relied less on references to previous problems as a prediction strategy than late solvers and non-solvers.

Table B10  
References to Previous Problems

Note: See pg.19 for type definitions

NON-SOLVERS (N=20)

TYPE	VALID					INVALID				?* ?
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>	<u>G</u>	<u>H</u>	<u>I</u>	<u>?</u>
COND										
I	0	4	0	1	0	1	9	8	1	48
II	0	0	0	0	0	1	17	5	0	29
III	1	0	0	0	0	0	1	0	0	5
IV	<u>7</u>	<u>5</u>	<u>1</u>	<u>0</u>	<u>1</u>	<u>0</u>	<u>10</u>	<u>2</u>	<u>0</u>	<u>10</u>
T	8	9	1	1	1	2	37	15	1	92

SOLVERS (N=20)

I	3	1	0	0	0	0	14	4	0	22
II	0	1	0	0	0	0	12	4	0	11
III	0	0	0	0	0	0	0	0	0	3
IV	<u>4</u>	<u>11</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>3</u>	<u>2</u>	<u>1</u>	<u>0</u>	<u>2</u>
T	7	13	0	0	0	4	28	9	0	38

	<u>TOTAL VALID</u>	<u>TOTAL INVALID</u>	<u>?</u>	<u>%CORRECT</u>
NON-SOLVERS	20	55	92	55%
SOLVERS	20	41	38	42%

TOTAL REFERENCES = 266

\* ? - Uncoded references

This is in contrast with Hardiman's (1983) suggestion that references to previous problems somehow helped subjects to solve. The finding here that greater reliance on references to previous problems is associated with late and non-solvers suggests that these subjects use this strategy to try to predict accurately each problem, rather than using the references as a source of information from which to induce a rule. That is, non-solvers seem to have a different goal than solvers -- that of immediate correctness for a given problem -- for which reference to specific prior instances is a reasonable strategy, while late solvers seem to vacillate between immediate correctness and rule-finding. Further evidence for this will be shown in the section analyzing subjects' verbal protocols.

The difference in uncodable references further suggests that solvers tended to be more explicit about why they were referring to a particular problem, a skill which may have helped them in general to solve.

Types of references. Table B10 also shows that other than references to identical problems (A), the most common type of valid reference was B, when subjects said that the current problem differed by a one unit change in either weight or distance from a previously learned balanced configuration. The two most common invalid types of references were G, where subjects believe that any change to

the lighter side of a tipped beam will cause that side to tip down or the beam to balance, and H, the multiple transformation in which subjects state that the outcome of two problems should be the same when equal weight or distance is added or subtracted from both sides. This was true of both solvers and non-solvers and is an indication that subjects' most common and deep-seated hypothesis about the beam is that weights and distances have an additive relationship.

In Condition III, there were only two references made to previous problems. Clearly the mixed sequence here provided few memory aids to help subjects remember previous configurations or transformational cues between problems to make the information from previous problems seem helpful.

The larger number of A and B valid references in Condition IV again reflects the repetition in this sequence of the simple conflict-balance problems.

### Proportional Reasoning

Proportional reasoning in this task is in essence reasoning about the problems that balance. For most subjects, explaining why a particular two-stack conflict-balance configuration balances is a far easier task than explaining why an unbalanced one tipped one way instead of another. All subjects seemed to realize that a balanced



beam implies a "numerical equality" of sides. By comparing the ratio of weights to the ratio of distances, a subject can predict whether or not any two-stack configuration balances by determining whether or not the ratios are inversely equal. This numerical equality, in a superficial sense, is not difficult to see in problems like 0010/2000 where the numbers are matched; i.e.  $X @ Y = Y @ X$ . Problems such as 0020/4000 require seeing the proportion as more than a one-for-one matching of numbers, to understand something about proportions or ratios in general. However, while ratio reasoning can easily predict balance or imbalance, predicting the direction of imbalance is more complicated.

What is interesting and varied are the ways subjects build on or extend their first proportional observation. For some it is like a template for compiling a list of proportions that balance -- whenever  $X @ Y$  equals  $Y @ X$ , say balance. For others it is the first step in a series of ratio rules that sometimes leads to the product-moment rule and sometimes does not. For a few subjects, complex ratio reasoning was developed that predicted accurately all problems in the experiment.

1. Specific two-stack balance problems are memorized but the inverse proportionality of the ratios is not explicitly stated.

Many non-solvers began to compile a list of two-stack problems that balanced (SCBs) and indicated they thought "...It has something to do with the ratios..." but few could explicitly state the way in which the weights and distances were proportional. Solvers, on the other hand, tended to generalize about the SCBs they had observed, to look for a rule to explain why they all balanced.

## 2. $X @ Y = Y @ X$ .

Any problem of this general form is predicted to balance

Six solvers and one non-solver explicitly stated this as a rule. Other subjects also appeared to be reasoning from this limited ratio rule but it was often not clear from what people said whether they were merely listing previously learned examples or had abstracted a general rule.

## 3. Twice the...

Over half of the solvers and one of the non-solvers described or explained problems such as 0010/2000, 0030/6000 or 0020/4000 by stating that the weights and/or distances on one side were "twice" or "double" those on the other side. The one non-solver restricted this type of reasoning to problems with multiples of

two, while solvers were more likely to extend it to problems involving other multiples, such as 0200/6000, 1000/4000 and 0100/3000.

#### 4. W1:W2::D2:D1

Subjects predict balance for any problem in which the ratio of weights is inversely proportional to the ratio of distances and state "the ratios are equal."

None of the non-solvers stated this rule; five solvers stated it, but, for the most part, they jumped so quickly from explicit statements of the types 2 and 3, above, to product-moment rule calculations that, while performance on SCBs indicated that they were consistently using this rule, explicit statements of it were rare.

#### 5. Critical difference from conflict-balance

Subjects compare a current problem to a known two-stack conflict-balance to determine the direction of imbalance. For example, given 0500/0003, the subject states that if the weights and distances were proportional, if it were 0400/0003, it would balance, and reasons correctly that one more block on the left will make it tip left. Many subjects used this means to make a prediction. It is difficult to differentiate between when this reasoning was based on general ratio

knowledge and when it relied on reference to specific prior learned SCBs.

### Combining Rules

A combining rule is any rule in which numerical values for weight and distance are arithmetically combined to compute "force" values for each side of the beam. Thirteen subjects (32.5% of all subjects), six non-solvers and seven solvers, explicitly hypothesized and tested one or more addition or subtraction rules of various forms, the most prevalent of which are listed here.

Addition Rules. The predominant combining rules hypothesized by both solvers and non-solvers were some form of addition of weights and distances. All thirteen subjects using a combining rule, other than the product-moment rule, tested at least one form of addition rule. In most cases, subjects assigned a value to the weight(s) and to the distance(s) on one side of the beam, added them and compared the total to the summed weight(s) and distance(s) on the other side. There were two subjects who tested a more complicated form of addition rule which involved adding the absolute value of differences in weight to one side. This is explained more fully below.

1.  $W1 + D1$  vs  $W2 + D2$

### A. 1 distance unit = 1 weight

This was the most common addition rule form and was used by five solvers and two non-solvers. Subjects indicated that they were trying to find the total force on each side by giving a distance unit the same 'worth' as a weight unit. Although subjects did not explicitly state the connection, it is easy to see that this type of reasoning might stem from the earlier, more qualitative awareness that 'a weight farther out is worth more than when it is closer to the center.' That is, for each distance unit out from the center the weight is, it increases, by some factor, in 'worth'. A logical and simple first hypothesis is that the factor is one; for each distance unit out the weight is from the center, increase the total weight on that side by one.

Thus, for each side of the beam, subjects count the weights in a stack and the number of lines the stack is from the center, add the two values, and compare the totals for each side. The side with the greater sum is the side that will tip down. If the sums are equal, the beam will balance. For example, the left side of 0030/0020 would have three weight units and two distance units --  $3+2 = 5$ , and the right would have two weight units and three distance units --  $2+3 = 5$ . Since the two sums are equal, the subject predicts, correctly, that the beam will balance. For a multi-stack problem such as 5000/0023, the left side would

be  $5+4 = 9$ , and the right side would be  $(2+3) + (3+4) = 12$ ; the prediction 'tip right' would also receive positive feedback.

This form of addition rule is understandably confusing to subjects. It gives a correct prediction for any simple conflict-balance situations of the  $X @ Y/Y @ X$  form, for some multi-stack situations such as the one above, and for many simple non-balancing conflicts, such as 0300/0400 and 0005/0002, where the difference in weights equals the difference in distances. In fact, in Conditions I - IV, respectively, the percentage of SCBs that this addition rule will correctly predict were 70%, 70%, 70%, and 75%; for SCNBs, 29.4%, 29.4%, 33.3%, and 62.5%; and for the multi-stacks, 23.3%, 23.3%, 34.3%, and 38.1%. Including the simple weight and distance problems, nearly half (47.2 average across conditions) of all problems can be correctly predicted using this rule. Because there are so many instances where it gives a correct prediction, it is often disconcerting when subjects finally encounter a problem that disconfirms the rule, causing some of them to ask if there are any "tricks" or to decide that it is one of a set of rules for the system, saying, for instance, "My rule works, but only part of the time."



What subjects using this addition rule fail to do is to take fully into account is the number of weights in the stack for which they are counting the distance from the center. If, for example, there are three weights on the fourth mark out, subjects do not realize that each weight has four times the "force" (is 'worth' four times as much) that it had on the first mark, and that the total worth of the stack is the sum total of each weight's 'worth' at a particular number of units from the center -- In this case,  $4 + 4 + 4 = 12$ . This, of course, is the multiplicative function of the product-moment rule -- to account for the distance of each weight in the stack from the center, instead of treating the stack as a whole.

There were other versions of this rule tested by subjects, all based on a distance unit being 'worth' one weight. Three subjects counted only the spaces between the stack of weights and the center (e.g. for 3000/0040, three distance units were added to the three weights on the left, and two distance unit were added to the four weights on the right, so that  $3 + 3 = 2 + 4$ ), or between two stacks on the same side (e.g. for 0202/6000, on the left there is one mark, but two spaces between the two stacks; the subject added  $2+2$  weights plus 2 for the spaces, for a total of 6 for the side.) Subjects were not always consistent in applying the rule. For 1100/2200, one subject computed the

left side as  $1+1$  for the weights plus 2 for the distance from center (2 marks instead of 2 spaces) for a total of 4; however, on the right side, he added  $2+2$  for the weights plus 1 for the space between them, for a total of 5 -- and predicted it would tip right.

B. 1 distance unit = 3 weight units

One subject, after finding that 0030/6000 balanced, hypothesized that each space between the weight and the fulcrum was worth 3 weights -- 3 weights plus 3 (for the one distance mark between) = 6 on the left.

C. 1 weight = 3, 1 distance unit = 1;

1 weight = 4, 1 distance unit = 1

The same subject as in B above also tested combining rules in which the number of weights was multiplied by 3 and the product added to the number of marks between the stack and the fulcrum. When this was disconfirmed, he tried multiplying the weights by 4.

D. Add  $[X(\text{the absolute value of } D_1 - D_2)]$  to the number of weights on one side.

Two subjects used a form of this reasoning to compute torque. In one version, the absolute difference between the distances was added to the side with the smaller total

weight (e.g., for 0200/5000, the difference in distances, two, is added to the two weights on the left, for a total of four and a tip right prediction; In another, the absolute difference was multiplied by two and added to the side with the lesser distance (e.g., for 0300/0500, the difference in distances, one, is multiplied by two and added to the five, so that the computation for the right side was  $[(3-2) \times 2] + 5 = 7$ . This was then compared to the given total weight on the left).

#### Subtraction rules.

While many subjects used subtraction to determine ordinality of distances on opposite sides or as a transform operator (see Transform section below), the explicit use of subtraction in a combining rule to compute the "force" for each side was rare.

In the one case that could be clearly documented, the number of distance marks between the weight and the fulcrum was subtracted from the number of weights on the same side. Thus, for 0300/0500, the "between distance" on the left, two, was subtracted from the three weights on the left ( $3-2=1$ ), and compared to  $5-1=4$  on the right, computed in the same way.

#### Multiplication and division rules.

There were no explicitly stated division or multiplication rules except

the product-moment rule, although a few subjects, when they were looking at a problem such as 0200/6000, mentioned that they "wondered if it had something to do with division."

### Transform Strategies

The term "transform strategies" will be used to refer to transformational rules used to simplify a complex configuration or to rearrange the configuration in a rule-governed manner. That is, weight changes are compensated with distance changes such that the original total "force" or the effect of the "forces" on each side is preserved, while the problem is made to look like a problem for which the subject thinks he knows the outcome. The three observed transform strategies in this study were cancelling, combining stacks and compensating changes.

Cancelling. The cancelling strategy is a means for simplifying a complex problem by cancelling "equalities" and basing predictions on what is left. For example, given the problem 0400/0031, by "cancelling" three weights at the third space on each side, the problem is reduced to 0100/0001 - a simple distance problem; given 5200/0310, by "cancelling" a previously learned equality -- in this case, the equality is the ratio of  $2 @ 3 = 3 @ 2$  -- , the problem

is reduced to 5000/0010, again a simple weight and distance problem.

Eliminating equalities in order to simplify a complex problem is, of course, not unique to the balance beam task. It is an often taught, well-learned, highly helpful strategy for doing many kinds of math, in particular solving of equations. It is not surprising, therefore, that this strategy is applied to the beam by so many subjects, in the sense that whenever the beam is in balance, the point of balance, the fulcrum, is analogous to an equal sign and whatever is done to one side, can and must be done to the other to preserve the state of balance.

Sixty-five percent (13) of non-solvers and 55% (11) of solvers used this strategy. Of the 24 subjects using this means to simplify, 20 used it under six times, 2 used it over ten times, and the two subjects who devised atypical ratio rule solutions to the task used this strategy extensively. For lack of a general rule and in the face of the complexity of the multi-stacks, this is a good strategy for making individual predictions. However, the goal was to find a general rule. The fact that solvers and non-solvers alike used this strategy suggests there might be something about how each regarded the simplified result of cancelling that helped one and not the other to succeed.



### Combining stacks and compensating changes

In order to combine stacks of weights on one side (e.g. to transform 0101/0400 into 0020/0400), subjects had to devise a rule to compensate changes in distance with changes in weight. The rules varied in complexity but were generally of the following form: for each distance unit a block is moved away from the fulcrum, decrease that block's weight by X, and for each unit a block is moved closer to the fulcrum, increase that block's weight by X. Four solvers and four non-solvers used this strategy.

A few solvers and non-solvers had a second use for making compensating changes. They applied rules of weight/distance compensations to both two- and multi-stack problems (e.g. 0200/5000 and 0310/6000) with the goal of having all weights be on the first distance unit on one or both sides of the fulcrum. For example, to transform 0200/, a subject might decide that for each unit the two blocks are moved toward the center, the weight should be doubled--  
 $0200/ = 0040/ = 0008/$ ; or, as in the add rules, each distance unit is worth one weight and  $0200/ = 0030/ = 0004/$ . In the case of the multi-stacks, 0310/, for example, with a 1 block = 1 distance unit rule, would be transformed:  $0310/ = 0302/ = 0007/$ .



### Patterns of Use: The Verbal Protocols

A major goal of this research was to characterize the differences in reasoning of solvers and non-solvers. The verbal protocols of subjects making and explaining their predictions for the balance beam problems were the source for documenting the diversity of reasoning rules listed in the previous section. They will also be the source for illustrating the patterns of rule use for subjects as they progressed through the different sequences of problems.

The protocols were coded as indicated below. Table B11 shows the coded protocols of non-solvers and Table B12 shows those of solvers. These coded protocols will be used to highlight commonalities and differences within and between the two groups.

Note: A subject was considered a solver when he or she stated the product-moment rule and correctly applied it to three successive multi-stack problems.

This section is organized into the following parts: 1) protocol coding, 2) sequencing considerations, 3) non-solvers, 4) solvers, and 5) atypical solvers.

Table B11

## Non-solvers: Coded Verbal Protocols

QUALITATIVE NON-SOLVERS

PROB.# 5 - - - - - 20 - - - - - 35 - - - - - 50 - - - - - 65

I, II  
 17 (-----Q-----) multi-stack → rf (---Q---) rf (-----Q-----)  
 43 rf (rf) Q ( rf ) Q (rf) Q (rf) (rf) (rf) (rf) (-----Q-----) rf  
 11 (-----Q-----)rf (-----Q-----)  
 95 (-----Q-----) (rf) rf (-----Q-----)

IV  
 160 D rf rf multi-stack → rf rf C r C (-----Q-----) C (---Q---)  
 175 (rf) (rf) (rf) CS C C CS rf rf rf Q C rf (---Q---)  
 178 (rf) r (rf) (rf) (rf) (C) (-----Q-----) C

III  
 36 Q C Q C (---Q---) D (-----Q-----) C (-----Q-----)  
 63 Q C Q C (---Q---) rf (---Q---) C C/CS rf CS C rf  
 87 r (-----Q-----) C r (-----Q-----)

RATIO NON-SOLVERS

I  
 41 D r rf r + C r rf r (rf) multi-stack → (r) (rf) (---Q---) CS rf (-Q-) rf r  
IV  
 22 D r rf r rf (- r -) multi-stack → r C CS rf (-----Q-----)  
III  
 45 Q D rf Q (r) rf r C C (-----Q-----) (r) C c/r Q (C/r) C r (C)

RULE FINDING NON-SOLVERS

I, II  
 29 D + Q +! rf t rf t +! +! +! multi-stack → (++)! ++! (C) rf CS (++)! rf (++)!  
 101 D(+) x/+ rf - +! +! +! ++ rf (-----Q-----)+! -- (++)  
 143 D (rf) - -- rf (--)! X rf X rf (-----Q-----)rf Q rf + +! rf  
 162 rf(\*) \*! D x/- (x/-)! rf (x/-)! (C) + C rf (C) rf (+)! C rf +! rf +!  
 48 D + rf rf (+)! (+)! rf Q +! rf +! +! rf rf C (rf) rf rf ++ rf  
 130 D rf \* (\*)! Q \*! rf (+) \*! (+)! rf C rf rf (+)! rf rf  
IV  
 18 D (rf) (rf) r multi-stack → (rf) C/r CS + C/r rf +! CS C +! +! C +! C (+)! CS

Table B12

## Solvers: Coded Verbal Protocols

EARLY SOLVERS

PROB.# 5 - - - - - 20 - - - - - 35 - - - - - 50 - - - - - 65

I, II  
4 D rf (\*)! PM ----- multi-stack →  
26 D rf (\*)! rf T PM ----- PM  
31 D (+)! (T) rf (R) PM ----- PM  
39 D T PM ----- PM  
42 D + rf (\*)! (T) PM ----- PM  
152 rf D \*! + (T) rf PM ----- PM

IV  
13 D (rf) T \* (R) multi-stack →  
28 D r rf PM ----- PM  
50 D (t) (T) C C CS PM

III  
25 D C r R PM  
64 D r r + X PM  
134 D r T PM  
158 D r X (R) PM

KEY  
D - distance encoded  
rf - ref. to prev. prob.  
! - retest of disconfirmed rule  
(-) - repeat use  
Q - qual. guess, perceptual, ordinal reasoning  
r - specific ratios recalled  
R - ratio rule  
\* -  $W_L - W_R = D_R - D_L$   
t - twice w/ordinal D  
T - twice w/twice D  
X - "multiples" "times"  
\*, -, X combining rules  
C - cancel strategy  
CS - combine stacks strategy  
PM - P-M rule stated

LATE SOLVERS

I, II  
30 D r (rf) r (rf) +/- (rf) multi-stack →  
98 (rf) D (rf) (+) (rf) r (rf) (rf) (+/-)! + rf - (+) r PM  
91 rf D \* r (rf) T (R) ++ (rf) (rf) + PM  
(C/R) PM

IV  
37 D (rf) r (rf) r multi-stack →  
CS C CS rf C r (C) r (C) PM

III  
100 D r X r rf r r r r + r PM

ATYPICAL SOLVERS

III  
27 rf D (r) (-----R-----) (R/move transforms-----)

IV  
82 (rf) D (rf) (rf) multi-stack →  
C CS (C/r) (C/r) (C/R) rf (move trans/r---

Table B12 cont.

SOLVER COMMENTS DURING BALANCE TRAINING (UNPROMPTED)  
OR AFTER TRAINING (PROMPT: "HOW DID YOU DISCOVER THE RULE?" )

	<u>SUBJ. #</u>	<u>COMMENT</u>
EARLY	4	(after) "Hunch."
	26	(after) "I just multiplied on (0040/0003."
	31	no reason
	39	(prob. #32) "I'll try multiplying."
	42	(after) "I began combining numbers."
	152	(prob. #33) "All I can figure out is to multiply weight times distance and compare sides."
	13	(prob. #26) "Maybe if you multiply..."
	28	no reason
	50	(prob. #32) "You times the number of blocks and the number of distance units."
	25	(prob. #26) "I just started being arithmetical."
	64	(prob. #26) When '+' rule disconfirmed: "Oh!... multiply."
	34	(prob. #11) "Maybe....multiply"
	158	no reason
LATE	30	(after) "I was stuck in a rut with '+' and '-'. Don't know why it took so long to multiply."
	98	(after) "It (the PM rule) just came into my head."
	91	Had complex working ratio rule; PM on prompt from E.
	37	(after) "I was stuck trying to get them right."
	100	(prob. #54) E prompt: "Rule?" (after) "When you asked if I had a rule, I began to look for a math rule."

### Coding of the verbal protocols in Tables B11 and B12

The verbal protocols were taken from audio tape recordings made of each subject's experimental session. The tapes were coded for:

- a) numerical encoding of distance
- b) combining rules
- c) explicit use of specific ratios
- d) explicit use of a general ratio rule
- e) accurate multiplicative statements ("twice the weight vs. twice the distance" or "it balances because  $2 \times 3$  equals 6")
- f) inaccurate multiplicative statements ("twice the weight vs. one more distance unit")
- g) transform strategies
- h) references to previous problems
- i) retests of disconfirmed rules.

Also shown in Table B12 are Solvers' comments, either during or after their run, about why they finally multiplied weight and distance.

A code in parentheses indicates repeated use of a rule or strategy at that point in the session.

Sequencing considerations. Because the sequences in Conditions I and II are identical, these will be considered in the Tables as a single group for this analysis. Conditions III and IV will be considered separately in order to determine if and how sequencing effects patterns of reasoning.

## Non-Solvers

Table B11 shows the coded patterns of the twenty non-solvers. Non-solvers divided into three groups: 1) Qualitatives: those who predicted predominantly on the basis of qualitative judgments (ordinal comparisons; see Rules section, pg.13), 2) Ratios: those who relied on qualitative rules but also recognized specific ratios and used transform strategies, and 3) Rule-finders: those who hypothesized numerical combining rules, but also relied heavily on instance-based reasoning and transform strategies.

Qualitative Non-solvers. There were ten subjects who reasoned mostly on the basis of qualitative rules. They never explicitly counted distance, but predicted by guessing, making ordinal (more, less, same) comparisons of weights and distances (e.g. "There's more weight on the left, but it's (the weight on the right) farther out on the right," or by perceptual assessment (e.g. "It just looks like the weights would offset the distance). Some subjects in this group also relied on references to previous problems or attempted to transform the problems into more familiar configurations.

As a group, they were plainly bewildered by the task and their statements during the session indicate that they had



no idea how to proceed nor that the rule they were instructed to find involved both weight and distance numbers. One subject, for example, said, one-third of the way through the session:

...no definite rule. It seems like if they're closer to the center on one side, on the other side they can be farther out from the center and still be able to balance it -- depending on how many there are. If there's a lot close to the middle, maybe a little bit further out on the other side might balance it off.

Ratio Non-solvers. There were three non-solvers who seemed to be particularly focused on the SCBs. They memorized specific two-stack ratios and used transform strategies to change multi-stacks into known SCBs. All three counted distance early in the session, but none attempted to combine weight and distance numbers into an algebraic rule.

#### Rule Finding Non-solvers.

Rules. Seven non-solvers, all in Conditions I, II, and IV, hypothesized and tested one or more numerical combining rules for prediction. All encoded distance numerically early in the session. Four tested two or more rules in which they either tried different ways to apply a single math operation (e.g. two ways to add weights and distances) or they varied the operator (addition, subtraction,

multiplication or a combination of two). Three tested only an addition rule.

References. Despite their rule-finding orientation, these subjects also relied heavily on references to previous problems and, in the multi-stacks, transform strategies.

Retests and ratios. Of particular interest is the repeated retesting of a disconfirmed rule, and the total absence of ratio reasoning for either specific ratios or in the form of a rule.

A set of rules? Not shown in the table are the comments of at least four of the eight combiners in which they indicated their belief that there was a set of rules for the balance beam, saying, for example, "My rule works, but only some of the time." This belief would explain the repeated retests of disconfirmed rules. Subjects who believed there were different rules for different types or classes of problems were possibly retesting rules in order to determine just which types of problems they could be applied to.

In sum, subjects who were rule-finders but did not solve appear to have been unsuccessful because a) they vacillated between a rule-finding approach and instance-based strategies, b) they had poor hypothesize-and-test skills, c) they did not discover the ratio rule or see the SCBs in

multiplicative terms, and/or d) they believed that there were multiple rules for the system.

### Solvers

Eighteen of the forty subjects were solvers. The solvers, however, were not a uniform group. As can be seen in Table B12, they divided clearly into two groups: early solvers and late solvers. There were thirteen early and five late solvers. In the sequenced conditions (I, II, and IV), all of the early solvers stated and correctly applied the most general form of the product-moment rule before the sixth multi-stack problem. In the mixed condition (III), they solved by problem thirty-three.

Late solvers in I, II, and IV solved after problem fifty, and, in III, the one late solver solved on the sixty-third problem.

Early Solvers. As can be seen in Table B12, early solvers' patterns were quite similar. They all encoded distance numerically early in the session, made few references to previous problems, seldom used transform strategies, and did not test many combining rules. Seven out of the nine early solvers in the sequenced conditions stated and used the product-moment rule before the end of the two-stack portion of the sequence, and then proceeded to

apply it accurately and without hesitation to the multi-stacks when they appeared.

The most interesting findings were related to 1) ratio reasoning, 2) rule-testing skill, and 3) subjects comments.

Ratio reasoning. In Tables B11 and B12, a "T", "X", or "R" indicates that on a SCB problem, a subject stated either that there was "twice (or double or half) the weight on one side and twice the distance on the other" (T), or that e.g. "two times three equals six" (X), or some form of correct ratio rule, e.g. "the ratio of weights is proportional to the ratio of distances" (R).

Eleven early solvers used multiplicative terms (T or X) or the ratio rule (R) to explain SCBs such as 0020/4000 and 0030/6000, prior to stating the product-moment rule. The other two solved so early in the session that it was impossible to determine their prior reasoning.

Within the sequenced conditions, the SCBs and SCNBs were mixed with simple weight and distance problems in the beginning two-stack series. When a SCB appeared in the sequence, early solvers refined or used ratio reasoning and as a group averaged seventy-five percent correct of those predicted prior to stating the product-moment rule. However, application was limited to SCBs. When a SCNB appeared,

these solvers were initially just as incorrect and perplexed as non-solvers.

Rule testing. In the table, a "!" indicates that a subject has retested a rule that has previously been disconfirmed. Solvers as a whole seldom retested a rule for which they had received negative feedback. In no case did they retest more than once.

Subject comments. Subjects were asked at the end of the session why they thought the rule they had induced worked. No subject could give an explanation other than that "the numbers worked" or similar calculation descriptions. The comments in Table B12 show that when subjects first thought to multiply, they were unaware of where the idea had come from. For most subjects, it appears to be a sudden insight, not tied to their immediately preceding train of thought. What the table cannot show, but is highly evident in the protocols, is the transition from confusion and bafflement to the "of course!" reaction subjects have once they have thought of multiplying. Even before they have tested it, they "know", with no uncertainty, that they have found the rule, and, with equal confidence, how to apply it to multi-stack problems. They are also greatly surprised that they



did not think of it before. Now that they have thought of it, it is patently obvious.

Sequenced vs. mixed differences: early solvers. The most salient difference between the four early solvers in the mixed condition and those in the sequenced ones is the nearly total absence of hypothesized combining rules in the mixed condition.

Otherwise, they looked just like the other early solvers: they encoded distance early; they were, on the average, correct on the SCBs almost as often as the solvers in the other conditions (70% correct vs. 76%); and they each gave multiplicative or ratio explanations of the SCBs. Interestingly, despite the complexity of the mixed sequence, the earliest solver was in this condition.

Clearly, there were subjects for whom the mixed sequence was no more difficult than the sequenced ones were for their solvers.

In sum, early solvers as a group were subjects who used multiplication or ratios to explain the SCBs and from there jumped to correct product-moment calculations for all problems. The fact that they made few references to previous problems and hypothesized only combining rules can be taken as evidence that these subjects consistently focused on looking for one general rule, not multiple rules for



multiple categories of problems. These subjects were also more efficient generators and testers of possible combining rules than non-solvers quickly abandoning them when disconfirmed.

While it seems logical that a subject should prefer to try to explain the equal "forces" on each side of a SCB configuration rather than deal with the more complex inequality of the SCNBs, it is not at all clear how a subject, having explained the SCBs in multiplicative or ratio terms, made the generalizing leap to prediction of all problems via the product-moment rule. One possibility is that simply having multiplication in current memory makes it more available to be one of the hypothesized rules to test on the more complex problems. However, this does not explain the leap to the product-moment rule calculations that occurred for some subjects solely from ratio reasoning.

Late Solvers. The five late solvers look at first glance very much like non-solvers. In addition to hypothesizing combining rules of limited generality, they relied extensively on non-analytic, instance-based reasoning, i.e. they made numerous references to previous problems in order to make current predictions, memorized specific ratios without generating a ratio rule, and used transform strategies.

However, those that hypothesized incorrect combining rules, did not often retest after disconfirmation, and, looked at one by one, each late solver's eventual success seems to have an explanation. Subject #30 was "stuck in a rut adding and subtracting," and then suddenly thought to try multiplying; #91 explained a SCB using "twice the.." terms and also stated the ratio rule; and #37 and #100 solved shortly after realizing, late in the session, that they should be looking for a math rule. The key to #98's eventual success was indeterminate.

One important factor was that all of the late solvers at some point explicitly stated that they were looking for a general combining rule, even though they were using non-rule oriented strategies to make predictions. They made statements such as "I'm trying to find a way to relate the weights and distances," "I'm trying to find a number thing," or "Going to try and form a mathematical equation." They seemed to understand the task, as evidenced by the numerical combining rules they hypothesized and tested, but to fall back on instance-based, get-this-one-right strategies whenever their hypothesized rules were disconfirmed.

Sequenced vs. mixed differences: Late solvers. There was only one late solver in the mixed condition, one in Condition IV, and three in I and II combined -- too few to

draw any substantial conclusions. It does appear, however, that in I, II and IV, subjects tried more different approaches than subjects in III.

### Atypical Solvers

There were two subjects who "solved" in the sense that they devised a ratio-based means for making consistently correct predictions, but never induced the product-moment rule. One subject, #27, applied the ratio rule (i.e. he compared the ratio of weight and distance on one side to the inverse ratio on the other) on all conflicts, single and multi-stack, using appropriate transform strategies first to simplify the configuration. The other subject, #82, developed a complicated system in which he moved all of the weights on one side to the first distance line, systematically increasing their weight for each unit moved in, and then comparing the simplified configuration to a specific, previously learned SCB.

In both cases, although their verbalizations of how they were applying ratio reasoning to a multi-stack problem were difficult to follow and much of what they were doing was not explicated, they proceeded systematically and confidently with their individual methods and were able to predict all problems. However, it seems doubtful that they would be able to predict as well if the problems were greatly

increased in complexity, in which case the product-moment rule would be far simpler to apply.

## CHAPTER VII

### GENERAL DISCUSSION

#### Task Cues vs. General Problem Solving Skills

The basic purpose of this research was to investigate the types of reasoning people use in trying to induce the rule for the balance beam and to determine key factors of success.

The major findings of this research were 1) adult subjects reason from a very diverse set of qualitative and quantitative rules, 2) sequencing and manner of presentation did not seem to have an effect on success, 3) ratio reasoning on simple conflict-balance problems appears to play a key role in cuing subjects to the multiplicative nature of the product-moment rule, and 4) that solvers and non-solvers show marked differences in general problem solving skills.

The proposed models of learners in this task assumed that the adult participants, given that they were college students, had the basic math knowledge necessary to understand the product-moment rule calculations; i.e. it was assumed that they 1) knew how to give numerical values to variables (weight and distance), 2) had had some experience with combining variable values in an arithmetic rule (e.g.

length  $\times$  width = area), and 3) knew the basic math operations for combining (add, subtract, multiply and divide). The underlying question for this research is whether the elements and structure of the balance beam task itself are the key factors in triggering the application of this assumed existing knowledge, or are there external factors underlying the differences between solvers and non-solvers?

While the findings from a study such as the current one can provide only a description of behavior and verbalized reasoning, the description that has resulted strongly supports the view that differences in success in this task reflect differences in previously learned general problem solving skills.

One way to characterize these differences is in terms of the extent to which subjects took an analytic approach to the task. Solvers were those subjects who took a predominantly analytic, rule-finding approach to the task; non-solvers were those who tended to rely more on non-analytic, instance-based reasoning and strategies for making predictions. Solvers stayed focused on the task of abstracting one general rule for prediction, while non-solvers were more intent on getting-this-one-right and tended to reason by analogy to specific, previously learned



problems, or to vacillate inefficiently between analytic and non-analytic approaches..

Brooks (1978) has made this distinction between rule-based and instance-based reasoning in relation to concept formation. In this framework, the choice between a rule-based analytic approach and an instance-based non-analytic approach depends on a person's perception of the complexity of the task materials. In the face of too much complexity, people default, often prudently, to a non-analytic approach as a means to reduce the complexity. That is, they develop multiple bases of categorization, learn the individual items that go into each category, and then reason about new exemplars by analogy to category instances. Given the complexity of the balance beam task, this seems a reasonable framework, on two levels, for interpreting the results of this study.

On the first level, the analytic-non-analytic distinction applies to the way in which subjects categorize problems, first by type (simple weight and distance and conflicts) and then by outcome (simple conflict-balance and simple conflict non-balance).

On the second level, given the instructions "Find a rule that will allow you to predict accurately one hundred percent of the time," a subject could take an analytic, one-rule view of the task or he could categorize it as a non-

analytic task with the emphasis on correctness via multiple bases for prediction.

All of the subjects in this study responded initially to the task on the first level in a non-analytic fashion. It was clear from the data that they all had appropriate rules for correctly predicting the simple weight and distance problems, and it was equally clear from their verbalizations that they recognized these problems to be in a separate class than the "tricky" ones, the conflicts, for which the simple weight and distance rules were inadequate. When a conflict problem appeared, most subjects made statements such as "Oh, this is a tricky one" or "Now there's more weight on the left but there's more distance on the right" or "Now we get to the hard ones."

All subjects indicated that when they were making predictions or, after making an error, were responding to the feedback answer in the display, they were further differentiating between conflict problems that balanced (SCBs) and those that did not (SCNBs). They repeatedly made statements about the SCBs that ranged in specificity from "It's proportional," "It just looks like it will balance because the bigger weight on the left compensates the greater distance on the right," and "This one won't balance because even though there are more weights on the left, they aren't far enough out to offset the ones on the right," to

"This will balance because of some ratio thing," to "This will balance just like the other ones because the weights on each side equal the distance on the other."

One of the most salient indicators of categorization of problems by type was the frequency with which subjects, and in particular non-solvers, explicitly listed the SCBs they had previously encountered. One subject, for example, said "I know that three weights at the second mark equals two weights at the third mark (0030/0020), one weight at four equals four at the first (1000/4000), one at the third equals three at the first (0100/3000), and one at the second equals two at the first (0010/2000); so three weights at the fourth mark should balance four at the third (3000/0040)."

This did not necessarily indicate that a subject had a ratio rule that would generalize to problems such as 0006/0020, but only that the subject had formed a category of SCB problems they had previously learned.

It is at this point that the differences between solvers and non-solvers become apparent. As the protocols show, some subjects referred to specific instances in the list to make current predictions ("0100/3000 balanced, so 0100/4000 should tip right'), while others tried to find a rule to explain the side-to-side equality of SCBs in general. Subjects who applied a rule-finding approach to SCBs and

recognized the multiplicative relationship of the weights and distances in the SCBs were more likely to solve, and to do so early. Late solvers either hypothesized combining rules, but failed to include multiplication as a possibility until late in the session, or relied on non-analytical reasoning until, again late in the session, it suddenly occurred to them to "become mathematical."

In this sense, the late solvers, up until just before they solved, look very much like the rule-finding non-solvers, until one compares the rule-testing patterns of the two groups. The late solvers, once they realized they were looking for a general, numerical rule, were far more efficient rule testers than were the rule-finding non-solvers and they seemed better able to keep in mind the goal of finding one general rule. Rule-finding non-solvers, on the other hand, could be characterized as adopting a rule-with-exceptions strategy wherein the goal is to find a "main" rule to handle most situations and a set of auxiliary rules and strategies to handle the exceptions (Holland, 1985). Since there is little about this task that could be considered training in hypothesize-and-test skills, (i.e. subjects did not receive any explicit feedback related to retesting of disconfirmed rules, nor were they given any overt clues as to what kinds of rules to test (numerical vs. qualitative), it seems reasonable to conclude that the

difference between the late solvers (and also the early solvers), and the rule-finding non-solvers is attributable, in large measure, to differences in previously learned problem solving skills.

#### Understanding the system vs. understanding the numbers

There remains the underlying question of determining on what level a subject must 'understand' the balance beam in order to induce the product-moment rule. What part does understanding the physical properties of the beam play in inducing a numerical rule to predict its action? What does a subject who knows the product-moment rule understand/know about the balance beam? The findings of this study argue for the view that solvers do not necessarily have a better understanding of the balance beam as a physical system than non-solvers, despite their knowledge of the rule and their ability to predict the behavior of the system under any conditions. But, rather, that solvers attended to the numbers in such a way as to discover how they could be combined.

Subjects begin the rule induction task by trying to understand and define the actual physical effects of the weights on the beam in terms of what they know about seesaws and weighing scales in the real world. On the basis of the examples in the task and memory of related real world



experiences, most subjects were able to verbalize a superficial level of understanding of the beam in qualitative terms. For example, "The weights have more force if they're farther out." For some subjects (ten non-solvers), this was their only level of reasoning, besides references to previous problems, and led to predictions based solely on ordinal comparisons; e.g. they knew that if the weights were equal, the side with the weight farther out would tip down (simple distance rule). But, when the weights were not equal, they reasoned that if the difference in weights was greater than the difference in distances, the side with the greater distance would go down. Their superficial, "more, less same" ordinal understanding of seesaws and balance beams was helpful only for simple weight and distance problems.

However, most subjects (thirty out of forty) progressed to more quantitative modes of reasoning in which they began to combine numerical values of weight and distance into rules for prediction. The unprompted comments of many subjects that they were looking for a "number thing," "a pattern," "a mathematical equation," and of solvers who suddenly said, for example, "I'll try multiplying..." and "Maybe you times the number of blocks and the number of distances," strongly suggests that at some point, they were no longer concerned with the beam as a physical system.



Rather, they were only concerned with how to make the numbers come out right. They had perhaps developed some hazy physical sense of why other combining rules did not work all of the time, but they were totally unable to explain why multiplication did fit the observed outcomes of all of the learning examples. Their comments, when queried at the end of the session, such as "It was just a hunch," and "I just started being arithmetical," further support the argument that induction of the product-moment rule does not imply that the inducer has a deep understanding of the system this rule represents. Rather, it indicates that the inducer has successfully figured out how to manipulate the numerical values of the variables.

This has important implications for teaching. It is yet another instance where a student may have learned a rule for a system without ever having a deep understanding of why the rule works. It is particularly easy to erroneously assume a deeper level of learning in a situation such as this experiment in which the student discovers or induces the rule himself from experiences with examples, without clues or instruction.

Fit to the Proposed Models

Earlier, two models of reasoning in the balance beam task were proposed: the "analytic model" and the "non-analytic model." The "analytic model" was used to predict successful and unsuccessful patterns of performance for subjects who maintained a predominantly rule-finding approach to the task, i.e. those subjects who focused on finding one general rule for making predictions. It predicted that successful analytic subjects would encode distance early, treat proportional problems analytically, make explicit statements indicating they were looking for a single mathematical rule, seldom retest a disconfirmed rule, and, after disconfirmation, quickly generate and test a new combining rule. It predicted that a subject using an analytic approach would be unsuccessful if he failed to encode distance properly, did not understand the SCBs in multiplicative terms, or persisted in retesting disconfirmed rules. The "non-analytic model" was used to make predictions about non-solving subjects who focused on making accurate predictions about individual problems from references to previous problems or to classes of problems. It predicted that these subjects would develop multiple rules for predicting different categories of problems, would

make more references to previous problems than solvers, and more retests of disconfirmed rules.

The findings of this study are that these proposed models present an oversimplified picture of the reasoning processes leading both to success and to failure in the balance beam rule induction task. As outlined below, the proposed model of the successful analytic subject fits the findings for the early solvers in this study, but fails to take into account the far less direct route of the late solver's reasoning. Nor does it adequately highlight the key role multiplicative explanations of SCBs play in triggering the test of a multiplication operation in a combining rule.

Early solvers. Early solvers fit the model in that they 1) encoded distance numerically early in the sequences, 2) explained the proportionality in the SCBs in an analytic manner -- either with a ratio rule or with an operational description such as "twice the weight vs. twice the distance," and 3) seldom retested disconfirmed rules.

The proposed model does not account for the suddenness and confidence with which early solvers "just thought to multiply" in a combining rule, and consequently how few other combining rules they tested.

Late solvers. Late solvers, like early solvers, also fit the model's predictions for 1) early distance encoding and 2) few retests of disconfirmed rules. Unlike the early solvers, however, they hypothesized many more combining rules before discovering the product-moment rule, vacillated between analytic and non-analytic reasoning and strategies, but made many explicit statements that they were actively looking for a "rule," a "pattern," or "some math thing."

Also unlike the early solvers, the product-moment calculations were triggered by multiplicative explanations of the SCBs for only one subject. For the other late solvers, success seemed to just "happen" out of their confusion and frustration.

Rule-finding non-solvers. A model of the rule-finding non-solver would look very much like the model for the successful late solver, with two important differences. One, it would have to account for the frequent retests of disconfirmed rules; and two, it would have to account for the fact that, although these subjects knew they were looking for a combining rule and tested many rules while in the two-stack part of the sequence, they were easily discouraged when confronted with the later multi-stack problems and were, thus, prone to fall back on non-analytic strategies.

Qualitative and Ratio non-solvers.

It would be difficult to model the qualitative non-solvers because, except for their references to previous problems, their reasoning for each prediction is hazy even to themselves; and the "Ratio" non-solver because their one category of problems, besides simple weight and distance ones, is of memorized configurations that balance. It is difficult to even speculate which element(s) of the models of success would lead these subjects to the product-moment rule.

## CHAPTER VIII

### CONCLUSIONS AND FUTURE DIRECTIONS

This study has lent support to the findings of previous studies that subjects attempting to induce the rule for the balance beam use many rules of limited generality and often rely on instance-based reasoning to make predictions. The current findings also documented patterns of reasoning for two descriptive models of solvers and three of non-solvers. Taken together, the five models might be seen as a continuum of reasoning progressing towards induction of the product-moment rule.

Taking the five models to represent successively more sophisticated stages of reasoning provides a framework for the many questions that were left unanswered in this study. The first question to arise is whether a solver progresses through all five stages en route to successful rule induction, or do some solvers bypass the more primitive modes of reasoning? Because the early solvers in this study moved so quickly to the product-moment rule, it was difficult to determine all of the steps that they took. However, the protocols suggest that early solvers' patterns are compacted versions of the late solvers and rule-finding non-solvers' patterns, but guided by more efficient general problem solving skills. In all three groups, problems were



categorized by type (simple weight and distance, simple conflict-balance, and simple conflict non-balance) and a different approach used to predict each type: the simple weight and distance problems by ordinal comparisons, the simple conflict-balance problems by reference to previous ratios or use of a ratio rule, and the non-balancing conflicts by a variety of combining rules. The primary differences between subjects in these three groups seem to lie in their varying abilities to remain rule-oriented in the face of complexity (to not resort to references to previous problems, particularly for the simple conflict-balance configurations), and their varying degrees of efficiency in testing possible rules. Research procedures that would more precisely unpack the early solvers' reasoning is necessary to verify this conclusion.

The second question is how to facilitate a person's refinement of more primitive reasoning into a more efficient and goal-focused approach to induction. What types of pre-task training would provide the appropriate problem solving skills to be applied and what task related manipulations would trigger subjects to apply existing or newly learned problem solving skills?

A future study might attempt, in an analogous task or by outright instruction, to pre-train subjects to be aware of heuristic pitfalls, such as were discussed above, for tasks

such as the one in this study. Pre-training might include experience with algebraic rules (e.g. simple area of a square computations or another rule induction task) or a task which emphasizes learning from errors (i.e. previous problems), as opposed to concentrating on reducing errors.

When asking what task-related variables might be manipulated to trigger application of the appropriate skills and reasoning, many aspects of the balance beam task become likely candidates for investigation.

Sequencing. It is possible that, given the finding that a multiplicative explanation ("twice the...", "double the...") of the simple conflict-balance problems is correlated with eventual discovery of the product-moment rule, a sequence that centered more around problems of this type would have an effect on success. If SCBs for which the addition rule gives a correct prediction were contrasted, in close succession, with SCBs for which it does not give a correct prediction, subjects should be cued to an earlier abandonment of the add rule. And if there were also more SCBs of the "multiple" type -- 0020/4000, 2000/0400, 0060/0003 --, subjects, looking for a new rule to test, should be cued by these problems to test a multiplication operation. However, "success" in this case would seem to be more the ability to understand the numbers than understand the physical system of the beam itself.

Hints. Would subjects perform better if they were given specific hints that a) numerical distance is important, b) finding one rule was the main goal, not making individual correct predictions, or c) there is one type of problem (SCBs) that will be most helpful in understanding the system?

Slegler and Klahr (1982) found that focusing attention on the numerical values of both weight and distance improved the ability of older subjects to induce the product-moment rule. This provides support for the use of this type of hint to move a subject out of a purely qualitative mode of reasoning into a combining mode, but it does not seem likely that it would help the rule-finding non-solvers, who are already encoding distance numerically, to become more efficient in their rule-testing procedures.

For these subjects, a hint to "stay on track" (i.e. to stay in a rule-finding mode) should help them to self-monitor and avoid the tendency to fall back on "get-this-one-right" strategies whenever the task seems too complex or frustrating. For the same reasons, such a hint should help the late solvers to solve earlier.

A hint to regard SCBs as key problems should have the same effect as highlighting them in the sequencing.

Memory aids. Would subjects perform better if they were given external memory aids such as a list of the

configurations or were allowed to make notes of the ones they thought important? Siegler and Klahr (1982) found that giving subjects printed schematic representations of the balance beam problems they were encountering, along with their outcomes, improved ability to induce the rule for the beam. Perhaps such aids helped subjects to make their own order out of the confusion they felt about the problems, to reduce the memory load of grouping similar types of problems together so that they could be analyzed. It would also seem likely that in the process of noting particular problems, people would devise a notation system to represent them which would make numerical encoding and relationships more salient.

Allowing subjects to make notes as they went along would provide still another way to unpack the way in which they categorize problems. It would also be interesting to discover the different coding systems that might be devised.

Diagnostic problems. A self-initiated means for setting up configurations on either a concrete beam or on the computer, might be helpful to those subjects who want particular types of diagnostic problems on which to test their current hypothesized rule. If, as a subject was moving through the presented sequence, he could branch off at any time to configurations he thought would confirm or disconfirm a rule, instead of having to wait for certain

types of problems to appear in the sequence, would he be less confused and more efficient? Intuitively, it seems that the subject who is asking for specific diagnostic problems would be actively hypothesizing to the same extent as the subject who is explaining his reasons for his predictions, although this does not assume that the same type of reasoning would be used for both types of active hypothesizing.

Concrete aids. Would subjects perform better if, in addition to the computer representation, they also had a concrete beam on which the problems were represented? Over a third of the subjects in this study mentioned a seesaw or said they wished they had a real beam to see the problem on. Perhaps, for some subjects, the computer representation is too abstract. A future study might compare computer vs. concrete vs. a combination of both.

Active hypothesizing. One final question not empirically addressed in the current study is that of the role of active hypothesizing. It was assumed that actively hypothesizing via verbalized explanations for each problem would have a positive effect on performance. This needs to be verified, for it is also possible that, left to their own devices, possibly with a more flexible means for generating their own problems, subjects would be more successful than



under the constraints and verbal demands present in this experimental design.

The overall conclusion drawn from this study is that heretofore too simplistic an approach has been taken to investigating induction in the balance beam task and to induction in general. The variability in reasoning rules, strategies and inductive problem solving patterns for different subjects that was found in this study indicates that future investigations be designed and interpreted in such a way as to allow the full spectrum of the learning continuum to be revealed.



## SECTION II

### TRANSFER

## CHAPTER I

### INTRODUCTION

The second focus of this research was on transfer of knowledge from one domain to another. The particular concern was with the transfer of balancing knowledge to understanding of the statistical concepts of the mean and the weighted mean. Many students, including college-age students, have only a superficial understanding of the mean and have considerable difficulty with weighted mean problems. Students learn to calculate the mean and have some notion of what an average is, but they do not fully understand why these calculations work or what the end product represents. Without a complete understanding of the mean, students are even more at a loss to understand the weighted mean (Hardiman, 1983).

It has been suggested that a full and integrated understanding of the concept of the mean involves three kinds of knowledge: 1) functional - understanding the mean as a real world concept, 2) computational - knowledge of correct computational formulas, and 3) analog - a visual-

kinesthetic image, which in the case of the mean might be a balance point (Hardiman, 1983). The primary concern in this study is with analog knowledge.

One possible source of analog knowledge useful for a deeper and more flexible understanding of means and, hence, weighted means, is a deep understanding of the concept of balancing. While the balance beam is often used in text books as an explanatory tool for teaching mean concepts, few students understand balancing well enough to benefit from the analogy. However, Hardiman (1983) found that subjects who received training in balancing concepts showed significant improvement in ability to solve weighted mean problems.

There were several variables in the Hardiman study which might account for the improved performance of her subjects. The current research attempted to replicate the Hardiman study and to determine the effective variables.

The issue for the current research was two-fold: 1) Does balance beam training that results in induction of the product-moment rule have an effect on ability to solve weighted mean problems, and 2) If so, which aspects(s) of the balance training are significant in improving performance in the weighted mean transfer task?

## CHAPTER II

### PREVIOUS RESEARCH

#### Hardiman: The Study

The Hardiman (1983) study is the only research dealing specifically with transfer of balancing knowledge to weighted mean concepts. The subjects in her transfer study were the same subjects who participated in her balance beam rule induction training described in Section I. College undergraduates were pretested for both balance knowledge and ability to solve weighted mean problems. Subjects who were classified as non-calculators and non-balancers (NC/NB) and subjects who were calculators and non-balancers (C/NB) were then randomly assigned within each group to either a control or a balance training condition.

#### Balance Training

Hardiman's balance training has been described in Section I. All subjects in her training condition met the criterion indicating that they had learned the product-moment rule and, on the average, took 49.0 trials to meet the criterion. Thus, her subjects had fairly extensive experiences with balance problems and it was assumed that, by meeting the criterion, they understood and were using the p-m rule to make their predictions. As was discussed in Section I, in a forced choice task such as this, using a

criterion of five correct predictions in a row of complex-conflict problems, it is possible some subjects met the criterion for p-m rule use by chance or by a combination of limited rule use, perceptual judgments and good guesses. However, most subjects gave other, verbal indications that they were in fact using the p-m rule.

#### Extra Judgement Training (EJT)

After subjects met the criterion for p-m rule use, they then were presented with an additional set of seven balance situations on the concrete beam used in the training phase (described in Section 1). For each configuration, they were told to predict which way it would tip and to state in which direction the system would have to be moved in order to put it in balance. In this part of the training, the acetate scale that was placed along the top of the beam was numbered continuously from one to ten and the blocks for a particular configuration were centered over the numbered marks. (In the first part of the training, the scale's units were marked with lines but were unnumbered.) Subjects were instructed to move the scale, with the blocks on it, in the direction they thought the configuration would need to move so that its balance point would be over the fulcrum. It is not clear from the study what effect this change in the scaling of the balance beam had on the transfer task, but it would seem that since a continuous number line more closely mimics

the type of scaling found in weighted mean problems, this part of the training was a significant factor in the transfer task.

### Transfer Test

In the transfer phase of Hardiman's study, trained and control subjects were given the five written problems in Table T1, which were designed to assess their understanding of the mean and of the weighted mean. They were asked to represent each problem on the beam and to also calculate the answer on paper, thinking aloud as they worked. Representing a mean problem on the beam involves making a number line in accordance with the values in the problem. Because subjects for the most part had no idea how to go about representing the weighted mean problems on the beam, they were presented first with a simple mean problem and the experimenter helped them represent it on the beam. After the first problem, subjects attempted to represent on the beam and solve the four remaining problems without help from the experimenter. They were allowed, however, to revise answers of previous problems. The transfer problems were designed to assess different types of mean and weighted mean knowledge. Control subjects, prior to the transfer test, were given unrelated statistical problems to solve in an interview situation, in order to equate interview experience.



Table 10: Hardiman's Transfer Test Problems

(Simple Mean) It is possible to view the mean of a set of numbers as the point at which the number line containing the set of numbers would balance if it were placed on a balance beam. You have a balance beam before you. Please represent the mean of 3 and 10 on the balance beam using the plastic scale as a number line and the blocks as weights.

1) A student attends College A for two semesters and earns a 3.2 GPA. The same student attends College B for four semesters and earns a 3.8 GPA. What was the student's overall GPA?

2) A number of people get on a large elevator. Three-fifths of the people are men and average 180 pounds. The remaining people are women and average 120 pounds. What is the average weight of the people on the elevator?

3) Person A and Person B are engaged in a weight maintenance program. Person A weights himself three times evenly spaced throughout the day and averages 185 pounds on a typical day. Person B weighs himself five times evenly spaced throughout the day and averages 211 pounds. What is the average weight of the two people?

4) A local shop employs several people who make the following salaries:

- 1 - owner/president.....30,000
- 2 - foreman.....10,000
- 3 - general workers.....8,000

The owner needed to calculate the average salary of a shop employee. She thought of two ways to do it: 1) add the three numbers together ---  $30,000 + 10,000 + 8,000$  ---, and divide by three, or 2) multiply each salary by the number of people paid that salary, add them together, and divide by fifteen. Which way would you calculate the average salary and why?

The findings from this study were that subjects who received balance training performed significantly better on the two weighted mean problems in the transfer test than did



untrained controls. However, there was no difference in ability to recognize that Problem 3, the weight maintenance problem, only looked like a weighted mean, but was, in fact, a simple mean problem. Balance training in this study also correlated with ability to represent problems on the beam and seemed to help subjects develop higher level rationales for their weighted mean calculations.

#### Hardiman: What Transferred?

The question for the current research was: What knowledge about balancing is helpful for understanding the mean and the weighted mean? The hypothesis tested was that subjects who do benefit from balance training (i.e. perform better on weighted mean problems after receiving the balance beam training), do so as a result of the extra judgment training and not from induction of the p-m rule.

The first phase of Hardiman's balance training, inducing the p-m rule, required and emphasized that the scale of the beam be one in which the distance units are measured on each side, beginning with zero at the fulcrum and incrementing symmetrically towards each end. In the second phase of the training, the extra judgment task, subjects were presented with multi-stack configurations of weights distributed along a continuously numbered scale.

If one looks at how one would represent a mean or weighted mean problem on the balance beam such that the beam could be used to find the answer, it is obvious that balancing experiences involving a continuous number line would be more analogous than one in which the scaling was from the center outward. Take, for instance, the GPA problem in Table T1. This would be represented on the beam by numbering the moveable scale from 3.0 to 4.0, in increments of .1 (i.e. 3.1, 3.2, 3.3...3.9, 4.0), and placing two blocks on the 3.2 mark, to represent the two semesters at College A, and four at 3.8 for the four semesters at College B. By moving the scale with the configuration of blocks on it until the beam balanced, the number that was then at the fulcrum would be the average grade.

It is difficult to find a visual correspondence between computing the p-m rule and the process just described of representing and solving a weighted mean problem on the beam. It is also difficult to imagine that a subject confronted with a mean or weighted mean problem would find anything in his representation of the problem that would trigger memory to access balancing concepts via an analogy to the p-m rule calculation.

Virtually nothing in the p-m calculation maps on to calculations of weighted means. To be sure, there is an

analogous underlying concept of a balance point between two equal (the simple mean) or unequal (a weighted mean) sets of weights (grades, heights, number per). But the analogy is only apparent after the problem is represented on the beam and the answer (the balance point) is seen to be equidistant from equal weights and not equidistant from unequal weights. The computational processes have little in common and the purposes of each type of calculation are antithetical to each other; i.e. the purpose of the p-m rule is simply to determine whether the system will balance, while the purpose of finding the weighted mean is to determine exactly where it will balance.

Furthermore, if one has an understanding, however incomplete, that a mean or average is a single figure to represent an entire set of numbers, then it seems that this knowledge is logically tied to a computation in which the total amount for the set is redistributed equally among the members of the set by dividing by the number of elements in the set. To expect people to make an analogy from this add-and-divide type of thinking to the multiply-and-add p-m thinking assumes an even more sophisticated knowledge of math than that needed to understand the mean and weighted mean in their simplest form.

The argument here is that knowing the p-m rule does not help people to understand and solve weighted mean problems.

However, this does not in any way preclude the possibility that the experience with the beam, and the resulting knowledge about balancing in general, that occurred during the process of inducing the p-m rule, did not have some positive effect on subjects' ability to understand the weighted mean concept. To the contrary, experience with two- and multi-stack conflict problems should be quite helpful in understanding the effects of unequal sets. But the relationship of this information about balancing to weighted mean problems is not easily seen and needs some type of bridging experience to connect the two concepts. The extra judgement task seemed to be just such a bridge.

Therefore, the current research attempted to replicate the Hardiman finding that balance training that results in induction of the p-m rule has an effect on ability to understand and solve weighted mean problems and to determine which aspect(s) of the training were significant. The hypothesis was that subjects exposed to both the balance training and the extra judgement task would perform better on the transfer task than those who only experienced the balance training. A second hypothesis was that there would be no difference in ability to solve weighted mean problems between those subjects who actually induced the p-m rule and those who did not, although experience with balance

problems, alone or in conjunction with the extra judgment training, might have a positive effect on the transfer task.

# CHAPTER III

## THE PRESENT TRANSFER STUDY

### Methods

The effects of two aspects of balance learning, rule induction and judging the direction of the balance point, on ability to solve weighted mean problems were investigated in a video-taped interview situation. Subjects were screened in a paper-and-pencil pretest for knowledge of the product-moment rule for the balance beam and for ability to solve weighted mean problems. Only non-calculators/non-balancers were used in the study. Subjects who received balance training were compared to control (no balance training) subjects in two weighted mean problem solving conditions: half received the extra judgment task from Hardiman (1983) before receiving the transfer test and half received no extra training.

### Pretest

Subjects. Subjects were the same undergraduates as in the Rule Induction (Section I) pretest and both pretests were administered in the same session.

Materials. The two weighted mean word problems in Table T2 below were presented on individual pages in a test



booklet that also contained instructions and the balance beam pretest problems.

Table T2: Pretest Weighted Mean Problems

1) Two boats of fishermen return from a weekend fishing trip. The four people on the first boat average 5 fish per person, while the two people on the second boat averaged 11 fish per person. What was the overall average number of fish caught?

2) There is a measure called Income Index. It ranges from low income to high income on a scale from 1 to 6. In one small town there are 200 families and the average income index is 2.8. In a second small town there are 400 families and the average income index is 3.6. What is the overall average income index for all the families in both towns?

Procedure. The pretest was administered to groups of 10-20 subjects. Subjects were given the test booklet and allowed to work at their own pace. No feedback was given on any part of the pretest.

Scoring. Subjects who answered one or both of the two weighted mean problems incorrectly were considered noncalculators.

### Transfer Phase

Subjects. Only subjects who are classified as both noncalculators (NCs) and nonbalancers (NBs) (see Section 1: Rule Induction Pretest for definition of nonbalancer) were used in all other phases of the study. One-half of the NC/NBs were randomly assigned to the balance beam rule

Induction training condition, and one-half were assigned to the control condition and received no rule Induction training. Each of the two groups were further divided into those who received the extra Judgment task (EJT) and those who did not. This yielded four groups for the Transfer Task: Trained/no EJT, Trained/EJT, Control/no EJT, and Control/EJT.

Materials and Procedure. All subjects were interviewed individually one to two weeks after the pretest. Trained subjects were those individuals who participated in the balance beam training described in Section I. Control subjects, prior to the transfer test, were interviewed in a 'think aloud' situation as they solved unrelated word problems, in order to equate them with the trained subjects for verbalization experience. The EJT and transfer tests were video taped with the subjects' consent.

EJT. In the EJT part of the study, subjects were seated at a table with the concrete balance beam and fifteen equal size wooden blocks before them. The experimenter (E) numbered a prescaled acetate strip from one to ten and placed it along the top of the beam. E then placed a configuration of blocks on the beam, holding the beam in a balanced position, and asked the subject to predict whether the beam would tip left, right or balance. After the subject made his prediction, the beam was released so that

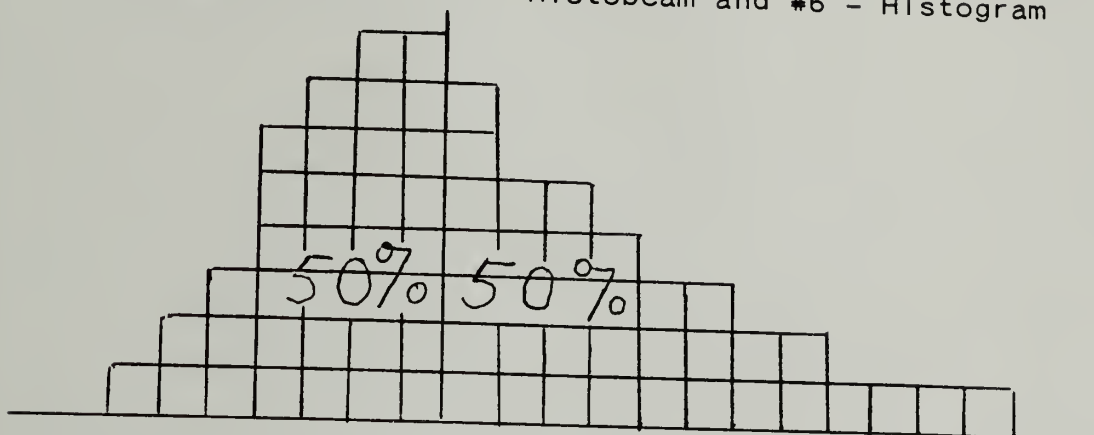
the outcome could be observed. The subject was then asked to slide the scale, with the blocks still on it, in the direction that would make the beam balance. This was repeated for each of the seven EJT configurations. All subjects in the EJT condition saw the same configurations in the same order.

Transfer Test Problems. The five written word problems in Table T1 (from Hardiman) plus the two problems in Table T3 were presented, one at a time, to the subject on individual sheets of paper. E sat across the table from the subject with the balance beam and blocks between them. The subject was supplied with scaled, but unnumbered acetate strips, a pen to mark the strips, and a pencil for calculating answers on the problem sheets. Subjects were instructed to read each problem aloud and to explain their solutions. For the first problem, the Simple Mean Problem, they were asked to represent the problem and its solution on the beam, numbering the strip and using the blocks, as well as to write down their calculations. If they could not figure out how to correctly represent it, the experimenter intervened. This was done to insure that all subjects had a basic idea of how to represent a mean problem on the beam. For the other six problems, subjects were given the same instructions except that they were told to "use the beam to solve the problems if you think it will be helpful," and

were also told that they would receive no further help from the experimenter.

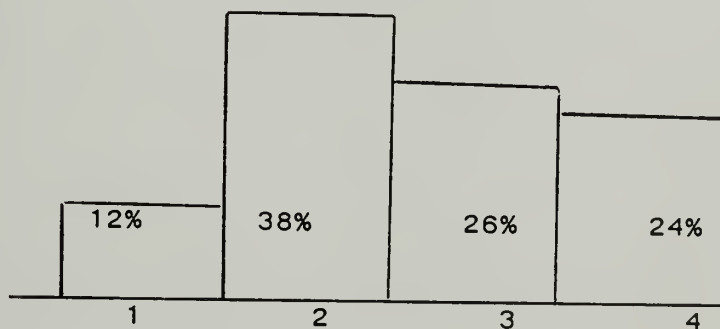
Table T3

Transfer Problems: #5 - Histobeam and #6 - Histogram



5) Imagine that this is a balance beam with a large number of blocks placed along it. The fulcrum is placed in such a way that half of the blocks are to the left of it and half are to the right. Would a balance beam with this configuration on it tip to the left, tip to the right, or balance? Please explain your answer.

Percent of  
Students at  
Each Quiz  
Score



Number Correct on Quiz

6) An Instructor gave a four question quiz. The grades were as above, with the percentage of students receiving each grade indicated in the graph. Would the mean (average) grade for the class be

a) less than 2.5?    b) equal to 2.5?    c) greater than 2.5?

## CHAPTER IV

### RESULTS

#### Analysis: Pretest and Transfer Data

The transfer data is presented in two forms: percentages of subjects who answered correctly and mean scores for each problem. Observations of the experimenter are included where appropriate.

The two weighted mean problems in the pretest and the six interview problems in the transfer session were scored as follows: In the pretest, a score of 1 (correct) or 0 (Incorrect) was given for each problem, for a possible total of 2; In the transfer interviews, a problem was given a score of 1 if both the calculation and the verbalized logic behind it were correct, a 0 if neither were correct, and .5 if one or the other was correct. For example, one interview subject incorrectly chose to compute the simple mean for the GPA problem, but said he thought the 4 semesters of 3.8 should count more than the 2 semesters of 3.2. For this problem, he was given a score of .5.

The first simple mean problem in the transfer session was not included in the scoring or the analyses, since its main purpose was to help subjects learn to represent problems on the beam. All subjects were able to compute the

simple mean on their own. However, most subjects required help from the experimenter to represent it on the beam.

### Pretest

The pretest for weighted mean calculation ability was administered at the same time as the pretest for balance knowledge. Of the subjects who were classified as non-balancers, 80 subjects were further classified as noncalculators. A subject was considered a noncalculator if he was incorrect on at least one of the two weighted mean problems in the pretest -- the fish problem or the Income Index problem. These non-calculator/nonbalancers were then randomly assigned half to the treatment and half to the control group.

Thirty-three (41.3%) of the 80 non-calculator/nonbalancer subjects solved only the fish problem correctly, 7 (8.75%) solved only the Income Index problem, and 40 (50%) were incorrect on both problems.

### Transfer Session

In the transfer session, weighted mean performance for the 40 subjects who received training on the balance beam was compared with that of the 40 untrained controls. The



two groups were further randomly divided into EJT and No EJT groups, resulting in four groups of 20 each: Trained/EJT (TE), Trained/No EJT (TN), Control/EJT (CE), and Control/No EJT (CN).

With the exception of the first simple mean problem, subjects were not required to represent the weighted mean problems on the beam, but were told they could do so if they thought it would be helpful in computing the answer to a problem. In this study, no more than two or three subjects attempted to represent problems on the beam and only one subject represented all of the problems.

The results of the transfer session will be presented as follows: the effects of 1) balance beam training, 2) EJT and 3) being a solver (inducing the product-moment rule for the balance beam). The effects will be considered first for the four numerical weighted mean problems (Problems 1-4) and then for the two graphic problems -- the Histobeam and the Histogram (Problems 5 and 6).

"Numerical" here refers to a word problem that presents numerical values to be calculated, as opposed to a "graphic" problem which presents a graphed, pictorial representation of the problem.

Problems 1-4

Balance beam training. Table T4 shows the percent (and number) of Trained and Control subjects who gave correct responses on the four numerical weighted mean problems. The difference between groups was negligible and it is clear that in this study balance beam training alone did not increase subjects' ability to solve weighted mean problems.

Table T4  
Percent Correct: Trained and Control Groups  
(Number correct in parentheses)

	<u>Trained</u>	<u>Controls</u>
GPA	20.0 ( 8.0)	22.5 ( 9.0)
Elevator	43.7 (17.5)	40.0 (16.0)
FoII	92.5 (37.0)	93.8 (37.5)
Shop	77.5 (31.0)	73.8 (29.5)
Histobeam	76.3 (30.5)	65.0 (26.0)
Histogram	26.2 (10.5)	17.5 ( 7.0)

EJT. Table T5 shows the percent (and number) of EJT and No EJT subjects, across and within training and control conditions, who correctly answered each of the problems. As a group, the subjects who received the EJT (TEs and CEs) did not perform significantly better than the No EJT subjects (TNs and CNs). Within the trained group, a t-test on the scored responses for the first four problems combined revealed no significant advantage for those who received the

EJT ( $t(38) = .0935$ , n.s.). Within the untrained Control group, the EJT seems to have had a reverse effect, but this also was not significant ( $t(38) = -1.21$ , n.s.). The data in Table T5 suggests a possible interaction between EJT and training. However, an analysis of variance on the scores showed the interaction to be not significant ( $F(1,76) = 2.39$ , n.s.).

Table T5

Percent Correct: EJT and No EJT Groups  
(Number correct in parentheses)

## EJT

	<u>TE</u>	<u>CE</u>	<u>Combined EJT</u>
GPA	22.5 ( 4.5)	10.0 ( 2.0)	16.3 ( 6.5)
Elevator	45.0 ( 9.0)	30.0 ( 6.0)	37.5 (15.0)
Foil	90.0 (18.0)	92.5 (18.5)	91.3 (36.5)
Shop	77.5 (15.5)	77.5 (15.5)	77.5 (31.0)
Histobeam	82.6 (16.5)	60.0 (12.0)	71.3 (28.5)
Histogram	25.0 ( 5.0)	15.0 ( 3.0)	20.0 ( 8.0)

## NO EJT

	<u>TN</u>	<u>CN</u>	<u>COMBINED NO EJT</u>
GPA	17.5 ( 3.5)	35.0 ( 7.0)	26.3 (10.5)
Elevator	42.5 ( 8.5)	50.0 (10.0)	46.3 (18.5)
Foil	95.0 (19.0)	95.0 (19.0)	95.0 (38.0)
Shop	77.5 (15.5)	70.0 (14.0)	73.8 (29.5)
Histobeam	70.0 (14.0)	70.0 (14.0)	70.0 (28.0)
Histogram	27.5 ( 5.5)	20.0 ( 4.0)	23.8 ( 9.5)

TE-Trained/EJT, CE-Control/EJT, TN-Trained/No EJT,  
CN-Control/No EJT.

Solvers vs. Non-solvers. Table T6 shows the percentage of balance beam solvers and non-solvers who gave correct responses on the transfer problems. Solver and non-solver responses are further divided into those who received the EJT and those who did not.

Table T6  
Percent Correct: Solvers and Non-Solvers  
(Number correct in parentheses)  
SOLVERS

	<u>EJT</u>	<u>NO EJT</u>	<u>COMBINED SOLVERS</u>
GPA	36.3 ( 4.0)	37.5 ( 3.0)	36.8 ( 7.0)
Elevator	50.0 ( 5.5)	68.8 ( 5.5)	57.9 (11.0)
FoII	90.9 (10.0)	100.0 ( 8.0)	94.7 (18.0)
Shop	81.8 ( 9.0)	68.8 ( 5.5)	76.3 (14.5)
Histobeam	81.8 ( 9.0)	75.0 ( 6.0)	78.9 (15.0)
Histogram	18.2 ( 2.0)	27.3 ( 3.0 )	26.3 ( 5.0)

#### NON-SOLVERS

	<u>EJT</u>	<u>NO EJT</u>	<u>COMBINED NON-SOLVERS</u>
GPA	5.5 ( .5)	3.8 ( .5)	4.8 ( 1.0)
Elevator	38.9 ( 3.5)	23.1 ( 3.0)	31.0 ( 6.5)
FoII	88.9 ( 8.0)	84.6 (11.0)	90.5 (19.0)
Shop	72.2 ( 6.5)	76.9 (10.0)	78.6 (16.5)
Histobeam	83.3 ( 7.5)	61.5 ( 8.0)	73.8 (15.5)
Histogram	33.3 ( 3.0)	19.2 ( 2.5)	26.2 ( 5.5)

The clear advantage of solvers over non-solvers that can be seen in Table T6 was tested in a 2 X 2 (solver/non-solver X EJT) analysis of variance on the scores for each subject for all four weighted mean problems. This analysis revealed a significant main effect of solving status; i.e.,

subjects who had solved in the balance beam task had significantly higher scores for the four numerical weighted mean problems than non-solvers ( $F(1,36)=5.81, p<.02$ ).

No interaction was found between solving status and EJT conditions.

From Table T6 it can be seen that most of the differences between solvers and non-solvers occur in the two calculation weighted mean problems (GPA and Elevator problems). This suggested an analysis of scores on just these two problems -- in relation to their pretest scores and in comparison to Hardiman's subjects.

Table T7

Mean correct responses per subject (0 - 2 scale) for Arriola pretest problems: Fish and Income Index combined; and Arriola and Hardiman transfer: GPA (#1) and Elevator (#2) combined.

	Arriola		Hardiman
	<u>Pretest</u>	<u>Transfer</u>	<u>Transfer</u>
Solvers	.58	.95	1.5
Controls	.53	.63	0.7
Non-solvers	.38	.36	---

Table T7 shows the mean scores (0 -2 scale) for solvers, non-solvers and controls on the two pretest weighted mean calculation problems (fish and Income Index combined) and on

the two transfer weighted mean calculation problems (GPA and elevator combined).

Pairwise *t*-tests confirmed the above analysis of variance results. Solvers in this study performed significantly better than non-solvers on the two transfer problems ( $t(38) = 2.916$ ,  $p < .005$ ). Comparisons between the scores of solvers and control subjects as well as between non-solvers and control subjects were not significant (S x C:  $t(57) = 1.4092$ ; NS x C:  $t(59) = 1.3724$ ).

The data in the table strongly suggests that balance beam training had more effect on ability to solve the transfer problems when the training resulted in inducing the product-moment rule. However, an analysis of variance on the pretest and transfer scores for solvers and non-solvers did not prove the interaction to be significant, although it was marginally non-significant ( $F(1,38) = 3.43$ ,  $p < .07$ , 1 tailed). As will be discussed later in more detail, this advantage for solvers could possibly be solely a result of the more positive attitude they had from their success on the beam when they later worked the transfer problems.

This difference did not appear on the foil problem (#3). One possible explanation is that all of the subjects came into the study with a similar high baseline understanding of simple means, knowledge acquired prior to and independently of the training provided in the experiment. In order to



look more closely at this, a foil problem would need to be included in the pretest. A more likely possibility is that solvers' and non-solvers' similarly high scores on the foil are attributable to very different causes. Solvers' high scores may reflect their understanding of the difference between weighted means and simple means, as evidenced in problems #1 and #2, while non-solvers' high scores might simply reflect their reliance on simple mean calculations for all "average" problems.

On the shop problem, #4, solvers may again have been exhibiting a better understanding of means while non-solvers, unsure of the correct answer to choose, may have adopted the heuristic that, if it wasn't obviously a simple mean problem, it must be the "other" answer. That is, when they were in doubt, they may have relied more on probabilistic test-taking strategies than on knowledge about means.

The equally high performance of solvers and non-solvers on #5, the histobeam could be attributed to all subjects having the same intuitive understanding of "a weight farther out has more 'force,' coupled with the effects of the training, which required that they be verbally explicit about their intuitions and for which they received confirming feedback.

The equally low performance of solvers and non-solvers on #6, the histogram, must be looked at in terms of the several different areas of knowledge, in addition to balancing, necessary to read and understand graphed frequency distributions. This is discussed more fully in the next section.

As shown by Tables T4-T6, the foil (#3: weight maintenance) problem and the answer choice (#4: shop) problems did not provide any means to differentiate between trained and untrained subjects (Table T1), EJT or No EJT conditions (Table T2), or solvers and non-solvers (Table T3). Problem #3, the foil problem, was a simple mean problem with a weighted mean surface structure. Over 90% of subjects in each group calculated this problem correctly and most also gave a correct rationale for choosing the simple mean calculation. Problem #4, the shop problem, asked subjects to choose between two calculation methods for finding an average salary for a shop employee. The percent of subjects in each group who chose the correct answer to this problem ranged from 70% to 78.8%.

#### Discussion: Problems 1-4

The main finding of this transfer study was that subjects who had solved in the balance beam task (i.e. induced the product-moment rule) performed significantly

better on the transfer weighted mean problems than non-solvers. Training alone (experience with the balance beam problems) did not improve performance nor was support found for the hypothesis that the key factor in Hardiman's transfer success was the extra judgement training she gave them on the beam after they had induced the rule and just prior to her transfer test.

The finding here that balance beam solving is associated with improved weighted mean solving ability appears at first glance to replicate and provide additional support for Hardiman's finding that balance beam knowledge transfers to understanding of the weighted mean. However, caution should be exercised before assuming that the same inference can be drawn from both studies.

First, Table T7 shows that, while solvers in the present study did in fact improve their pretest calculation scores, performance on the two calculation transfer problems was relatively poor compared to that of Hardiman's subjects. (Note that all of Hardiman's subjects are included in the analysis as solvers, since all of them eventually met the criteria for inducing the balance beam rule.)

Second, the two studies differed in several important ways such that the gains in performance in each study might be attributed to very different factors. The most noticeable difference was that Hardiman required her

subjects to represent both the GPA and elevator problems on the beam, while in this study subjects were told it was optional. Also, in the video tapes of Hardiman's transfer sessions, it was possible to detect instances where the experimenter responded helpfully or was not entirely neutral to a subject's difficulty in representing these problems.

The required representation in Hardiman's study may have provided a strong situational cue to use balance beam knowledge in the transfer task. Or, it is possible that working through a representation of a weighted mean problem, possibly with help from the experimenter, gave Hardiman's subjects the chance to test several approaches to thinking about means in a concrete situation that provided a certain degree of feedback. Furthermore, when a problem was properly represented on the beam, the subject then had both a visual and numerical answer before him that could be then used to guide paper and pencil calculations.

One other factor that could account for the higher average scores of Hardiman's subjects was that her subjects were permitted to change their answers to problems they had calculated earlier. Thus, later problems, such as the shop problem, might have clarified their thinking about weighted means and triggered a correct recalculation of a previous problem.

This is in strong contrast to the current study in which only two or three subjects opted to represent the problems on the beam either before or after calculating their answers on paper, and no subjects changed a previous answer. Without being forced to represent problems, subjects may have simply used the first calculation method that came to mind, usually the simple mean algorithm.

The high level of performance on the foil (weight maintenance) problem, in contrast to the overall low level on the GPA and elevator problems, could possibly be attributed to subjects' greater familiarity with simple mean calculations. It seemed to the experimenter that, by the time subjects were presented with the foil problem, they were generally alerted to the fact that not all "average" problems were the same, but not equally knowledgeable about when and how to calculate anything other than a simple mean. Once they had decided that the foil problem was a simple mean situation, they could easily apply the more familiar simple mean algorithm. However, the observation of the experimenter was that when subjects suspected a problem was not a simple mean situation, and were consequently unsure how to proceed, they sometimes correctly weighted the numbers and sometimes, by default, fell back on the simple mean algorithm. Alternatively, the fact that forty-one (17 trained and 23 controls) of the subjects who were correct on

the foil problem gave wrong answers on the preceding GPA and elevator problems suggests that subjects might have been using the simple mean calculation by rote. This would mean that they were correct on the foil by default. However, while a detailed analysis of the video tapes of interviews has not been completed, it was the experimenter's observation that many subjects gave reasonable rationales for deciding that the foil was a simple mean situation. For example, several subjects said "It doesn't matter how many times a day they were weighed -- a person only has one weight."

It is doubtful that balance knowledge was influencing their thinking on this problem since no subject gave any indication that he or she was using balancing knowledge as a basis for reasoning and no subject attempted to represent the problem on the beam. Rather it is more likely that subjects were relying only on their algorithm knowledge of means.

Interestingly, the shop problem seemed to be more confusing for subjects than the weight maintenance problem. Only 75% of the subjects answered the former correctly. Intuitively, one would expect that, since such a high percentage of subjects were able to see through the weighted mean surface structure of the immediately preceding foil (weight maintenance) problem, they would perform as well or



better on the shop problem where the difference between simple and weighted means is clearly defined in the problem's two answer choices. Subjects spent much more time deliberating over the two choices in this problem than in solving the foil problem, and expressed much less confidence in their answer. Subjects gave various reasons for choosing the simple mean description, the most common of which was that they thought the two calculation methods would result in the same answer and chose the first, the simple mean, because it was "faster." Again, there were no indications in the data or interviews that balance knowledge was playing any part in their thinking.

It was argued earlier that virtually nothing about calculating the product-moment rule maps onto the calculations for the weighted mean. And yet Hardiman found a strong relationship between inducing the rule and improved ability to solve weighted mean problems. Whether this was a mapping of balance rule knowledge onto weighted means or an effect of training in general is difficult to determine since she did not have trained non-solvers with whom to compare her solvers.

In the present study, the advantage of solvers over non-solvers in the transfer task, while significant, is still not a large enough gain over pretest performance to conclude

that balance beam rule knowledge has a substantial effect on subjects' reasoning. Solvers may have been simply the smarter subjects or, as was argued in Section 1, they may have been the subjects who possessed good general problem solving skills and, hence, were more likely to improve regardless of prior balance training.

Other factors that should be considered include the fact that 75% of the solvers solved within thirty minutes. When they returned, after a five-minute break, to begin the transfer session, they were far less fatigued than the non-solvers, most of whom took over an hour and a half to complete the balance beam session. They were also not feeling frustrated and defeated like the non-solvers. Degree of fatigue and feelings of self-confidence are not trivial factors and it does not seem unreasonable to attribute a large portion of the solvers' success in the transfer task to these motivational variables (Dweck, 1975).

#### Problems 5 & 6

An analysis of variance did not reveal any significant main effects of balance beam training on ability to solve the histobeam (#5) and histogram (#6) problems ( $F(2,78)=1.41$ , n.s.). Nor did solvers perform significantly

better than non-solvers ( $t(38) = .3027$ , n.s.) or controls ( $t(57) = 1.4167$ , n.s.). However, Table T4 shows a trend favoring training for Problem #5, the histobeam. There was also a tendency for trained-plus-EJT subjects (TE) to be correct more often on #5 than either the CE, TN or CN subjects (Table T5).

On Problem #6, the histogram, trained subjects showed only a small, non-significant advantage over the controls ( $t(78) = .9216$ ,  $p < .2$ , 1-tailed), EJT had no effect (Table T5), and, as Table T6 shows, there was no difference between solvers and non-solvers.

#### Discussion: Problems 5 & 6

Problems #5 and #6, the histobeam and histogram, were included to test the breadth of transfer of balance beam knowledge. At issue was whether experiences with the balance beam or inducing the rule for the beam would provide a helpful analogy for reading the mean of a graphed frequency distribution such as the histogram. The histobeam is of interest as a possible graphic bridge between the balance beam and the histogram.

Histobeam (#5). The histobeam problem specifically asked subjects to imagine that the pictured distribution was "a balance beam with a large number of blocks placed along it," making the connection to the balance training more than

obvious. One would expect, therefore, that trained subjects would perform significantly better than controls. The finding that there was no difference between trained and untrained subjects may possibly reflect that, in both groups, there were equal proportions of subjects who knew, prior to the experiment, that "weight farther out has more force" to subjects who did not have firm hold of this concept. In the trained group, for example, there was verbal evidence during the balance training from at least four subjects that they were not at all sure in which direction a block would have to move (towards or away from the center) to gain in "force."

Histogram (#6). Overall, performance on the histogram was extremely poor. Only 21.9% of all subjects responded correctly on this problem, well below chance level. Neither training nor EJT improved performance and, within the trained group, solvers were wrong nearly as often as non-solvers. Clearly, in order to understand the histogram, subjects needed more knowledge than the balance beam training supplied.

The main factor that appeared to contribute to subjects' inability to answer the Histogram problem correctly was a disregard of the values in the 1-4 scale on the graph (i.e. the "Number Correct on Quiz" axis). Their comments suggest that, as far as the graph was concerned, "1", "2", "3", and

"4" were just labels for locations on the axis; their actual value had no meaning or importance. Thus subjects reasoned that the midpoint or mean should be determined solely by finding the point that divided the number of students in the problem exactly in half -- a simple mean calculation.

Of the 17.5 correct responses (out of 40 trained plus 40 controls), only four actually calculated (on paper) the weighted mean for the Histogram problem. The others gave no reason or simply said the average grade "seemed like" it should be greater than 2.5. On no occasion did the experimenter observe a subject refer to the balance beam as a basis for responding. Obviously, a more revealing method of investigation is necessary to determine what role, if any, the balance training was playing in their reasoning.

### Conclusions

The conclusion to be drawn from this study is that the balance beam may indeed be a good analogy for understanding weighted means, but the distance between such disparate domains is too "far" to render the analogy useful just from experiences with it. Hardiman had fairly dramatic transfer effects, but the question is raised as to just what role representing played in causing these effects. In the

present experiment, where representing was not a requirement and no hints or help were given to help subjects utilize balance knowledge, it does not seem surprising that transfer did not occur. Unless the mapping is more direct between analogous situations, or a second, structurally similar analogy is provided, the connection between the analogous situations may never become apparent (Gick & Holyoak, 1983; Holyoak & Koh, 1986). These would be interesting avenues for further empirical studies, as would be the question as to which aspects of balance knowledge transfer.

Therefore, consistent with the literature and taking into consideration the questions raised by the Hardiman study, the conclusion here is that more than just a "deep understanding" of an analogous task is necessary to facilitate "far" transfer.



# APPENDIX A. Sequences for Conditions I - IV

COND. I & II			COND. III			COND. IV		
FR#	PROB	CODE	FR#	PROB	CODE	FR#	PROB	CODE
1	0020	0100	1	1110	4200	1	0020	0300
2	0020	0200	2	0400	1410	2	0100	0020
3	0200	0200	3	0200	0300	3	0021	1200
4	0200	0030	4	0203	0030	4	2000	0001
5	0200	0300	5	0010	2000	5	0301	1030
6	0300	0400	6	0111	4000	6	0001	0010
7	3000	0040	7	0300	0300	7	0002	0010
8	0001	4000	8	0030	6000	8	0003	0010
9	0010	4000	9	0111	6000	9	0004	0010
10	0100	4000	10	0020	0200	10	0003	0010
11	1000	4000	11	5000	0060	11	0020	2000
12	3000	0004	12	0100	4000	12	0020	3000
13	3000	0040	13	1110	5200	13	0020	4000
14	0300	0400	14	0005	0002	14	0020	5000
15	0030	4000	15	0040	0201	15	0020	4000
16	0030	6000	16	0033	0002	16	0010	4000
17	0300	0300	17	0400	0310	17	0010	3000
18	0300	0400	18	0300	0031	18	0010	2000
19	0300	0500	19	3000	0004	19	0010	1000
20	0030	3000	20	0010	2000	20	0010	2000
21	0030	4000	21	0100	3000	21	0200	0200
22	0030	5000	22	0030	6000	22	0200	0300
23	0030	6000	23	3000	0040	23	0200	0400
24	0005	0002	24	0200	0200	24	0200	0300
25	0005	0020	25	0400	1320	25	0100	0200
26	0005	0200	26	0200	6000	26	0101	0200
27	0005	0300	27	2020	0600	27	0022	0002
28	2000	0030	28	1100	3200	28	1020	4200
29	3000	0040	29	2100	0600	29	0030	2300
30	4000	0050	30	0300	1120	30	1010	0020
31	5000	0060	31	0300	0400	31	0013	0001
32	0010	2000	32	0020	0100	32	0022	0002
33	0100	3000	33	0040	0111	33	0102	0020
34	0011	3000	34	0040	1011	34	0200	1300
35	0111	4000	35	0040	0110	35	1100	2200
36	0111	6000	36	0400	0031	36	0040	0310
37	0200	6000	37	0011	3000	37	0012	0200
38	0400	0310	38	0005	0200	38	0030	1200
39	0400	0320	39	0001	4000	39	0006	0120
40	0400	1320	40	0040	0011	40	0040	0011
41	0400	1410	41	0300	1030	41	0203	0030
42	0300	0031	42	0005	0020	42	0030	0011
43	0400	0031	43	0111	4000	43	0220	0002
44	0500	0031	44	0220	1030	44	0024	0201
45	0401	0031	45	1000	4000	45	5000	0023

## Appendix A cont.

46	0101	0001	46	0032	0002	46	0030	0011
47	0040	0110	47	0101	0001	47	0040	4200
48	0040	0111	48	0030	3000	48	5001	0630
49	0030	0011	49	0220	0030	49	0040	0110
50	0030	0110	50	0030	4000	50	0011	3000
51	0030	1110	51	2101	0600	51	0040	0011
52	5000	0023	52	0300	0400	52	0400	0320
53	5000	1023	53	0030	0011	53	0500	2600
54	5000	0123	54	3000	0040	54	0060	0012
55	2000	0600	55	0011	3000	55	0400	0031
56	2010	0600	56	0200	0030	56	0040	0201
57	2100	0600	57	0111	6000	57	5000	0610
58	2101	0600	58	0023	0002	58	0032	0002
59	2020	0600	59	0030	4000	59	0022	0002
60	0222	0600	60	2000	0600	60	0023	0002
61	0400	0600	61	0400	0320	61	5000	0630
62	0022	0002	62	0030	5000	62	0400	0310
63	0023	0002	63	0040	0210	63	0300	1030
64	0024	0002	64	0010	4000	64	0040	0210
65	0033	0002	65	0500	0031	65	0030	0110

## REFERENCES

- Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch and B.B. Lloyd (Eds.), Cognition and categorization. Hillsdale, N.J.: Erlbaum.
- Dweck, C.S., (1975). The role of expectations and attributions in the alleviation of learned helplessness. Journal of Personality and Social Psychology, 31, 674-685.
- Hardiman, P.T. (1983). The importance of analog knowledge in understanding the mean. Unpublished manuscript, University of Massachusetts, Amherst.
- Hardiman, P.T., Pollatsek, A., and Well, A.D., (1986). Learning to understand the balance beam. Cognition and Instruction, 3, 63-86.
- Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1986). Induction: Processes of Inference, learning, and discovery. Cambridge, Mass.: MIT Press.
- Holyoak, K.J., and Koh, K. (1986). Surface and structural similarity in analogical transfer. Unpublished manuscript, University of Pittsburgh, Pittsburgh.
- Inhelder, B., and Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic Books.
- Karmiloff-Smith, A. (1984). Children's problem solving. In M.E. Lamb, A.L. Brown, and B. Rogoff (Eds.), Advances in developmental psychology: Vol. 3. Hillsdale, N.J.: Erlbaum, 39-90.
- Klahr, D., and Slegler, R.S., (1978). The representation of children's knowledge. In H.W. Reese and L.P. Lipsett (Eds.), Advances in child development and behavior, Vol. 12, New York: Academic Press, 62-116.
- Lesgold, A.M. Acquiring expertise (Tech. Rep. PDS-5). Pittsburgh: University of Pittsburgh, Learning Research and Development Center, January 1983.

- Lewis, M.W., and Anderson, J.R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. Cognitive Psychology, 17, 26-65.
- Mervis, C.B., and Rosch, E. (1981). Categorization of natural objects. Annual Review of Psychology, 32, 89-115.
- Siegler, R.S. (1976). Three aspects of cognitive development. Cognitive Psychology, 8, 481-520.
- Siegler, R.S., and Klahr, D. (1982). Why do children learn? The relationship between existing knowledge and the acquisition of new knowledge. In R. Glaser (Ed.) Advances in Instructional psychology: Vol. 2, Hillsdale, N.J.:Erlbaum, 121-211.
- Sweller, J., and Levine, M. (1982). Effects of goal specificity on means-end analysis and learning. Journal of Experimental Psychology: Learning, Memory and Cognition, 8, 463-474.
- Sweller, J., Mawer, R.F., and Ward, M.R. (1983). Development of expertise in mathematical problem solving. Journal of Experimental Psychology: General, 112, 639-661.
- Welsberg, R.W. and Alba, J.W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. Journal of Experimental Psychology: General, 110, 169-192.



