Title

Individual Differences In Mechanical Ability

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 9(0)

Authors

Hegarty, Mary
Just, Marcel Adam
Morrison, Ian R.

Publication Date

1987

INDIVIDUAL DIFFERENCES IN MECHANICAL ABILITY

Mary Hegarty
Marcel Adam Just

Carnegie-Mellon University

Ian R. Morrison

Interact R & D Corporation

Abstract

People who understand mechanical systems can infer the principles of operation of an unfamiliar device from their knowledge of the device's components and their mechanical interactions. Individuals vary in their ability to make this type of inference. This paper describes studies of performance in psychometric tests of mechanical ability. Based on subjects' retrospective protocols and response patterns, it was possible to identify rules of mechanical reasoning which accounted for the performance of subjects who differ in mechanical ability. The rules are explicitely stated in a simulation model which demonstrates the sufficiency of the rules by producing the kinds of responses observed in the subjects. Three factors are proposed as the sources of individual differences in mechanical ability: (1) ability to correctly identify which attributes of a system are relevant to its mechanical function, (2) ability to use rules consistently, and (3) ability to quantitatively combine information about two or more relevant attributes.

Introduction

We generally associate mechanical ability with a person's understanding of how machines work, the ability to build a machine out of its elementary components, and the ability to determine why a machine is not working. To understand a machine in this way, a person has to be able to identify the elementary components of the machine, know which properties of these elementary components are relevant to their function in the system, and also understand how these elementary components interact to accomplish the machine's function. This paper explores the mental models of individuals with different levels of understanding of machines.

One approach to understanding mechanical ability used by psychometricians is to measure the correlations between tests of mechanical ability and tests of other basic cognitive traits. Studies using this approach suggested that there were several components of mechanical ability, such as general reasoning ability, and knowledge acquired through

---

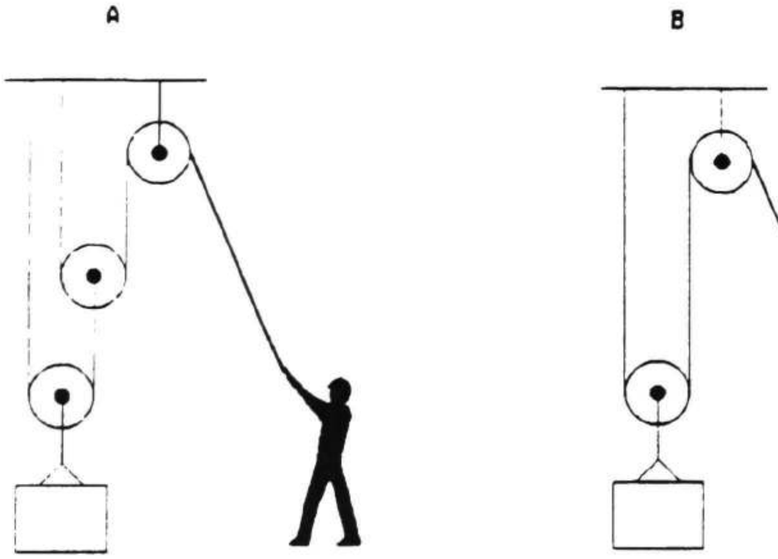
This work was supported in part by contract number N00014-85-K-0584 from the Office of Naval Research, grant number MH-9617 from the National Institute of Mental Health, and an NSERC post-doctoral fellowship to Ian Morrison. We thank Patricia Carpenter and Carolanne Fisher for their comments and suggestions.

experience with machines (Cronbach, 1984). This suggests that mechanical ability is not a static trait, but can develop as a result of experience.

Our approach includes an analysis of verbal protocols as well as an analysis of the response patterns obtained during the performance on test items. This approach allows us to examine the mental models of different individuals as reflected in which attributes of a mechanical system people consider relevant to its function, their rules relating the attributes to the function, their preferences among different rules, and their methods for combining rules pertaining to different attributes. The resulting models of high and low ability subjects are instantiated as two computer simulation models, whose performance on the test items produces patterns resembling those of human subjects.

We studied performance on pulley problems of the type used in the Bennett test of Mechanical Comprehension (Bennett, 1969). Paper-and-pencil texts such as this have been found to be highly predictive of performance in a number of technical fields such as machine assembly, mechanical repair and vehicle operation (Bennett, 1969, Ghiselli, 1955, Vernon and Parry, 1949). Our focus on pulleys permitted us to construct a large number of pulley problems which systematically varied the number and type of attributes that distinguished the two systems depicted in each problem. The Bennett type of pulley problems were at an appropriate difficulty level for our college-student subjects, allowing measurement of a range of individual differences in performance. Restricting the experiments to pulley problems does not compromise the generality of the research, since previous analyses of the Bennett test (Cronbach, 1984) and our own pilot study have shown that separate scores for different types of items are highly correlated. Thus, our examination of the mechanical ability that deals with pulleys should apply to reasoning about other types of mechanical systems.

Figure 1: A typical pulley problem.



With which pulley system does the man have to pull with more force to lift the weight?

A
B
If no difference, mark C.

Method

Problems. We analyze performance on 17 pulley system problems, including some items from the Bennett Mechanical Comprehension Test (Bennett, 1969) and other similar items which were constructed especially for this study. All of the items were multiple choice, requiring a selection among three response alternatives. Each problem depicted two pulley systems lifting a weight and asked which pulley system required more force to lift the weight (see Figure 1).

The two pulley systems depicted in each item differed on one or more of the following dimensions: mechanical advantage, weight to be lifted, height (rope length), and pulley diameter. Pulley systems that differed in mechanical advantage, also differed on some other attributes (relevant attributes), which are correlated with mechanical advantage, such as the number of load-bearing ropes and the number of pulleys.

Three types of problems differed in the kinds of attributes that distinguished the two systems depicted in the problem. In one type of problem, the two systems differed only on attributes irrelevant to the mechanical advantage of a pulley system (height or pulley size). In the second type of problem the two pulley systems differed in mechanical advantage, while the weights they lifted were equal. In the third type of problem, both the mechanical advantage and the weights were different for the two systems.

Subjects. The subjects were 43 undergraduate students, 27 students at Carnegie-Mellon University and 16 students at the Community College of Allegheny County. Fourteen of the students had taken two or more courses in physics at college level, while the remainder had taken no college level physics courses.

Procedure. Thirty-eight subjects were administered the test in a group setting, while five other subjects were tested individually and gave verbal protocols while they solved the problems. Two of the five protocol subjects had taken college level physics.

For the purposes of comparing different levels of ability, the data from the 38 subjects who performed the test in a group setting were divided into two groups, a high-scoring group and a low-scoring group, on the basis of their overall scores. A discontinuity in the distribution of scores defined the boundary between the high and low ability subjects. Twenty five subjects solved less than 59% of the problems correctly while thirteen of the remaining subjects scored more than 65% of the problems correctly. The high-scoring group therefore consisted of the top third of the distribution.

Results

General Solution Processes. An analysis of the subjects' verbal protocols suggested the following general account of how they solved the test items. The subjects decided which of the two pulley systems' distinguishing attributes (such as the number of pulleys) were relevant to reducing the effort required to lift the weight. They then compared the two systems using rules which relate these attributes to the amount of effort required.

The repertoire of rules used was inferred from the five subjects who gave verbal protocols. The rules pertain to those attributes that the subjects described as relevant, which were all attributes of the visible components of the systems - either their number, size, or attachments to other components. As Table 1 shows, most of the rules were based on system attributes that are correlated with mechanical advantage. Two of the rules were based on irrelevant attributes (height and pulley size). Two of the rules were quantitative, i.e., they expressed the effort as the ratio of the weight to some some attribute of the pulley system. The remainder of the rules were qualitative. A qualitative rule could state that pulley system with a higher value of some attribute requires less effort or that a system with a lower value the attribute requires less effort.

When two or more of a subject's rules were applicable in a problem, the rule that was used to generate the answer reflected a preference ordering among the rules. The preference ordering among rules implies that even if a subject knows a rule, he will not use it to generate the answer to the problem unless it is the most preferred in the situation.

Table 1: Rules used by the Protocol Subjects

Rule	Number of Subjects who used the Rule.
Qualitative Rules (relevant attributes):	
A pulley system with ... requires less effort	
less weight	5
more pulleys	4
more load-bearing ropes (tensions)	3
more attachments to the ceiling	2
more free pulleys	2
Qualitative Rules (irrelevant attributes):	
A pulley system with ... requires less effort	
larger pulleys	1
less height	1
Quantitative Rules (relevant attributes):	
A pulley system with ... requires less effort.	
less weight per pulley	1
less weight per attachment	1

Individual Differences. The response patterns of a large proportion of subjects could be classified as consistent with the rules observed in the protocols. These response patterns revealed that three factors accounted for individual differences in mechanical ability: (1) ability to discriminate relevant from irrelevant attributes, (2) consistency of rule use and (3) ability to quantitatively combine information about two attributes within a single rule. We will discuss each of these factors in turn.

High-scoring subjects were better able to discriminate relevant from irrelevant attributes of pulley systems. The majority of high-scoring subjects (92%) correctly identified height and pulley size to be irrelevant, while 52% of low-scoring subjects considered height to be relevant and 44% of low-scoring subjects considered pulley size to be relevant. This was reflected in the answers that they chose. High-scoring subjects chose a significantly higher proportion of correct responses (.90) than did low-scoring subjects (.44) in problems that varied the height of the system ($t(36) = 4.00$, $p < .001$). In problems that varied pulley size .98 of high-scoring subjects' responses and .35 of low-scoring subjects' responses were correct ($t(36) = 6.09$, $p < .001$).

High-scoring subjects used rules more consistently in problems that varied mechanical advantage and the rules that they used were more likely to be correct. If consistency is defined as having at least four out of six responses that are consistent with one rule, twelve of the thirteen high-scoring subjects responded consistently. In contrast only eleven of the twenty-five low-scoring subjects responded consistently. Seven of the thirteen high-scoring subjects were classified as using the rule that a system with more load-bearing ropes requires less effort, which gives the correct answer to all of the problems of this type. High-scoring subjects answered a significantly higher proportion (.77) of these problems correctly than did low-scoring subjects (.47), ($t(36) = 4.48$, $p<.001$)

High-scoring subjects also demonstrated the ability to quantitatively combine information about two attributes within a single rule. In problems involving both mechanical advantage and weight differences, the responses of ten of the high-scoring subjects (77%) were consistent with rules expressing a ratio of the weight to some attribute of the system, such as weight per load-bearing strand, attachment, or pulley. The low-scoring subjects, on the other hand, were more likely to base their comparisons of the systems either on weight or on a single attribute of the system, but did not combine the consideration of weight and the system attribute into a single rule. The most common rule used by these subjects was that more effort is required to lift a heavier weight. High-scoring subjects answered a much higher proportion (.62) of these problems correctly than did low-scoring subjects (.33) ($t(36) = 4.03$, $p<.001$).

In summary, high-scoring subjects are better able to identify the attributes relevant to the operation of a pulley system, they are more consistent in their use of rules, and they are more likely to use rules that indicate a quantitative understanding of pulley systems. Not only do the three factors have significant effects on performance, but they are also similarly related to the total scores, as assessed by the following procedure. Each subject was given a score of 1 or 0 on each of the three factors. A score of 1, based on the response pattern on the relevant problems, indicated that the subject had the ability measured by a given factor, while a score of 0 indicated that the subject did not have this ability. Each of the factors had a correlation with the overall score which lay between .49 and .51. Thus the three factors are of approximately comparable importance in predicting an individual's performance. Together the three factors accounted for 38.6% of the variance among the total scores.

A Model of Performance.

In order to specify mechanisms which can underlie the individual differences identified in the experiment, we developed a simulation model, written in Soar (Laird, Newell, and Rosenbloom, in press). The model simulates the performance of one high-scoring subject and one low-scoring subject who gave protocols in the experiment. It simulates the response choices that the subjects gave to the problems, as well as stating the rationale for each choice.

Representational Format. The model operates on a problem description for each of the 17 problems in the experiment. Each problem description contains all the information that is directly available to a human subject through visual inspection. However, not all of the information in the problem description is necessarily used by by the model or by the subject it simulates.

The format of a problem description is a structured description list which consists of identifiers and lists of attributes and values. There are four types of attributes: properties, relations, comparisons, and questions. The simplest type of attribute is a property of a pulley system or component of a pulley system, such as the number of pulleys in the system. The second type of attribute is a relation between two objects. For example, a relation might state that a particular pulley is fixed to the ceiling. The third type of attribute, a comparison, compares two properties or two relations. The fourth type of attribute, a question, contains an attribute with a missing value and states that the value should be obtained. The requirement in each item of the test, namely to compare the relative efforts required to lift the weights with the two depicted pulley systems, is represented as a question about the comparison of the effort attribute.

Production Rules. The simulation model uses a set of productions that can be divided into two subsets, one subset common to all subjects, and a second subset unique to the individual whose solutions were simulated. The common productions control the operators that seek information about the problem and the operators that generate answers to the question posed, express the reasons for producing these answers, and stop the processing when the final answer has been selected. The subject-specific productions determine what information an individual seeks and how he reasons from that information to generate an answer to the problem. These productions reflect the rules that a subject possesses relating attributes of pulley systems to their function (reducing the effort required to lift a weight).

The model can evoke one of two types of operators, elaboration operators and hypothesis operators. When a value in a question is missing, elaboration operators look for information in the problem statement that might be relevant to answering the question. Hypothesis operators suggest values for attributes that are sought by elaboration operators and use these values to suggest tentative answers to the problem. Each suggested answer is accompanied with a reason for this answer. For example, a hypothesis operator may suggest pulley system A requires a greater effort than system B because the weight that system A is lifting is heavier.

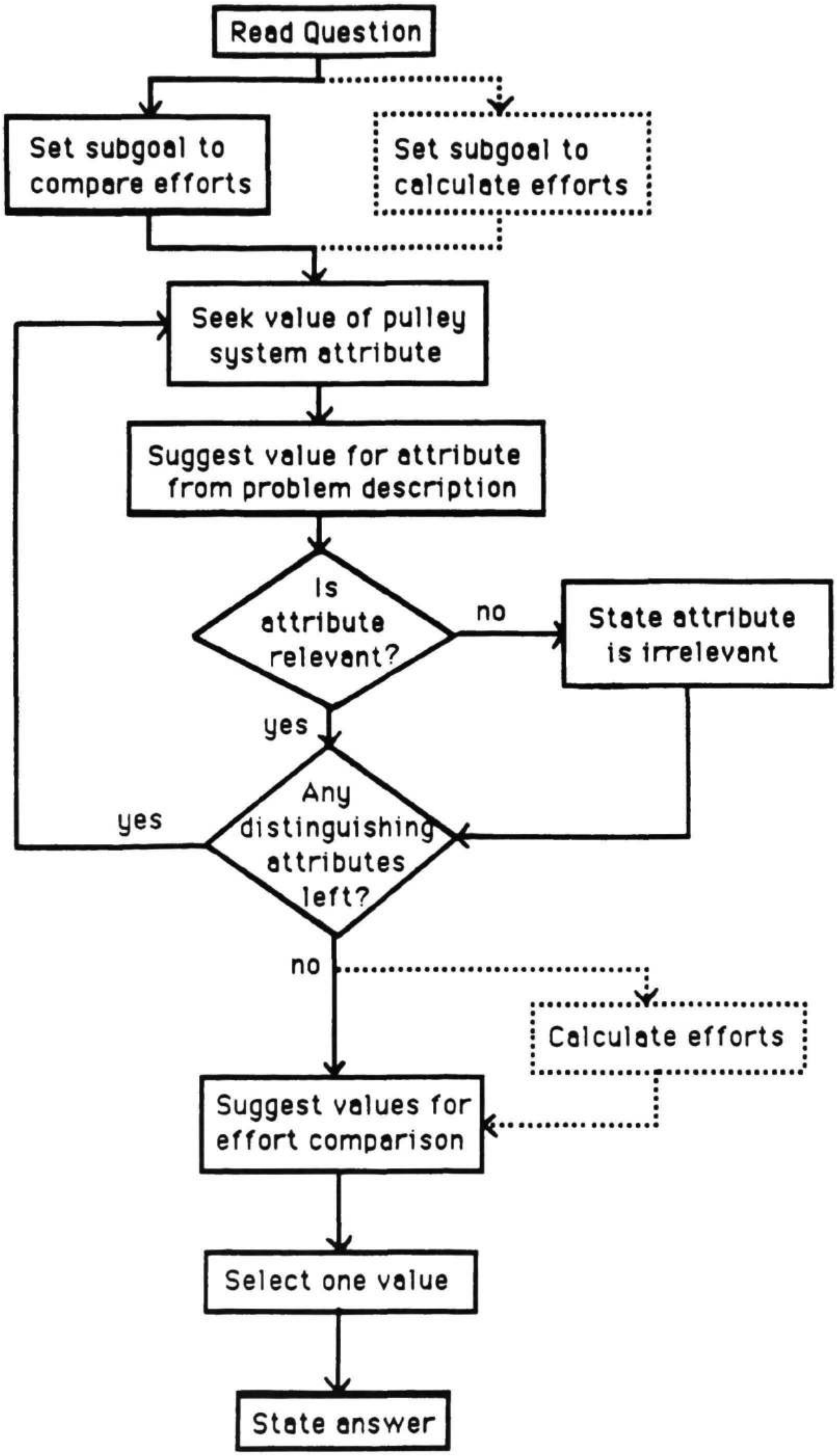
The productions choose among elaboration and hypothesis operators on the basis of preferences, expressed in Soar as special data elements. A preference might favor an answer supported by a particular reason. For example a preference might favor an answer based on the amount of weight to be lifted by a system over an answer based on the number of pulleys in a system. Alternatively, a preference might express a response bias. For example a preference might favor a hypothesis operator stating that the efforts required to lift the loads of the two pulley systems are different over an operator stating that the efforts are the same.

Flow of Control. The model proceeds from the problem description and question to its ultimate response by evoking a sequence of operators which derive information from the problem description and suggest answers on the basis of the obtained information (see Figure 2). When the question is first interpreted, an elaboration operator is evoked to seek the information that the question interrogates. The question ("With which pulley system does the man have to pull with more force to lift the weight?") interrogates a comparison of the effort attributes of the two pulley systems. Because there is no information available that allows this comparison to be made directly, additional elaboration operators are evoked to seek other information that might be relevant to the answer. For example, information about the number of pulleys or ceiling attachments in the two pulley systems might be sought at this point. In addition, if the person being modeled has sufficient knowledge to calculate the efforts required by the two pulley systems, a subgoal is generated to calculate the efforts. (The dotted lines in Figure 2 indicate components of the model that are present for subjects with this knowledge). Hypothesis operators use the information obtained by elaboration operators to suggest answers to the question. If no answer is suggested, the model chooses randomly among the possible answers. If only one answer is suggested, it becomes the response of the model for that problem. If more than one answer is suggested, a subgoal is created to resolve the tie. To satisfy the subgoal of resolving the tie, one hypothesis operator may be selected over another as a result of a preference. Otherwise, a random choice is made among the operators.

Modeling individual differences in performance. The three sources of individual differences observed in the experiment are modeled in the simulation in the following ways. To account for the differences among subjects in what they consider to be relevant, the model for a given subject relates the effort required in the case of a particular pulley system to precisely those attributes of the system that the subject considers relevant. That is, the attributes that were considered relevant were in the conditions of the productions embodying the mechanical rules. To account for the differences among subjects in how consistently they use one rule, the model varies or keeps constant its preferences among hypothesis operators across the different problems. If there is a preference for one hypothesis operator over all other hypothesis operators in a situation, the model will always choose the answer and the reason given by that operator in any similar situation. If there is no preference among operators, then the model chooses randomly among applicable operators, producing the same type of inconsistent behavior as observed for low-scoring subjects in the experiment. Finally, to account for the differences among subjects in their ability to quantitatively combine information from two relevant attributes, the model can either contain or not contain productions that suggest values for the effort based on a ratio of the weight of the system to some other relevant attribute.

The model simulated the performance of one high-scoring subject and one low-scoring subject who gave protocols in the experiment. The simulation for both subjects provided the same response and the same explanation of the response as the human subject in 16 of the 17 problems.

Figure 2: Flow of control of the simulation model through a problem.



The simulations for the high-scoring and the low-scoring subjects differed in the following ways:

1. The high-scoring simulation produces fewer suggested answers than the low-scoring simulation. This is because the low-scoring model makes suggestions about an irrelevant attribute in addition to making suggestions about a number of relevant attributes. All of the answers suggested by the simulation of the high-scoring subject were based on relevant attributes.
2. The high-scoring simulation calculated numerical values for the the efforts required to lift the loads of the pulley systems by quantitatively combining two attributes; weight and number of pulleys. The low-scoring did not attempt to determine the efforts directly and did not quantitatively combine attributes. When two or more attributes produced the same answer, the low-scoring simulation combined them into a single answer justified by the several explanations. If two or more suggestions produced contradictory answers, these comparisons canceled each other and the low-scoring simulation gave an answer of equality.
3. The high-scoring simulation organized the search for information so that numerical values for the efforts were calculated only when the answer could not be determined by qualitative comparisons. The order of search for information in the low-scoring model was random.

In summary, the model specified mechanisms which can account for the individual differences identified in the experiment. It successfully simulated the performance of a high-scoring and a low-scoring subject indicating the sufficiency of the theoretical proposal. The model suggested that the process of applying rules is similar for high-scoring and low-scoring subjects, but the content of the rules changes with increases in mechanical ability.

Discussion

The research reported in this paper provided both a general model of the processes involved in solving items from tests of mechanical ability and identified sources of individual difference in performance on these tasks. It was found that subjects encoded mechanical systems in terms of attributes of systems that they considered relevant to their function. Comparison of the pulley systems by different subjects was based on rules which expressed a relation between one or more of these attributes and the attribute in question i.e. the effort required to lift the weight with the pulley system.

According to the description of individual differences presented in the paper, low-scoring subjects are characterized as using qualitative rules based on both relevant and irrelevant attributes of pulley systems, and have no clear preferences among their rules so that their responses appear inconsistent with any particular rule. High-scoring subjects, on the other hand, can use quantitative rules when the problem demands the use of these rules, their rules are based on relevant attributes, and they prefer rules based on attributes that are highly correlated with mechanical advantage.

A striking feature of the range of solution processes used by subjects with different mechanical ability is their similarity to the developmental stages observed by Siegler (1978, 1981) in his analysis of young children's understanding of a balance beam. The parallel between our findings and developmental findings such as Siegler's suggests the intriguing hypothesis that the processes that underlie the development of mechanical abilities also characterize differences along an individual difference dimension. Our results suggest that mechanical ability should not be thought of as a static trait but as an ability that can develop with increased experience in this domain. This view of mechanical ability is consistent with the dominant view of mechanical ability in the psychometric literature, i.e., that it is a measure of understanding acquired through general exposure to tools and machinery (Cronbach, 1984).

Our results demonstrate that qualitative mental models of pulley systems precede quantitative models. In a qualitative model, attributes of pulley systems are coded by comparison with corresponding attributes of other pulley systems. A qualitative model also includes rules relating attributes of mechanical systems, situations in which it is appropriate to apply these rules, and preferences among these rules. Preferences can resolve conflicts between qualitative rules that are equally applicable in a situation, but there is no simple way in a qualitative model to resolve conflicts between rules with equal preference. In a quantitative model, on the other hand, attributes are given numerical values so that mathematical operations can be applied to these values to resolve conflicts.

The development of understanding of a physical system such as the balance beam or the pulley can be seen as a progression of mental models in which each model elaborates and refines the earlier models, rather than replacing them. The progression from low ability to high ability in mechanical ability involves advancing along a number of different dimensions. One dimension involves adjusting preferences between different rules so that rules based on relevant attributes are preferred to rules based on irrelevant attributes and at higher levels of mechanical ability, preferences among rules based on different correlated attributes correspond to how highly these attributes are correlated with mechanical advantage. A second advance is the progression from a qualitative to a quantitative model of mechanical advantage that enables the subject to quantify the extent to which mechanical advantage can reduce the effort required to lift a weight.

References

Bennett, C. K. (1969). *Bennett Mechanical Comprehension Test.* New York: The Psychological Corporation.

Cronbach, L. J. (1984). *Essentials of Psychological Testing (4th edition).* New York: Harper and Row.

Ghiselli, E. E. (1955). The measurement of occupational aptitude. *University of California Publications in Psychology.* 8(2), 101-216.

Laird, J. E., Newell, A and Rosenbloom, P. S. (in press). Soar: An architecture for general intelligence. *Artificial Intelligence*,

Siegler, R. S. (1978). The origins of scientific reasoning. In R. S. Siegler (Ed.), *Children's Thinking: What Develops?*. Hillsdale, NJ: Erlbaum.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46*(2), 1-149.

Vernon P. E. and Parry, J. B. (1949). *Personnel Selection in the British Forces.* London: University of London Press.