DOCUMENT RESUME

ED 282 751 SE 048 175

AUTHOR Collins, Angelo
TITLE A Description of the Strategic Knowledge of Experts Solving Transmission Genetics Problems.
PUB DATE 24 Apr 87
NOTE 41p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (60th, Washington, DC, April 23-25, 1987).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Processes; *College Science; Computer Assisted Instruction; Computer Simulation; Computer Uses in Education; Educational Technology; *Genetics; Higher Education; *Problem Solving; Science Education; *Science Instruction; Scientists; Secondary Education; *Secondary School Science
IDENTIFIERS *Science Education Research

ABSTRACT

Descriptions of the problem-solving strategies of experts solving realistic, computer-generated transmission genetics problems are presented in this paper and implications for instruction are discussed. Seven experts were involved in the study. All of the experts had a doctoral degree and experience in both teaching and doing research in genetics. Two types of data were available for analysis and for the description of the strategic knowledge that was used by the experts. These were the transcripts of the think aloud protocols and the computer printouts of the sequence of crosses for each genetics problem. Tables are provided which summarize the experts' strategies. Implications for instruction in solving genetics problems are reviewed in the areas of: (1) the utility of the model used for the study of problem solving; (2) the content knowledge of expert problem solvers; (3) clear and explicit information on use of computer-generated problems; and (4) what strategic knowledge to teach and when and how to teach it. A flowchart of the solution path used by experts to solve genetics problems is also included. (ML)

A DESCRIPTION OF THE STRATEGIC KNOWLEDGE OF EXPERTS SOLVING TRANSMISSION GENETICS PROBLEMS

Angelo Collins
251 Bluemont Hall
Kansas State University
Manhattan, KS 66506

Paper presented at the Annual Meeting of the National Association for Research in Science Teaching - Washington, D.C. April 23-25, 1987

A DESCRIPTION OF THE STRATEGIC KNOWLEDGE OF EXPERTS SOLVING TRANSMISSION GENETICS PROBLEMS

Introduction

Problem solving is an essential aspect of critical thinking, a topic currently receiving attention from both educators and the public. If reports such as Science and Mathematics in the Schools: Report of a Convocation (National Academy of Science, 1982) are any indication, problem solving is a topic of special concern to science educators. Concurrent with this interest is the problem solving research of cognitive scientists that provides science educators with insights into the nature of problem solving and which holds promise for educational practice.

One research approach used by cognitive scientists has been to study the problem solving performance of experts in content rich domains, especially physics. In an early study, Bhaskar and Simon (1977), studying an expert in thermodynamics, noted the consistent use of a single problem solving strategy, means/ends analysis. They also noted that the expert was consistent in performing a check of the solution. Chi, Feltovich, and Glaser (1981), comparing experts and novices solving mechanics problems, found that experts describe a problem in terms of the concepts of mechanics rather than in terms of incidental surface features. Larkin (Larkin & Rainhard, 1984; Larkin & Reif, 1979) claims that physics experts begin solving a problem by constructing descriptions of the problem at several levels. These

levels include a basic description taken from the facts of the problem statement, a scientific description which converts the facts to scientific concepts, and a computational description which reduces the relationships of the concepts to mathematical formulae. In a summary of their research on the problem solving performance of physics experts, Larkin, McDermott, Simon, and Simon, (1980) identify four characteristics of expert performance: 1) the conceptual knowledge of the expert is stored and retrieved hierarchically; 2) experts have ancillary knowledge of when and how to use the conceptual knowledge; 3) they begin to solve a problem by redescribing the data given in the problem statement in conceptual terms and mathematical relationships; and 4) experts, solving typical problems, use a forward-working, knowledge-producing strategy such as setting subgoals.

Synthesizing much of the research in problem solving in physics and providing a framework for further research, Reif (1983a; 1983b) has designed a comprehensive model for understanding and teaching problem solving in any natural science discipline. The comprehensive model includes a model of desired performance derived from descriptions of expert performance, a model of novice performance, a model of learning and a model of teaching. The two components of the performance models are the two types of knowledge required to solve problems, which Reif designates as content knowledge and strategic knowledge. He identifies three aspects of content knowledge: 1) the concepts and principles of the discipline; 2) the ancillary knowledge of when and how to use this conceptual knowledge; and 3) the structure of this knowledge. He also identifies three categories of strategic knowledge: 1) data

redescription strategies which enable the solver to identify the essentials of a problem and limit the problem space; 2) solution synthesis strategies by which the solver plans and executes ways to search the problem space; and 3) solution assessment strategies by which the solver decides if the answer is as complete and accurate as possible.

Although physics was the first science discipline in which pro solving was studied, transmission genetics is another area that is receiving increased attention from science education researchers. Paralleling the research in physics, Smith & Good (1983, 1984a, 1984b) have described the strategies of experts solving genetics problems. They identified 32 tendencies that can be used to differentiate between expert (or successful) and novice (or unsuccessful) problem solving performance in genetics. Among the tendencies of successful solvers that they identified are: 1) that they perceive a problem as a task requiring analysis and reasoning; 2) that they use knowledge-producing (forward-working) strategies, including setting subgoals; 3) that they begin solving the problem by investing initial time in qualitatively redescribing the problem; 4) that they make frequent checks of their work; and 5) that they use accurate bookkeeping procedures. Smith and Good found that experts also have a fund of accurate genetics knowledge which includes models of procedures for problem solving.

The problems studied by Smith and Good were challenging since they required the solver to analyze data about offspring and infer the genetic causes of the data, but the problems were taken from textbooks. Typically, textbook problems tend to be well-structured and require the

students to use relatively few, recently-taught concepts to obtain solutions. Textbook problems are limited to the amount of data in the text. Real problems in science tend to be ill-structured and the solver must determine what conceptual knowledge is needed to obtain solutions. An area in which the performance of experts solving real problems has been studied is medical diagnosis. Shulman, Elstein and Sprafka (1978) have identified several characteristics of medical diagnosticians who were judged by their peers to be highly successful. These characteristics include: 1) that they are not limited to the cues (data) in the original problem situation but continuously produce additional data; 2) that the strategy used most often to make a diagnosis (solve a problem) is hypothesis testing; 3) that expert diagnosticians entertain several hypotheses simultaneously; 4) and that hypotheses are confirmed, revised or discarded in light of additional data.

Computer simulations make it possible to create realistic problem-solving environments in which the problems are ill-structured, like real problems, yet without the difficulties, such as cost and time, usually associated with real problems. Real problems in transmission genetics are not only ill-structured but also differ from typical textbook problems in form. In textbook problems, the solver is presented with a description of a trait (for example, height in pea plants) and variations (for example, tall and short) of parents and the inheritance pattern (for example, simple dominance) controlling the production of offspring. Given the limited, static data, the solution is to predict the distribution of the variations among the offspring

(3/4 of the offspring will be tall and 1/4 of the offspring will be short). To reach a solution requires cause to effect reasoning, that is, from the inheritance pattern to the distribution of variations among the offspring. In real genetics problems the researcher begins with observations about a population of organisms. The researcher selects parents with traits and variations of interest (decides what the problem is) and produces generations of offspring (data) until an inheritance pattern can be inferred. To reach the solution requires effect to cause reasoning. Realistic, computer-generated problems in genetics, such as problems generated by GENETICS CONSTRUCTION KIT (Jungck & Calley, 1984), provide an opportunity for students to learn to solve problems with the form and lack of structure of real problems.

Stewart (in press) claims that learning to solve realistic problems provides students with the greatest potential for achieving four important learning outcomes. These are: 1) knowledge of the concepts of a discipline; 2) the ability to recognize and use general problem solving strategies; 3) the ability to use these general problem solving strategies in instances specific to a discipline and to recognize and use problem solving stratgeies that are discipline specific; and 4) to understand aspects of the nature of science. In genetics, solving realistic problems provides students with opportunities to pose the problem, to use their knowledge of genetics to generate and evaluate data, and to arrive at justifiable explanations of their solutions.

A description of the strategic knowledge of experts solving realistic transmission genetics problems can contribute to the theoretical knowledge about problem solving in science by providing

insights into the characteristics of successful problem solving performance in realistic genetics problems. A description of the strategic knowledge of experts can also provide science educators with initial help in designing instruction to enable students to learn to solve realistic problems.

The primary purpose of this paper is to describe the problem solving strategies of experts solving realistic, computer-generated, transmission genetics problems. A secondary purpose is to suggest implications for instruction in solving realistic genetics problems.

Methods

GENETICS CONSTRUCTION KIT (GCK) (Jungck & Calley, 1984) was the strategic simulation program used to generate realistic transmission genetics problems. The simulation begins by displaying a population of field collected organisms with the sex and phenotype of each individual identified. The solver then selects individuals for parents and crosses them to produce offspring. Generations of offspring can be produced until the solver is able to infer the inheritance pattern operating on the population. Inheritance pattern is the term used to summarize the genetics knowledge required to match a phenotype (the trait and variation observed, for example green pea pods) with the genotype (the abstract, theoretical genetic factors causing the variation, often a pair of alleles expressed as paired symbols such as 'Gg'). A problem must have an inheritance pattern for each trait and the inheritance patterns are mutually exclusive. The most common inheritance patterns

taught in introductory biology are simple dominance, codominance, and multiple alleles. After the inheritance pattern has been inferred, the solver may decide that a modifier is also operating on the population. Modifier is the term used to describe a condition that may alter the distribution of phenotypes within an inheritance pattern without affecting the genotype to phenotype match. For example, the position of the alleles on the chromosome may result in some traits frequently being inherited together. Modifiers cannot exist independently of an inheritance pattern and more than one modifier may affect a single inheritance pattern at the same time. The modifers usually taught in introductory biology include sex linkage and autosomal linkage.

GCK can be programmed to generate populations of many types of organisms. In this study the phenotypes of the organisms were traits with the variations of insects. In a GCK problem an individual may have up to four traits. GCK organisms are diploid with homogametic females and heterogametic males. With GCK it is possible to construct problems with the following phenomena within the domain of classical Mendelian or transmission genetics: 1) simple dominance (dominance-recessiveness); 2) codominance; 3) sex linkage; 4) pleiotropy; 5) epistatsis and other gene interactions; 6) lethality; 7) multiple alleles; 8) penetrance; 9) autosomal linkage, synteny, coincidence and interference; 10) multifactorial inheritance with and without environmental effects; and 11) complex combinations of most of the preceding phenomena (Jungck & Calley, 1986).

The parameters actually used to construct classes of problems were: number of traits - two; inheritance pattern - simple dominance,

codominance, or multiple alleles; modifier - sex linkage or autosomal linkage. These classes of problems were chosen because they are typical of those used in high school and undergraduate biology instruction.

Seven experts solved realistic GCK generated problems. All of the experts have a doctoral degree and experience in both teaching and doing research in genetics. Experts were chosen to represent a variety of interests within genetics: population genetics, clinical genetics, molecular genetics, genetics and evolution, viral genetics, genetics and paramecium behavior. Each expert spent an hour with the researcher learning the mechanics of the computer program. At this time the experts were given the list of phenomena possible for problems generated by GCK, but were not told the parameters actually used in constructing the problems they were about to solve. After the initial hour, in order to eliminate discomfort and/or silent clues possible if the researcher were present, each expert spent four additional hours alone solving problems. Because the experts worked at their own pace and because the problem generator was random, every class of problems was not addressed by every expert and some experts did more than one problem in a class. The classes of problems attempted by each expert are presented in Table 1.

Table 1 Here

In the initial session with the researcher, the experts were also asked to think aloud while solving the problems. They were given written directions on thinking aloud such as "Don't mumble". On the

written directions were questions to ask themselves, such as "Why are you making the cross you are making?" with suggestions of points in the problem solving process to remind themselves to think aloud, such as while the program is producing offspring from a cross. It was also emphasized that the transcripts of the tapes of them thinking aloud provide part of the raw data of educational research, and that too much data is preferable to too little data. Evidence that this idea was readily understood by the experts is that all of the tapes have an almost continuous, relaxed flow of comments. Without direction, all of the experts addressed the researcher while thinking aloud. The transcripts were a rich data source.

Two types of data were available for analysis and the description of the strategic knowledge used by the experts: 1) the transcripts of the think aloud protocols and 2) the computer printouts of the sequence of crosses made by each expert for each problem (The printout includes the expert solver's solution and the computer-generated solution). These data are termed research data to distinguish them from the data about offspring generated by the expert while solving the problem, which are termed problem data. A sample protocol and a sample printout for a problem are found in Figures 1 and 2 respectively. The class of problems from which the protocol and printout are taken is a two trait problem with a simple dominant inheritance pattern and no modifers. This problem and this class of problems will be used as examples in the analysis.

Figures 1 & 2 Here

Analysis

The analysis and reduction of the data gathered from the performance of experts solving realistic genetics problems occurred in four stages. The first stage was to express the research data in terms of the concepts and principles of transmission genetics and group them into one of three categories: 1) about the problem data; 2) about an hypothesis that explains the results of a single cross, called a specific hypothesis; and 3) about an hypothesis about the inheritance pattern that could explain all the crosses and predict the results of additional crosses, called a general hypothesis. This first stage of data reduction required four steps. The four steps of the first stage of data reduction for the initial population and first cross for the example problem are shown in Table 2. Step 4 was to illustrate the dynamic, non-linear nature of the solution process.

Table 2 Here

The second stage in the reduction of the research data was to tabulate all the data refined in the first stage for all solvers for one class of problems. A table was constructed for each cross. Table 3 is the table for the first cross for all experts for the simple dominant problems they did.

Table 3 Here

Comments about problem data are coded in the row labeled redescription. If there was a comment on the number and types of variations, the code is 'v'. Comments on the number of classes of phenotypes are coded 'c'. Comments on missing classes of phenotypes are coded 'm'. If the expert used symbols such as letters instead of words to discuss the traits or variations, the symbol row is marked. For example, in Table 3, in the first column, the solver quoted refers to the straw, lobed class of phenotypes as the 'SL group'. Comments about general hypotheses were coded. For example, SD is the code for simple dominant. To code the research data about the specific hypotheses, a chart was constructed of six possible crosses based on the phenotypic variations of the parents and the offspring produced. Each cross was assigned a letter which was used for coding. For example, specific hypothesis C is the cross of an homozygous (individuals with like alleles, aa) recessive parent with another homozygous recessive parent producing offspring with one variation the same as the parents. Specific hypothesis F is the classic Mendelian cross of heteorzygous (individuals with unlike alleles, Aa) parents producing offspring with two variations in a 3:1 ratio. The row labeled type of cross was a quick reference to the parents having the same variation (L for like) or different variations (U for unlike). Observations about the research data that were not easily coded were noted in abbreviated form in the last row.

In the third stage of analysis, the data tabulated in the second stage were combined to describe the performance of all the experts for each class of problems. The descriptions were grouped into the three categories of strategic knowledge. Table 4 is the summary of the

research data about problem data redescription for simple dominant problems; Table 5 is the summary of research data about hypothesis testing, the solution synthesis strategy used in simple dominant problems; and Table 6 is a summary about confirmation, the solution assessment strategy used in simple dominant problems.

Tables 4,5, & 6 Here

The fourth stage of the analysis was to combine all the research data about the strategic knowledge of experts solving all the classes of problems considered in this study. The result of this analysis is the description of the strategic knowledge of experts solving realistic computer generated transmission genetics problems which follows.

Data Redescription Recall that the function of data redescription is to isolate the essentials of the problem and limit the problem space. The experts include in their data redescription statements about the number and name of the traits and variations. They also combine individuals with the same phenotypic variations and consider classes of phenotypes. Identifying the number of variations for each trait and the number of classes of phenotypes is helpful in hypothesizing about the inheritance pattern. In addition, the experts note any missing classes of phenotypes. For example, one expert says "...there are eleven different kinds, we've got eyes and bristles. There are only two types of bristles, hairless and singed, but for eyes we've got apricot, red, plum...Now what combination is not there...Let's count up...There are 1, 2, 3, 4, 5, kinds of females and 6 kinds of males. So we're missing a

class of females." A missing class of phenotypes by sex among the offspring of a cross may indicate that the sex linkage modifier is operating in that population. A missing class of phenotypes by variation or an unbalanced distribution of individuals by variation is an indicator that the autosomal linkage modifier might be operating in the population.

Data redescription always precedes the formulation of an hypothesis about an inheritance pattern or modifier. Therefore, for example, data redescription occurs at the beginning of the problem. One person begins "In this problem I suppose that all three genotypes are expressed as different phenotypes for tiny, specked and sable which would mean codominant or else that there are more than two alleles at the locus." Experts also redescribe the problem data in the course of the solution synthesis whenever an alternate hypothesis is formulated. Alternate hypotheses are formulated 1) when a cross produces new data that alters the essentials of the problem; 2) when the solver is unable to infer or confirm an inheritance pattern; and 3) when solvers realize they have made an error in data interpretation. One example of new data altering the problem is, "Even before I begin I am suspicous that there is something funny because there are no b (blistery wing) males...I'll do a bs (blistery wing, sepia eye) female with an ss (short wing, sepia eye) male cross...Oh, there are b (blistery wing) males, so much for that hypothesis. Now there are 8 groups and it looks like it is simple." Data redescription also occurs when a solver considers a hypothesis about a modifier and, in a multi-trait problem, when the solver begins to focus on the inheritance pattern of a different trait. In

considering a modifier one expert says, " I crossed an sc (scarlet ocelll, crinkled antennae) by a wb (white ocelli, blunt antennae) and Wow, yeah I got - wc's (white ocelli, crinkled antennae) are 2, sb's (scarlet ocelli, blunt antennae) are 1, sc's (scarlet ocelli, crinkled antennae) are 20 and wb's (white ocelli, blunt antennae) are 11. I can see clearly that I got an excess of parental types contributing to the heterozygotes that I used in the cross which suggests strongly that these are not independently assorting but linked."

By redescribing the data, the solver is able to limit the problem space to reasonable general hypotheses and consolidate and recall knowledge that has been obtained from the crosses that have been done so far.

Solution Synthesis Solution synthesis strategies are those used to plan and execute a search of the problem space and enable the solver to infer a solution. In realistic transmission genetics problems the solution strategy that is used by all experts is hypothesis testing. Experts formulate two types of hypotheses -- general hypotheses about the inheritance patterns and modifiers and specific hypotheses about the distribution of variations to offspring for each cross. Because new data is continuously produced, there is an interaction between the problem data, the specific hypotheses and the general hypothesis. One expert begins, "I've got 4 classes each of males and females so there is no reason not to think it is simple so I'll cross the dw's (dumpy wing, white eye) with the sc's (shiny wing, cinnabar eye) and all the offspring are dw (dumpy wing, white eye), so if d (dumpy wing) and w (white eye) are dominant, the offspring are all heterozygotes..." In

the example, the initial population data presents an organism with two variations for each of two traits. The redescription allows the expert to retrieve the knowledge needed to formulate an initial, tentative general hypothesis of simple dominance. The expert then chooses to cross parents with unlike variations, using the specific hypothesis that if the genotype of one parent is homozygous dominant and the genotype of the other parent is homozygous recessive, the offspring will be heterozygous and have a dominant phenotype to predict the distribution of variations among the offspring. This cross is then performed, and the results agree with the prediction. The newly generated data supports the specific hypothesis and the specific hypothesis helps the solver infer the general hypothesis. This interaction between data, specific hypotheses, and general hypotheses continues throughout the synthesis of the problem solution. Also, in the solution synthesis, for each inheritance pattern and modifier, there is a cross or class of crosses that, once performed and explained, assures the solver that the solution is justifiable. This cross is being termed the definitive cross. In simple dominance and codominance this definitive cross is the F(2) cross; in multiple alleles the class of crosses used to justify the solution includes two F(2) crosses. An F(2) cross is between two parents that are known to be heterozygotes with the distribution of variations to the offspring in a 3:1 (dominant:recessive) ratio. In the example begun earlier in this paragraph the expert continues solving the problem by using the offspring from the first cross, assuming they are heterozygotes, as parents in the second cross. This is an F(2) cross for both traits. The definitive cross in all classes of problems except

sex linkage requires the identification of heterozygous individuals. In this problem the expert has constructed heterozygous individuals by crossing parents with unlike phenotypes.

Once the inheritance pattern has been inferred, the expert continues to do crosses to decide if a modifier is operating on the population. Either because of indicators in the problem data and/or to assure themselves the solution is complete, experts usually consider both sex linkage and autosomal linkage modifiers. In testing for modifiers, the interaction between the problem data, the specific hypotheses and general hypotheses continues. There is also a definitive cross to justify each modifier. In sex linkage the definitive cross is between a dominant male and a recessive female, producing recessive male and dominant female offspring. In the two-trait autosomal linkage problems the definitive cross is between a parent that is heterozygous for both traits and another that is homozygous recessive for both traits. The indication that the traits are not independent is that the ratio of the distribution of the variations to the offspring is not the expected 1:1:1:1 ratio.

By formulating two types of hypotheses, and by generating additional data that are either explained by a hypothesis or predicted from a hypothesis, experts are able to infer solutions to genetics problems that are justifiable.

Solution Assessment Solution assessment strategies are used to assure the solver that the solution is as complete and accurate as possible. While determining the presence of a modifier in the problem,

the experts are assuring themselves that the solution to the problem is complete.

Experts assure themselves that the solution is accurate by confirmation, by collecting additional evidence beyond the definitive cross that they have reasonably inferred the inheritance pattern or modifier. Although the Chi square test is the statistical test to determine if the observed distribution of variations to offspring agrees with the distribution expected from the principles of transmission genetics, experts seldom use the Chi square test. Rather, they compare the ratios of the distribution of the variations by intuition, without the formal mathematical test. Experts also increase their confidence in the accuracy of the inheritance pattern and modifier hypotheses by doing additional crosses that are explained by or predicted from the general and specific hypotheses. Whenever possible, experts use more than one method of confirmation. One example of confirmation is, "I think now I'll do its reciprocal." Another expert says, "...this is basically the 9:3:3:1 - 20:9:5:2, which is very, very, very close. So I'm sure I know what is going on already. Might as well confirm it by a test cross." A third example of confirmation is the expert who says, "I think I'll just repeat that cross a few times to jack up the numbers before I pull out my calculator...Oh, the ratio is getting closer all the time."

By using mathematical tests and by generating additional data, solvers increase their confidence in the completeness and accuracy of the solution to each problem.

Summary The description of the strategic knowledge of experts used to solve introductory level realistic transmission genetics problems is

summarized in Table 7. The strategy of data redescription consists of identifying traits, variations and classes of phenotypes and their distribution. It occurs prior to the formulation of tentative general hypotheses. The strategy of solution synthesis is hypothesis testing. Hypothesis testing in the classes of problems considered requires a definitive cross usually using heterozygotes. The strategy of solution assessment consists of producing additional evidence to confirm the inferred hypothesis. Experts in genetics know when and how to use these strategies to successfully solve realistic problems. The solution is the identification, by inference from the problem data generated, of general hypotheses about inheritance patterns and modifiers. The expert, having tested and confirmed the hypotheses by using them to explain and predict data, has a high degree of confidence that the they are justifiable from the data.

Table 7 Here

Table 8 summarizes the genetics feature of each category of strategic knowledge used by the experts to infer the solution for each class of problems. For data redescription this feature is the characteristic of the problem data that the solver uses initially to limit the problem space. For solution synthesis this feature is the definitive cross used by the experts to justify the inheritance pattern or modifier. For solution assessment this feature is the methods of confirmation most frequently used for that inheritance pattern.

Table 8 Here

From the analysis of the research data on the performance of experts solving realistic computer generated transmission genetics problems, the description of this performance can lastly be summarized as a flowchart, Figure 3. In this flowchart there are many paths and many feedback loops but the three categories of strategic knowledge used in solving genetics problems - data redescription, hypothesis testing and confirmation - regularly recur. From the flowchart it is also evident that the opportunity to continuously produce problem data is essential for the solution of these realistic problems.

Figure 3 Here

Implications

From the description of the strategic knowledge of experts solving realistic transmission genetics problems one implication can be made about the utility of the model designed by Reif as a starting place for the study of problem solving in science. The categories of strategic knowledge identified by Reif to describe problem solving in physics - data redescription, solution synthesis and solution assessment - have been used to describe problem solving in transmission genetics. The details within each category are different for genetics problems and physics problems, but this is expected since the disciplines are different, and the realistic problems studied in genetics are not like the textbook problems studied in physics in structure and form. Among the differences are: 1) that in the physics problems the data is limited to what is given in the problem statement while in the genetics problems continuous data production is possible; 2) that in the physics problems the solution requires a mathematical formula while no mathematical formula exists for the solution of the genetics problems; and 3) that in the physics problems the solution has a numerical value while in the genetics problems the solution is a confirmed hypothesis. The fact that the genetics problems and physics problems are not similar but that the same categories of strategic knowledge can be used to describe problem solving performance in both disciplines, supports the utility of the model.

A second implication is about the content knowledge of expert problem solvers in genetics. This implication may be important both to

the study of problem solving and to the design of instruction in problem solving in science. Although content knowledge is not the emphasis of this study, it is evident that expert problem solvers in genetics have a large store of highly organized, easily retrievable information available for problem solving. The use of strategic knowledge could not be described without reference to the content knowledge - for example, of inheritance patterns and modifiers, of specific crosses, of traits and variations, of dominant and recessive variations, of phenotypes and genotypes, of homozygotes and heterozygotes. It is also evident that this content knowledge includes information of when and how to use the strategic knowledge. For example, the experts know that an F(2) cross yields data useful in testing the simple dominant inheritance pattern hypothesis, and that this cross requires heterozygous individuals. In the study of problem solving, further research is needed to analyze and explicate the content knowledge required for successful problem solving in genetics. Likewise, instruction designed to teach problem solving strategies in genetics cannot be independent of instruction in the content of the discipline.

Another implication important for the design of instruction in problem solving in genetics is the need to include clear and explicit information on the use of each of the three categories of strategic knowledge. Teaching problem solving using realistic, computer-generated problems is currently an atypical experience for an instructor. Even though the instructor may have more knowledge and experience than the students, the instructor does not know the correct answer before beginning the problem. The instructor becomes a co-researcher with the

students. In the context of solving realistic problems, the instructor and the students explore the problem together. However, since stratgeic simulation programs have only recently become available, neither instructors nor students have much experience in solving realistic problems. This is a realistic time to try to improve instruction. Data on how students solve realistic problems without instruction can contribute to the design of new instruction. Research by Albright (1987) and Slack (1987), using GCK problems, is in process to describe the strategic knowledge of novices at the high school and undergraduate biology levels. They are finding, for example, that novices do not begin GCK problems by identifying important aspects of the problem data (redescription). It is reasonable that instruction in solving realistic genetics problems include knowledge of the general strategy of redescription and specific details for redscription in solving genetics problems. If students are to realize the full benefits of learning to solve realistic genetics problems, it will not be sufficient for the instruction to merely identify strategic knowledge being used in the process of seeking a solution, reasons for its use will have to be clearly and explicitly identified. For instance, students may learn that it is important to identify the name, number and distribution of traits, variations and classes of phenotypes for data redescription at the beginning of a problem, but to be successful problem solvers, students also need to learn the content knowledge that explains why this information is useful in limiting the number of possible justifiable solutions.

As science educators work to design instruction for solving realistic problems, the development of artificial intelligence computer programs will result in new instructional strategies. MENDEL (Streibel, Stewart, Koedinger, Collins, & Jungck, 1987), is an artificial intelligence computer tutoring system for genetics problem solving. MENDEL has two computer program components: the GCK problem generator and a TUTOR. The TUTOR, in turn, includes a SOLVER program and an ADVISOR program. The SOLVER consists of frames that contain content knowledge and rules, derived from this study of expert performance, for the use of strategic knowledge. The design of the ADVISOR addresses some of the same instructional issues as the design of traditional classroom instruction. These include what strategic knowledge to teach and when and how to teach it and how to integrate instruction in strategic knowledge and content knowledge.

The advent of realistic, computer-generated problems has created opportunities for students to achieve important learning outcomes in science. As models for understanding and teaching problem solving develop and as technology makes the computer a powerful and available instructional tool, science educators need to continue to design instruction to provide students with improved learning experiences in problem solving. One step on toward achieving the goal of improved instruction and learning in problem solving is to describe the performance of succcessful problem solvers.

REFERENCES

Albright, W., (1987). Problem solving strategies used by university students, Master's thesis in progress, University of Wisconsin, Madison.

Bhaskar, R., & Simon, H.A., (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. Cognitive Science, 1, 192-215.

Chi, M.T.H., Feltovich, P.J., & Glaser, R., (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5(2), 121-152.

Elstein, A., Shulman, L., & Sprafka, S. (1978). Medical problem solving: An analysis of clinical reasoning. Cambridge: Harvard University Press.

Jungck, J., & Calley, J. (1985). Genetics: Strategic simulations in Mendelian genetics - User's manual, COMPress, Wentworth, N.H.

Jungck, J., & Calley, J. (1984). GENETICS CONSTRUCTION KIT, COMPress, Wentworth, N.H.

Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Larkin, J.H., & Rainhard, B. (1984). A research methodology for studying how people think. Journal of Research in Science Teaching, 21(3), 235-254.

Larkin, J.H., & Reif, F. (1979). Understanding and teaching problem solving in physics. European Journal of Science Education, 1(2), 191-203.

National Academy of Science & National Academy of Engineering. (1982). Science and Mathematics in the Schools: A Report of a Convocation, Washington, DC: National Academy Press.

Reif, F. (1983a, June). Understanding and teaching problem solving in physics. Lectures at the International Summer School on Physics Education, La Londe Les Maures, France.

Reif, F. (1983b). How can chemists teach problem solving? Journal of Chemical Education, 60(11), 948-953.

Slack, S., (1987). Problem solving strategies used by high school students to solve realistic genetics problems, Master's thesis in progress, University of Wisconsin, Madison.

Smith, M. (1983). A comparative analysis of the performance of experts and novices while solving selected classical genetics problems. Paper presented at the meeting of the National Association of Research in Science Teaching, Dallas, TX.

Smith, M., & Good, R. (1984a, April) A proposed developmental sequence for problem-solving ability in classical genetics: The trial and error to deductive logic continuum. Paper presented at the meeting on the National Association of Research in Science Teaching, New Orleans, LA.

Smith, M., & Good, R. (1984b). Problem solving and classical genetics: successful vs. unsuccessful performance. Journal of Research in Science Teaching, 21(9), 895-912.

Stewart, J.H., (In press). Potential learning outcomes from problem-solving instruction: Examples from genetics. Science Education

Streibel, M., Stewart, J., Koedinger, K., Collins, A., & Jungck, J.R.,(1987) MENDEL: An intelligent computer tutoring system for genetics problem-solving, conjecturing, and understanding. Machine-Mediated Learning, 2(1&2), 129-159.

The research was supported in part by grants from the Dean of the University of Wisconsin-Madison, School of Education and the FIPSE (Fund for the Improvement of Post Secondary Education) program of the United States Department of Education (Grant # G008301471).

Table 1: Classes of Problems Attempted by Each Expert

EXPERT	1	2	3	4	5	6	7	TOTAL
PROBLEM								
Simple Dominance	2	2	2	2	2	2	2	14
Codominance	2	2	1	2	2	2		11
Multiple Alleles	4	1		1	1	2		9
Sex Linkage	1	1	2	1	1		1	7
Autosomal Linkage	2	1	1	1		1	1	7
TOTAL	11	7	6	7	6	7	4	48

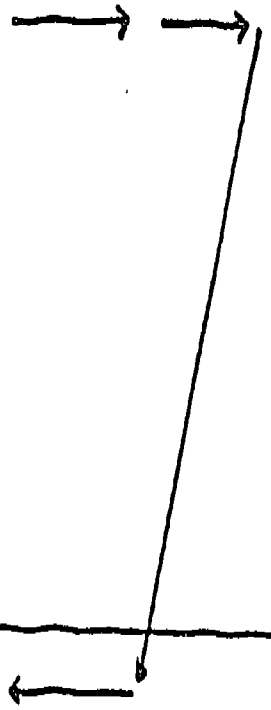
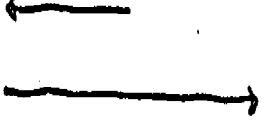
CROSS	Step 1 Read the transcript and mark it to correspond with the crosses	Step 2 Place the phrases of the transcript in groups depending on whether they refer to problem data (PD), specific hypothesis (SH), or general hypothesis (GH)	Step 3 Reduce the phrases of the transcript to transmission genetics concepts and add notes	Step 4 Draw arrows to represent the sequence and relationship of (PD), (SH), and (GH)
0	we're back to 8 phenotypes & 2 groups of characteristics yellow & straw & red & lobed. Start with a simple dihybrid cross. we'll just for fun assume that the least frequent phenotype is going to be doubly recessive & do it.	PD: 8 pheno 2 group charact yellow & straw red & lobed SH: least frequent is doubly rec. GH: simple di-hybrid	PD: classes traits varia-tions SH: aa x aa → aa double GH: Simple Dom note lfp = rec	PD SH GH 
1	I'll start with an SL by SL mating & we got all SL's. That's helpful.	PD: all SL SH: SL x SL mating GH: helpful	PD: traits SH: aabb x aabb → aabb GH: confirm SD	

Table 2 -- Stage One:
Data Reduction - Simple Dominance

Table 3 -- Stage Two:

Data Tabulation - Simple Dominance

Cross 1

Person	1		2		3		4		5		6		7	
Problem	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Redescription		m	v		c v not like real	c							m	
Symbol	x		x	x							x	x		
General Hypothesis	SD	other SD			SD		SD		SD		SD	SD		SD
Specific Hypothesis	C	C D			E		c		F		C	C		C/A D
Type of Cross	L	L U	U	U	U	L	L	U	U	L	L	L	L	L
Notes		separate traits ratio sex ratio	series	separate traits sex ratio create het	separate traits ratio create h cross error	ratio sex ratio cross error		separate traits ratio create het create miss class	separate traits	have hypo but no confirm	separate traits missing class		sex ratio series	separate traits

Table 4: Data Redescription -- Simple Dominance

1. Details of Initial Redescription

   -- 14 of 14 problems have some type of initial redescription

      -- 10 include comments on traits, variations and classes of phenotypes

      -- 2 include comments on traits and variations

      -- 2 include comments on the number of classes of phenotypes

      -- 5 note missing classes

      -- 4 note least frequent phenotypes; of these, 1 also notes most frequent phenotype

2. Additional Occasions of Redescription

   -- 2 problems are redescribed when the attention of the solver is focusing on the second trait

   -- 6 problems are redescribed whenever an alternate hypothesis is considered

   -- 4 problems are redescribed at the end of the problem

Table 5: Solution Synthesis -- Simple Dominance

1. Origin of the General Hypothesis

   -- 6 problems have the simple dominant inheritance pattern stated from the redescription of the initial population
   -- 6 problems have hypothesis stated after 1 or 2 crosses
   -- 2 problems have hypothesis stated after beginning a series of 4 or 5 possible crosses

2. Definitive Cross

   -- In 8 of the 11 successfully solved problems a monohybrid or dihybrid F(2) cross is used to match genotype to phenotype

      -- In 2 of these the heterozygote is constructed
      -- In 6 an obligate heterozygote is located

   -- In 3 of 11 successfully solved problems the linkage cross is used to match genotype to phenotype

      -- In 3 an obligate heterozygote is used

3. Alternate Hypotheses

   -- In 11 problems autosomal linkage as a modifier is considered and rejected

      -- 11 times after the inheritance pattern is confirmed
      -- 7 times by the linkage cross
      -- 4 times by a dyhybrid F(2) cross

   -- In 10 problems the sex linkage modifier is considered and rejected

      -- 6 times after the inheritance pattern is confirmed
      -- 2 times after the second cross
      -- 2 times it is rejected by the sex linkage cross
      -- 8 times the hypothesis is rejected because there is nothing to support it

   -- In 1 problem lethality is rejected because there is nothing to suggest it

   -- In 4 problems other hypotheses are considered -- sex influence, sex limited and interaction

Table 6: Solution Assessment - Simple Dominance

1. Mathematical

   -- In 8 of the 8 problems that use an F(2), ratios are used to confirm the inheritance pattern and genotype to phenotype match

      -- In 1 problem Chi square is used
      -- In 7 problems the solver says the ratio "looks ok"
      -- In 3 problems Chi squared is mentioned but not used

2. Strategic

   -- In 6 problems both an F(2) and a linkage cross with an examination of their ratios are used to confirm simple dominance

   -- In 4 problems the definitive cross is repeated with different individuals, in 1 case the reciprocals of the F(2) cross

   -- In 9 of 11 problems at least two methods of confirmation are used

Table 7: Summary of the Characteristics of Strategic Knowledge

1. Data Redescription

   -- Consists of

      -- number and name of variations
      -- number and name of traits
      -- number of classes of phenotypes
      -- missing classes of phenotypes
      -- unequal distribution of individuals to classes of phenotypes

   -- Occurs prior to formulation of a general hypothesis

2. Solution Synthesis

   -- Consists of hypothesis testing

      -- general hypotheses about inheritance patterns and modifiers
      -- specific hypotheses about crosses

   -- Occurs by

      -- using hypotheses to explain data generated by crosses
      -- predicting new data by crosses from hypotheses

   -- Requires

      -- interaction of data, specific hypotheses and general hypotheses
      -- performing a definitive cross using heterozygotes

3. Solution Assessment

   -- Consists of confirmation

   -- Occurs by collecting additional evidence

      -- through Chi square and other informal mathematical tests
      -- by doing additional crosses

   -- Includes more than one form of confirmation if possible

Table 8: Summary of Details of Strategic Knowledge

	REDESCRIPTION	SOLUTION SYNTHESIS	SOLUTION ASSESSMENT
	CHARACTERISTICS	DEFINITIVE CROSS	CONFIRMATION
Simple Dominant	2 variations/ trait	F(2)	Chi square linkage
Co-Dominant	3 variations/ trait	F(2)	Chi square linkage
Multiple Alleles	3-6 variations /trait	Series of crosses with an F(2)	Match all phenotypes to a genotype
Sex Linkage	Missing class of phenotype of one sex	Dominant m X recessive f	None
Autosomal Linkage	Missing or low frequency class of phenotypes	Linkage	Repeat cross with different individuals

Figure 1: Transcript of Think Aloud Protocol for Simple Dominant Problem

Well, fortunately we're back to 8 **phenotypes** and two groups of **characteristics.**

Yellow and straw and red and lobed.

Start with a **dihybrid** cross.

We'll just for fun assume that the least frequent **genotype, phenotype** is going to be doubly **recessive** and do it.

That means it's SL.

I'll start with an SL by SL mating.

And we got all SL's.

That's helpful.

Let's try a YR by SL cross and then do an **F(2).**

If it works the way I'm expecting.

OK YR by SL gives uh only YR's.

So presumably I happened to pick up a **homozygous** YR and now I have just **heterozygous** YR's.

So we should get a nice distribution by crossing them.

Let's see if this new line is basically a 9:3:3:1.

20:9:5:2 which is very, very close.

So I'm sure I know what is going on already.

Might as well confirm it.

Doing a **test cross.**

Let's see Vial 2 by Vial 3.

That gives a 14:10:8:8 which I'm sure is near enough to 1:1:1:1.

Y and R are **independently segregating** and are **dominant** over S and L.

Figure 2: Computer Printout of Simple Dominant Problem and Solution

```
********************* N E W   P R O B L E M *********************

                          Problem Type #1
************************************************************************

  Contents of Vial #1 (field collected population):
 8 F  Yellow      Red                    7 M  Yellow      Red
 1 F  Straw       Lobe                   1 M  Straw       Lobe
 3 F  Straw       Red                    1 M  Straw       Red
 2 F  Yellow      Lobe                   1 M  Yellow      Lobe

Entering CROSS....
       Vial #1  Phenotype #3  Individual #1  (f SL x m SL)
       Vial #1  Phenotype #4  Individual #1

  Contents of Vial #2 (offspring from cross above):
16 F  Straw       Lobe                  11 M  Straw       Lobe

Entering CROSS....
       Vial #1  Phenotype #1  Individual #2  (f YR x m SL)
       Vial #2  Phenotype #2  Individual #2

  Contents of Vial #3 (offspring from cross above):
20 F  Yellow      Red                   28 M  Yellow      Red

Entering CROSS....
       Vial #3  Phenotype #1  Individual #7  (f YR x m YR)
       Vial #3  Phenotype #2  Individual #8

  Contents of Vial #4 (offspring from cross above):
10 F  Yellow      Red                   10 M  Yellow      Red
 3 F  Yellow      Lobe                   2 M  Yellow      Lobe
 1 F  Straw       Lobe                   1 M  Straw       Lobe
 2 F  Straw       Red                    7 M  Straw       Red

Entering CROSS....
       Vial #2  Phenotype #1  Individual #8  (f SL x m YR)
       Vial #3  Phenotype #2  Individual #5

  Contents of Vial #5 (offspring from cross above):
 6 F  Straw       Red                    8 M  Straw       Red
 6 F  Yellow      Red                    2 M  Yellow      Red
 5 F  Yellow      Lobe                   5 M  Yellow      Lobe
 5 F  Straw       Lobe                   3 M  Straw       Lobe
```

Solver's Solution:

Dihydrid. Alleles Y and R are dominant over S and L, respectively. They appear to be completely independently segregating.

Program Solution:

```
Trait #1 (Body): There are 2 alleles.
Genotypes map to phenotypes as follows:
  1,1  IS  Yellow    2,2  IS  Straw    1,2  IS  Yellow

Trait #2 (Eyes):
Genotypes map to phenotypes as follows:
  1,1  IS  Red       2,2  IS  Lobe     1,2  IS  Red
```