DOCUMENT RESUME

ED 271 288 SE 046 634

AUTHOR Collins, Angelo
TITLE Problem-Solving Rules for Genetics.
PUB DATE 86
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Biology; Cognitive Structures; *Computer Assisted Instruction; *Genetics; *Problem Solving; *Programed Tutoring; Science Education; *Science Instruction; Secondary Education; *Secondary School Science; Skill Development
IDENTIFIERS *Knowledge; Science Education Research

ABSTRACT

The categories and applications of strategic knowledge as these relate to problem solving in the area of transmission genetics are examined in this research study. The role of computer simulations in helping students acquire the strategic knowledge necessary to solve realistic transmission genetics problems was emphasized. The Genetics Construction Kit (GCK) was the simulation program used to generate transmission genetics problems in this study. The two sources of strategic knowledge required to solve these problems came from the rational analysis of each class of problems and from the analysis of the performance of experts solving such problems. The rational analysis and the analysis of the performance of experts were used to modify and supplement each other in order to construct a description of desired performance for solving realistic, computer-generated transmission genetics problems. Information presented on these two sources included: (1) a general description, (2) data redescription, (3) solution synthesis, and (4) solution assessment. It appears feasible that a computer tutoring system together with an expert system can be designed and implemented to enable students to solve realistic computer generated transmission genetics problems. (ML)

Problem-Solving Rules for Genetics

Angelo Collins
University of Wisconsin, Madison

Paper presented at the Annual Meeting
of the American Educational Research Association
Symposium on Concept Development and Problem Solving in Genetics
San Francisco, April 18, 1986

Introduction

As early as 1916, John Dewey, in the first issue of Science Education urged science educators to recognize problem solving as a valuable scientific tool and a worthwhile educational goal. Documents such as Science and Mathematics in the Schools: Report of a Convocation (National Academy of Science, 1982) indicate that problem solving is once again in the forefront of the interests of science educators. Concurrent with this interest is the problem solving research of cognitive scientists which provides science educators with insights into the nature of problem solving which hold promise for educational practice. Reif (1983a, 1983b), a cognitive scientist and physicist, while studying problem solving in physics, has proposed that to understand and teach problem solving in any natural science discipline, the following models must be described: a model of desired performance (which includes expert performance, but modified and supplemented to be as clear and explicit as possible), a model of novice performance, a model of learning and a model of teaching. He further states that the performance models must include descriptions of two types of knowledge required for problem solving which he designates as content knowledge and strategic knowledge. He continues by identifying three aspects of content knowledge: the concepts and principles of the discipline, the ancillary knowledge of when and how to use this conceptual knowledge, and the structure of this knowledge. He also identifies three categories of strategic knowledge: data redescription strategies which enable the solver to identify the essentials of a problem and limit the problem space; solution synthesis strategies by which the solver plans and executes means to search the problem space; and solution assessment strategies by which the solver determines that the answer is as complete and accurate as possible. While acknowledging the importance of content

knowledge and its structure, the focus of this research is on the three categories of strategic knowledge.

Much of the recent research in problem solving has been in physics and mathematics. However, transmission genetics is another natural science discipline that is receiving increased attention from science education researchers. Transmission genetics can be considered an important area to study for at least three reasons. First, transmission genetics is important to biologists. It is essential to understanding evolution theory, a basic unifying theme in biology. Recent advances in reproductive technology and in the causes of human genetic disorders are based on an understanding of transmission genetics. And the principles of transmission genetics influence agriculture and horticulture. Secondly, transmission genetics is important to science educators. It is included in the curriculum of almost all high school and undergraduate biology courses (Hurd, Bybee, & Yager, 1980). Moreover, in a survey of high school biology teachers, genetics was ranked among the five most important topics for students to learn (Finley, Stewart, Yarroch, 1982). Lastly, the study of transmission genetics problems has the potential to contribute to the knowledge of problem solving. Transmission genetics problems have both qualitative and quantitative elements and, with the use of computer programs, can be constructed to be similar to the real problems that geneticists address.

Traditionally in genetics, two types of experiences have been offered to students to enable them to learn the strategic knowledge required for problem solving: textbook problems and laboratory problems. Textbook problems are unlike real problems because they are well-structured and require relatively few, recently-taught concepts. In transmission genetics, the form of the textbook problems is also unlike real problems. In typical textbook problems, the solver is presented with a description of traits and

variations of parents and the inheritance pattern controlling the production of offspring. The solution is to predict the distribution of variations among the offspring. A genetics researcher begins with observations about a population of organisms. The researcher selects parents with traits and variations of interest and produces generations of offspring until an inheritance pattern can be determined. Problems of this type are provided to a student in laboratory experiences, but constraints of time and cost usually allow a student to address only one or two problems in a course. In these two approaches, students are taught to solve unrealistic problems with the result that their strategic knowledge has little connection with real genetics problem solving or they have real problems in the laboratory but are unable to solve many problems.

Computer simulations make it possible to offer students experiences that will help them acquire the strategic knowledge necessary to solve realistic transmission genetics problems without the difficulties associated with laboratory experiences. GENETICS CONSTRUCTION KIT (Jungck & Calley, 1984) is a computer program that generates problems that require the continual production of data and interaction between the problem and the solver. Having access to such a realistic problem solving environment is not sufficient for students to become successful problem solvers. Students need help in constructing the strategic knowledge required to solve realistic problems. Such help can be provided to students by tutorial assistance. While a human tutor would be most helpful, it is feasible that a computer program with the strategic knowledge required for genetics problem solving, tutorial expertise, and access to the same computer-generated data as a student, could assume some of the responsibilities of a tutor.

A project to develop such a computer tutor (the Interactive Genetics

Tutor of Project Trochos) is currently underway at the University of Wisconsin (Streibel & Stewart, 1984). To design and construct such an intelligent tutoring system the strategic knowledge required for desired performance in solving realistic computer-generated transmission genetics problems must be described and then presented in a form that can be used for computer programming, in this research as problem solving rules.

METHODS

GENETICS CONSTRUCTION KIT (GCK) was the strategic simulation program used to generate transmission genetics problems in this study. The parameters 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. Problems such as these are typical of problems used in high school and undergraduate biology instruction.

GCK generates populations of imaginary organisms. In this study the organisms had traits with discrete variations similar to bugs. The organisms are diploid with homogametic females and heterogametic males. The simulation begins by presenting to the solver a field collection of organisms with the sex and phenotype of each individual given. The solver then chooses individuals for parents and makes crosses until he or she is able to determine the inheritance pattern, match the genotype to phenotype, and identify a modifier if one is present.

The description of the strategic knowledge required to solve these realistic problems comes from two sources: a rational analysis of each class of problems and the analysis of the performance of experts solving such problems. In the rational analysis, the classes of problems were examined to identify the characteristics of each class which are the essentials required for data redescription. Then the most efficient,

knowledge-producing path for solution synthesis was determined. Lastly, a mathematical test was identified to be used for solution assessment.

For the analysis of expert performance, the problem solving of seven experts was studied. All of the experts have a doctoral degree and experience in both teaching and 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. The experts were asked to think aloud as they solved the problems and make notes as they normally would. Each expert spent an hour with the researcher learning the mechanics of the computer program. Then, in order to eliminate discomfort and/or silent clues possible if the researcher were present, each expert spent four additional hours alone solving problems. However, because each expert worked at their own pace, and because the problem generator is a random problem generator, every problem was not addressed by each expert. The classes of problems attempted by each expert are presented in Table 1.

Three types of data were available for the analysis of the performance of experts: 1) computer printouts of the sequence of crosses made by the expert along with the expert's solution and the correct solution (Appendix A), 2) the transcript of the think aloud protocol, 3) and the notes made by the expert in the process of solving each problem.

The rational analysis and the analysis of the performance of experts were used to modify and supplement each other to construct a description of desired performance for solving realistic, computer-generated transmission genetics problems.

ANALYSIS

Rational Analysis

The purpose of the rational analysis was to identify the strategic

knowledge required to solve realistic problems without referring to the performance of experts. The rational analysis provided the details needed to supplement the description of expert performance so as to create the description of desired performance.

Data Redescription To identify the essential characteristics of a class of problems required for data redescription, the researcher asked questions such as: what makes a particular problem a member of the class of simple dominant problems; what makes the class of codominant problems different from the class of sex linkage problems. As a result, the classes of problems were divided into problems with inheritance patterns -- simple dominant, codominant, and multiple alleles -- that exist independently and mutually exclusive of each other, and modifiers -- sex linkage and autosomal linkage -- which cannot exist independent of an inheritance pattern. For each class of problems an identifying characteristic of the data was also determined. For example, problems with the simple dominant inheritance pattern have two variations for a trait; those with the codominance inheritance pattern have three variations for a trait; and those with multiple alleles have three to six variations for a trait.

Solution Synthesis To identify the most efficient means to reach a solution, the researcher asked questions such as: Given the initial population, what are all the possible crosses that could be made; what inferences could be drawn from each cross; which sequence of crosses produces the most knowledge. From this analysis it was determined that the most knowledge producing first cross is to cross parents with unlike phenotypic variations and that the most efficient second cross is to use offspring of the first cross as parents and construct a second generation by crossing individuals with like variations.

Solution Assessment To be sure the solution was accurate, the

researcher used the mathematical tradition of doing a Chi square test to compare the observed distribution of variations among the offspring with the expected distribution based on transmission genetics principles for all classes of problems except sex linkage, where a standard test cross was employed.

Figure 1 is a flowchart of the rational analysis of the simple dominant class of problems.

Expert Performance

The analysis and reduction of the research data on expert performance had four stages. The first was to match the computer printouts of the solvers' crosses with the transcripts and reduce both to the concepts and principles of transmission genetics. It was also important to maintain the dynamic nature of the solution process. The four steps required for the first stage of data reduction for one person, for the beginning of one simple dominant problem is illustrated in Figure 2. Figure 3 illustrates the second stage of data reduction in which all the data from each cross for all solvers for a class of problems -- in this instance the first cross in a simple dominant problem -- was tabulated. In the third stage of data reduction, the performance of the experts for each category of strategic knowledge for each type of problem was summarized. Table 2 is the summary of data redescription for simple dominant problems; Table 3 is of the solution synthesis strategy, hypothesis testing, used in simple dominant problems; and Table 4 is of the solution assessment strategy, confirmation, used in simple dominant problems. The last stage of the analysis was to describe of the performance of experts solving realistic computer generated transmission genetics problems for each of the categories of strategic knowledge.

Data Redescription Recall that the function of data redescription is

to isolate the essentials of the problem and limit the search space. The experts include in the data redescription statements about the number and name of the traits and variations. Identifying the number of traits for each variation is necessary to determine an inheritance pattern. While describing the data, the experts also note missing class of phenotypes. A missing class of phenotypes by sex among the offspring of a cross is an indicator that the sex linkage modifier might be operating on the population of organisms. A missing class of phenotypes or an unbalanced distribution of individuals by variation is an indicator that the autosomal linkage modifier might be operating on the population.

Data redescription always precedes the formulation of an hypothesis about an inheritance pattern or modifier. Therefore data redescription occurs at the beginning of the problem and in the course of solution synthesis whenever an hypothesis is formulated. Alternate hypotheses are formulated when a cross produces a new variation that alters the essentials of the problem; when an expert solver is unable to determine or confirm an inheritance pattern; and when an expert solver realizes he has made an error in data interpretation. 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.

Solution Synthesis Solution synthesis strategies are those used to plan and execute a search of the problem space and enable the solver to determine 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 continually produced, there is an interaction between the problem data, the

specific hypothesis and the general hypothesis. For example, the initial population data may present an organism with two variations for a trait. The redescription that identifies this characteristic of the problem would lead the expert to formulate an initial, tentative general hypothesis of simple dominance. The expert might then choose to mate 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. If this cross is then performed, and the results agree with the prediction, the data will support the specific hypothesis and the specific hypothesis will confirm the general hypothesis. This interaction between data, specific hypotheses, and general hypotheses continues throughout the synthesis of the problem solution. In addition, 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 has been determined. In simple dominance and codominance this definitive cross is the F(2) cross; in multiple alleles the class of crosses used to determine 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. The definitive cross in all classes of problems except sex linkage requires the identification of heterozygous individuals.

Once the inheritance pattern has been determined, the expert continues to do crosses to determine 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 consider both sex linkage and autosomal linkage modifiers. In determining modifiers, the interaction between the

problem data, the specific hypotheses and general hypotheses continues. There is also a definitive cross to determine each modifier. In sex linkage the cross is between a dominant male and a recessive female, producing recessive male and dominant female offspring. In autosomal linkage the definitive cross is between a parent that is heterozygous for two 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.

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 expert is assuring himself 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 determined the inheritance pattern or modifier. Although the Chi square test is the statistical test to determine if the actual 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 test. Experts also increase their confidence in the inheritance pattern and modifier hypotheses by doing additional crosses that are explained or predicted by the hypotheses. Whenever possible, experts use more than one method of confirmation.

Table 5 is a summary of the details of each strategy for each class of problems.

A general description of desired performance was constructed from a description of desired performance for each class of problems. The description of desired performance includes the richness of the expert

performance and the detail of the rational analysis. Figure 4 is a flowchart of the general description of desired performance.

IMPLICATIONS AND IMPLEMENTATION

Two implications of this research are important to problem solving theory. First is that, although the details within each category are different, the categories of strategic knowledge identified and used by Reif to describe problem solving in physics, can be used to describe problem solving in transmission genetics. This suggests that describing the details of these categories of strategic knowledge could be a fruitful model for problem solving studies in other natural science disciplines. The second implication is that the two characteristics of the strategic knowledge of an expert solving textbook physics problems identified by Larkin, McDermott, Simon and Simon (1980) are seen in the problem solving performance of experts solving realistic transmission genetics problems. These characteristics are the use of redescription at the beginning of the problem to identify the essentials of the problem and the use of forward-working, knowledge producing strategies in the synthesis of the solution. However, the characteristics are the same, but the details are different. Expert physicists solving textbook problems redescribe once and use setting subgoals as the solution strategy; expert geneticists redescribe with each hypothesis and use hypothesis testing as the knowledge producing strategy.

Intelligent Tutoring Program The innovative implementation of the description of desired performance for solving realistic computer generated transmission genetics problems is to construct an intelligent tutoring system to help students learn to solve such problems. The design and implementation of such a tutoring system is the goal of the Interactive Genetics Tutor (IGT) research project of Project Trochos currently being conducted at the University of Wisconsin, Madison. The IGT will eventually

have three components - the genetics problem generator, GENETICS CONSTRUCTION KIT, which is in operation; an intelligent tutoring system which will consist of a model of desired performance, a model of student performance, and a tutoring program; and a videodisc component of visuals to accompany the tutoring program.

The description of desired performance is being used to design an expert system component as the model of desired performance of the intelligent tutoring system of the IGT. An expert system is a program that represents and applies factual knowledge of a specific area of expertise for domain-specific problem solving. The format that has been chosen for the prototype of the expert system is a combination of frame and rule-based knowledge representation. In writing the program, the details and the implications of the description of desired performance are analyzed and clearly stated; this knowledge is then converted to frames or rules; and lastly, the frames and rules are converted to program code by a computer programmer. Various categories of rules have been determined - for example hypothesis generating rules and hypothesis testing rules. The most general category of knowledge is in the form of an agenda, which directs the program to access a category of specific rules or frames. The agenda is presented in Table 6 in the form that immediately precedes the rule form.

Item 1 in the agenda corresponds with the data redescription strategy, Items 2, 3, and 4 correspond with the hypothesis testing strategy, and Item 5 corresponds with the confirmation strategy. The agenda parallels the model of desired performance.

The frames or rules accessed by the computer from each item of the agenda are illustrated with examples from the simple dominant inheritance pattern and presented in an understandable form close to the frame or rule form.

Data Redescription - Item 1 The program uses frames for the knowledge representation of the number and name of each trait and variation. Table 7 is a sample trait frame. Enough frames have been identified and designed to represent the knowledge needed to formulate inheritance pattern hypotheses from problem data. Currently the expert system completes the frames by asking questions of a person that has data available by running GENETICS CONSTRUCTION KIT. Eventually the problem generator will interface with the expert system.

Solution Synthesis - Agenda Items 2, 3, and 4 Once the data has been redescribed, it is possible to formulate and test an hypothesis. Hypothesis generating rules, Item 2 on the agenda, are presented in Table 8.

The next item on the agenda, Item 3 - Make a Cross to Test the Hypothesis - uses frames to represent the variations of the parents and the offspring from a cross. A sample cross frame and offspring frame are presented in Table 9.

Agenda Item 4 - If Possible, Explain the Cross - consists of two sets of rules. The first set of rules matches the phenotypes of the parents and offspring from the frames with the genotypes of the specific crosses. Table 10 illustrates this set of matching rules for the simple dominance inheritance pattern. The pattern matching rules are then accessed by a set of hypothesis testing rules. Table 11 illustrates the hypothesis testing rules for the simple dominance inheritance pattern.

Agenda Items 2, 3, and 4 for the simple dominant, the codominant, and the multiple alleles inheritance patterns are in operation in a prototype expert system program. Item 5 on the Agenda will implement the confirmation strategies. Although not yet written, Item 5 will probably take the form of rules and will require the use of more than one form of confirmation. Item 6, repeating the Agenda for each trait, and Item 7, repeating the Agenda for

modifier hypotheses are also not yet written.

The agenda format parallels the description of desired performance. The specific rules and frames of the expert system employ the details of the description of desired performance. It seems feasible that an Interactive Genetics Tutor with an expert system based on the description of desired performance can be designed and implemented to enable students learn to solve realistic computer generated transmission genetics problems.

REFERENCES

Dewey, J. (1916). Method in science teaching. Science Education, 1(1), 1-10.

Finley, F., Stewart, J., & Yarroch, W. (1982). Teachers' perceptions of important and difficult science content. Science Education, 66(4), 531-538.

Hurd, P., Bybee, R., Kahle, J., & Yager, R. (1980). Biology education in secondary schools of the United States. American Biology Teacher, 42(7), 388-410.

Jungck, J., & Calley, J. (1984). GENETICS CONSTRUCTION KIT

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

National Academy of Science & National Academy of Engineering. (1982). Science and Mathemetics 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.

Streibel, M., & Stewart, J.H. (1984). Interactive genetics tutorial project. In Project Trochos, (pp. 53-56). Madison: University of Wisconsin, IBM Advanced Education Project.

APPENDIX A

COMPUTER PRINTOUT OF PROBLEM AND SOLUTION

```
Operations

**********New Problem**********
  PROBLEM TYPE 1
**********************************
  Contents of Vial #1
 1 F  Straw     Delta
 8 M  Straw     Delta
 3 M  Straw     Ocelliless
12 F  Yellow    Delta
 2 M  Yellow    Delta
Listing Vial Number 1
Entering CROSS...
Vial #1  Phenotype #1   Individual #1
Vial #1  Phenotype #3   Individual #1
  Contents of Vial #2
 9 F  Straw     Delta
12 M  Straw     Delta
 2 F  Straw     Ocelliless
 7 M  Straw     Ocelliless
Entering CROSS...
Vial #2  Phenotype #3   Individual #1
Vial #2  Phenotype #2   Individual #2
  Contents of Vial #3
 6 F  Straw     Ocelliless
 9 M  Straw     Ocelliless
 8 F  Straw     Delta
10 M  Straw     Delta
Entering CROSS...
Vial #1  Phenotype #4   Individual #4
Vial #1  Phenotype #5   Individual #1
  Contents of Vial #4
14 F  Yellow    Delta
 7 M  Yellow    Delta
 4 M  Straw     Delta
Entering CROSS...
Vial #4  Phenotype #2   Individual #2
Vial #4  Phenotype #3   Individual #3
Vial #4  Phenotype #1   Individual #3
Vial #4  Phenotype #3   Individual #3
  Contents of Vial #5
 8 F  Straw     Delta
13 M  Straw     Delta
 7 F  Yellow    Delta
12 M  Yellow    Delta
```

Expert's solution:

TWO ALLELIC PAIRS -- Y IS DOMINANT AND O IS DOMINANT TO THEIR PARTNERS.
BOTH ARE AUTOSOMAL AND THEY APPEAR TO BE UNLINKED.

Program solution

TRAIT NUMBER 1 (Body):

This trait is sex-linked.

There are 2 alleles.

Genotypes map to phenotypes as follows:

1,1 IS Yellow
2,2 IS Straw
1,2 IS Yellow

TRAIT NUMBER 2 (Ocelli):

There are 2 alleles.

Genotypes map to phenotypes as follows:

1,1 IS Delta
2,2 IS Ocelliless
1,2 IS Delta

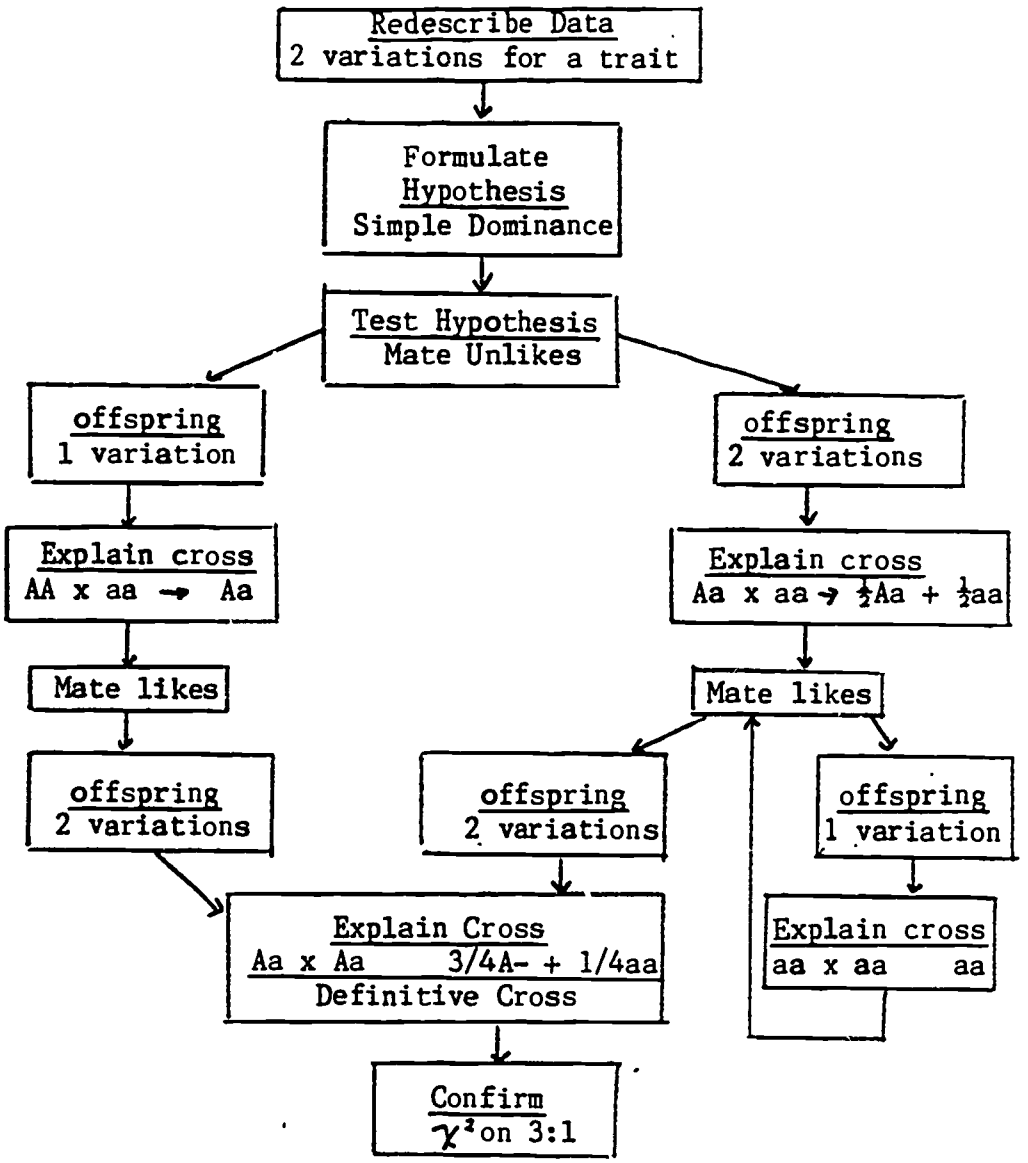


Figure 1 Rational Analysis of Simple Dominance.

Figure 2 Data Tabulation

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

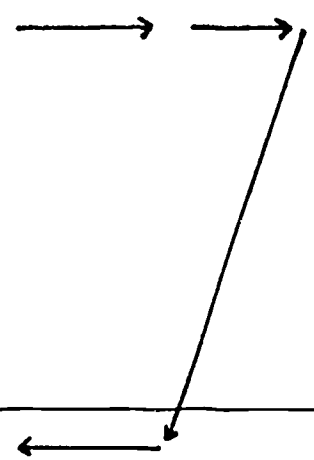
CROSS	Step 1 Read the transcript and mark the frames	Step 2 Place the phrases of the transcript in groups depending on whether they refer to problem data (PD), specific hypothesis (SH), or general hypothesis (GH)	Step 3 Reduce the phrases of the transcript to transmission genetics concepts and add notes	Step 4 Draw arrows to represent the sequence
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 dihybrid	PD: classes traits variations SH: aa x aa → aa double note lfp = rec GH: Simple Dom	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	

Figure 3 Data Reduction.

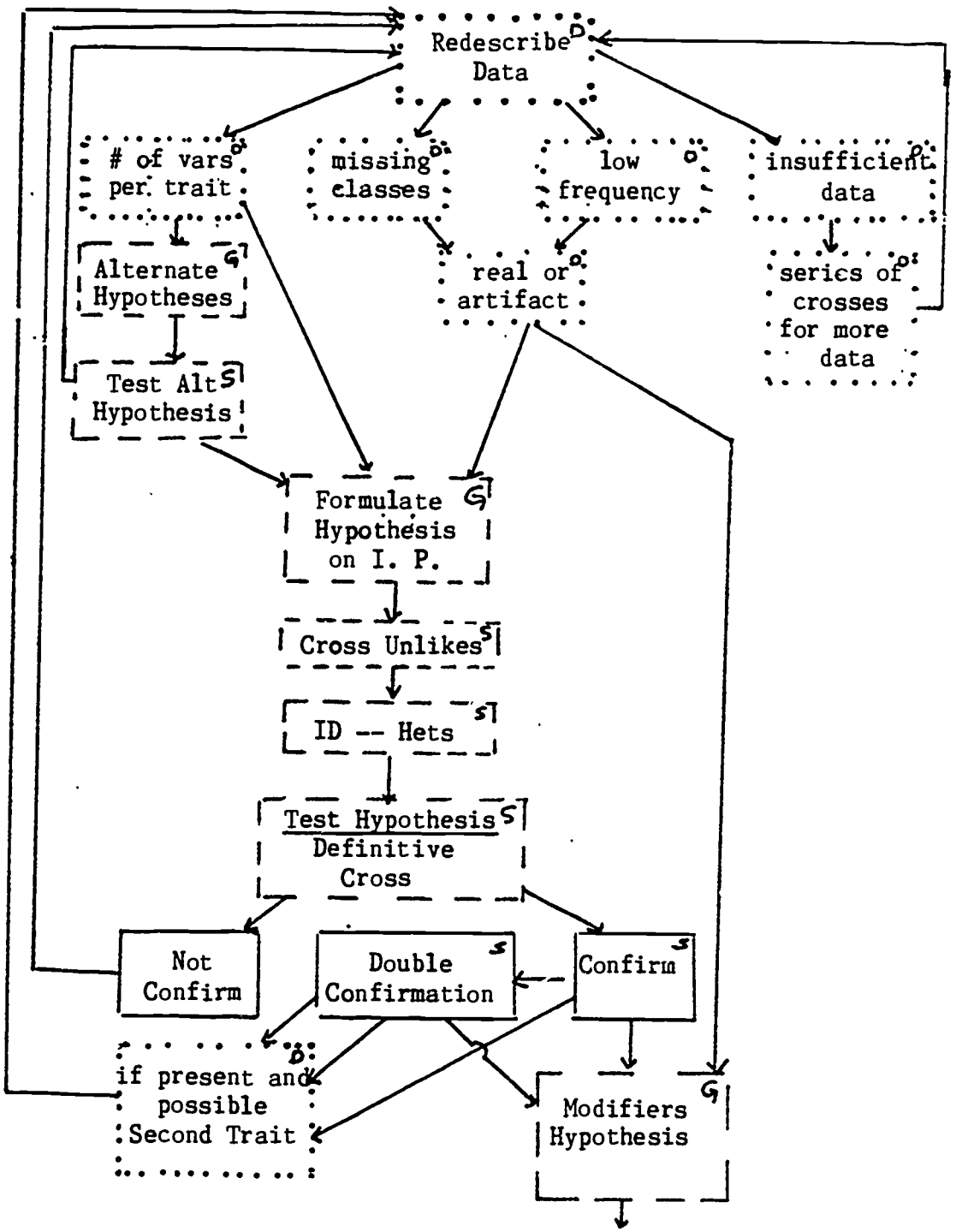


Figure 4 Desired Performance.

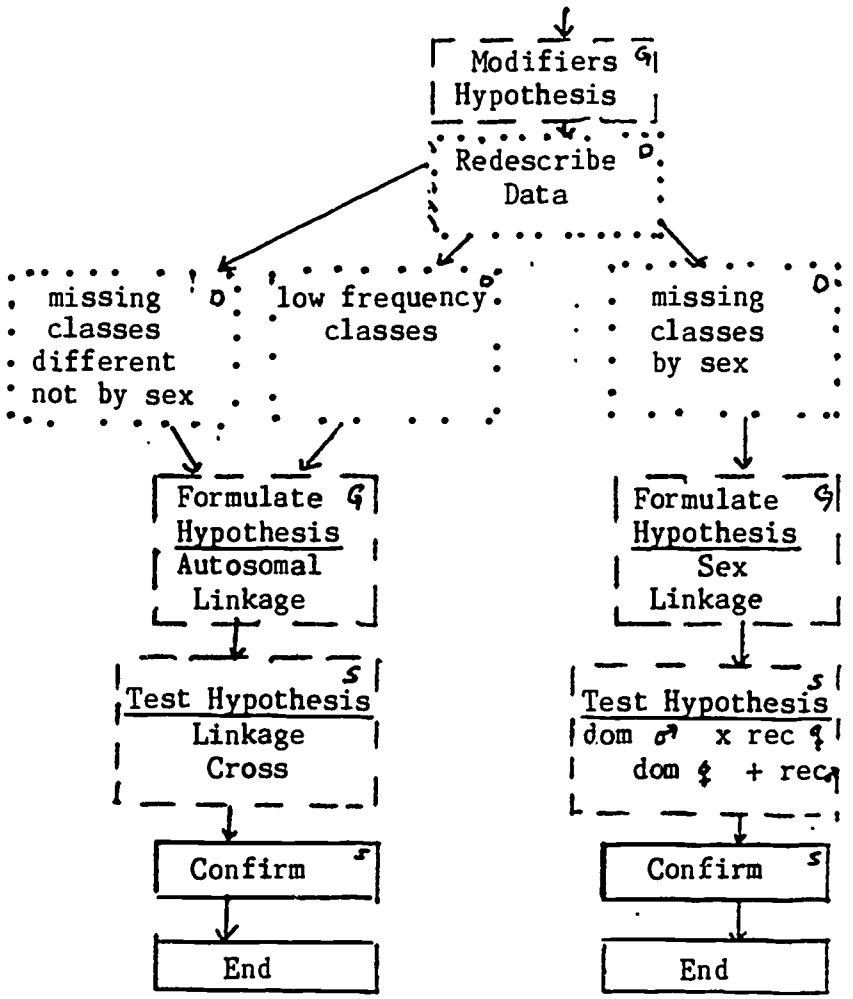


Figure 4 (continued).

Table 1 Classes of Problems Attempted by Each Expert

EXPERT PROBLEM	1	2	3	4	5	6	7	Total
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

Table 2 Simple Dominance — Data Redescription.

1. Details of Initial Redescription
   -- 14 problems have some type of 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 phenotype; of these, 1 also notes most frequent phenotype

2. Additional Occasions of Redscription
   -- 2 problems are redescribed when the attention of the solver is focusing on the other trait
   -- 6 problems are redescribed whenever an alternate hypothesis is considered
   -- 4 problems are redescribed at the end of the problem

3. Symbols
   -- In 12 problems the solver uses the first letter of the trait or variation instead of the full name

Table 3 . Simple dominance -- Hypothesis Testing.

1. Origin of the General Hypothesis
   - -- 6 problems have the simple dominant inheritance pattern stated from the initial population
   - -- 6 problems have this hypothesis stated after 1 or 2 crosses
   - -- 2 problems have this hypothesis stated after beginning a series of 4 or 5 possible crosses of phenotype classes described in the initial population

2. Interaction of General and Specific Hypotheses
   - -- In 73% of the frames, at least one general hypothesis is considered; in 50% of these, more than one is considered
   - -- In 84% of the frames, at least one specific hypothesis is considered; in 35% of these, more than one is considered
   - -- In 64% of the frames both general and specific hypotheses are considered
   - -- the D-S-G-S-D-S-G pattern is used 34 times in 12 of 14 problems, at least once by each expert

3. 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 heterozygotes are created
     - -- 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

4. Initial Cross
   - -- In 5 problems the initial cross is unlikes
   - -- In 8 problems the initial cross is likes
     - -- In 3 of the 8 problems that begin with likes, the data is produced to confirm the hypothesis that the least frequent phenotype is recessive
     - -- In 3 of 8 problems that begin with likes the purpose is to create a missing class of phenotypes
       - -- In 3 of 6 problems that begin with likes for a reason, the next cross is unlikes
   - -- The 3 unsuccessful problems begin with like crosses

Table 3 (continued).

5. Alternate Hypotheses
   -- In 11 problems autosomal linkage as a modifier is considered and rejected.
      -- 11 times after the inheritance pattern is confirmed
      -- 8 times in the last frame
      -- 7 times by the linkage cross
      -- 4 times by a dihybrid 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 linked 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

6. Separate Traits
   -- In 9 problems the two traits are discussed separately in separate frames

7. Simple First
   -- not applicable

TABLE 4 Simple Dominance -- Confirmation.

1. Mathematical
   -- In 8 of the 9 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 3 problems Chi square is mentioned but not used because the solver says it is an appropriate test but he/she is too lazy.
      -- In 2 problems an elaborate method of calculating fractions is used
      -- In 5 problems the solver says the ratio "looks ok"

2. Strategic
   -- In 6 problems both an F(2) and a linkage cross and 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 5. Summary of Details of Strategies.

IP	REDESCRIPTION	DEFINITIVE CROSS	CONFIRMATION
SD	2 vars/trait	F(2)	χ^2, linkage
COD	3 vars/trait	F(2)	χ^2, additional X
MA	3-6 vars/trait	series + F(2)	all geno to pheno matches
SL	missing class of one sex	dominant ♂ X recessive ♀	
AL	missing or low frequency class	linkage	different individuals

Table 6 The IGT Agenda.

1.	Redescribe the data.
2.	Entertain a hypothesis about the inheritance pattern.
3.	Test the hypothesis by making a cross.
4.	If possible, explain the cross using the information of the six possible specific hypotheses.
5.	Confirm a solution.
6.	Repeat items 1-5 for additional traits.
7.	Repeat items 1-5 for the modifiers of sex linkage and autosomal linkage for each trait.

Table 7 Trait Frame.

Empty Frame	Completed Frame
Trait #_____	Trait # 1
Trait_____	Trait eyes
Variations ___	Variations 2
Var 1 _____	Var 1 red
Var 2 _____	Var 2 apricot

Table 8 Correspondence of Data Redescription with Inheritance Pattern Hypotheses

Hypothesis-Generating Rules:

-- if the trait has two variations, then test the simple dominant inheritance pattern

-- if the trait has three variations, then test the codominance inheritance pattern

-- if the trait has three variations and if codminance has not been confirmed, then test the multiple alleles inheritance pattern

-- if the trait has four, five, or six variations, then test the multiple alleles inheritance pattern.

Table 9 Cross and Offspring Frames.

CROSS FRAME	
Empty Frame	Completed Frame
Parent # _____	Parent # 1
Var # _____	Var # 1
Parent # _____	Parent # 2
Var # _____	Var # 1
Var # _____	Var #____
Offspring Frame	Offspring Frame

OFFSPRING FRAME	
Empty Frame	Completed Frame
Offspring #_____	Offspring # 1
Var # _____	Var # 1
Individuals ____	Individuals 57

Table 10 Rules Matching Parent and Offspring Phenotypes with Genotype Patterns.

If the parent phenotypes are		and the offspring phenotypes are	then the genotype pattern is
like X like	->	1 type, like	AA x AA -> AA
like X like	->	1 type, like	AA x A- -> A-
like X like	->	1 type, like	aa x aa -> aa
unlike x unlike	->	1 type, like	AA x aa -> Aa
unlike X unlike	->	2 types, unlikes	Aa x aa -> Aa + aa
like X like	->	2 types, unlikes	Aa x Aa -> A- + aa (definitive cross)

Table 11 Hypothesis-Testing Rules.

-- if there is one and only one pattern explanation for a cross, then determine if the pattern is the definitive cross

  -- if the cross was the definitive cross, then go on to Agenda Item 5

  -- if the cross was not the definitive cross, then identify a heterozygote.

    -- if the parents were unlikes and there is only one type of offspring, then the offspring are heterozygotes so use the offspring to repeat Agenda Item 3

    -- if the parents were unlikes and there are two types of offspring, then repeat Agenda Item 3 with offspring of this cross with like variations