

Genetics Inquiry: Strategies and Knowledge Geneticists Use in Solving Transmission Genetics Problems

NORMAN THOMSON

Science Education Department, University of Georgia, Athens, GA 30602, USA

JAMES STEWART

Department of Curriculum and Instruction, 225 N. Mills Street, University of Wisconsin, Madison, WI 53706, USA

Received 2 April 2001; revised 17 December 2001; accepted 9 January 2002

ABSTRACT: Scientists are increasingly challenged to solve problems that require conceptualizing and understanding dynamic complex natural systems. Computer simulations have become an integral tool in helping solve complex problems because they can be effectively used to model dynamic systems. Today's reform standards advocate that science teachers and students be involved in scientific inquiry that is consistent with the practice of science. The goal of this research is to determine methods of inquiry geneticists' use in solving dynamic complex computer-generated transmission genetics problems, specifically, their strategies and conceptual knowledge. Six geneticists representing three areas of genetics (transmission, molecular, and population) solved four problem types. Using their solution protocols, and through interviews, a hierarchical framework and pathway for solving the problems were developed. For science education researchers, teachers, and students, it is hoped the geneticists' insights and the framework developed can provide a guide for inquiry-based problem solving that extends beyond genetics. © 2003 Wiley Periodicals, Inc. *Sci Ed* 87:161–180, 2003; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/sce.10065

INTRODUCTION

In the National Science Education Standards, two themes are emphasized as important goals for developing science curricula (National Research Council [NRC], 1996, 2000). The first theme is that science should be understood and experienced as a method of inquiry. Accordingly, students should (a) “have the abilities to do” and (b) “understand the nature of scientific inquiry” (p. 23). The second theme is that students need learning experiences that include the “unifying concepts and processes in the sciences: (a) systems, order, and organization, (b) evidence, models and explanation, (c) change, constancy, and measurement,

Correspondence to: Norman Thomson; e-mail: nthomson@coe.uga.edu

(d) evolution and equilibrium, and (e) form and function” (p. 104). These themes do not differ greatly from those proposed as Benchmarks for Scientific Literacy in Project 2061 (American Association for the Advancement of Science, 1993).

The implication of the Standards and Benchmarks for science teachers is that they are challenged to design and/or implement inquiry-based instruction in the hope that their students will become sufficiently “scientifically literate” to make informed decisions about problems that integrate scientific, technological, and societal issues (NRC, 1996, p. 23). Today, informed decision-making in science requires understanding the nature and substance of complex problems and systems. Most decision-making will involve being able to make choices between alternative solutions.

Scientific interest and inquiry into complex dynamic systems and problems is not a recent endeavor. For well over a century, evolutionary biologists (Darwin, 1866; Futuyma, 1998; Mayr, 1982), and more recently, meteorologists (Peixoto & Oort, 1992), ecologists (Allen & Starr, 1982; Odum, 1971), and chaos and quark theorists (Gell-Mann, 1994; Gleick, 1987) have created conceptual models and analytical tools to explore, describe, and understand natural complex dynamic systems. Through a historical analysis, Darden (1991) has constructed an account of the strategies and knowledge geneticists used from 1900 to 1926 to solve complex problems that led to a theory of the gene. Her research resulted in the development of a generalized framework that she hopes other researchers can utilize to investigate strategies and methods of inquiry scientists’ implement during theory change in science. The historical development of scientific inquiry is one method teachers may use to construct thoughtful lessons that can provide insight as to how the nature of inquiry has addressed past complex problems in science. The limitation of historical studies, however, is that it informs us as to what scientists did in the past, rather than providing insight into what they currently do.

The translation of ongoing scientific inquiry by teachers into classroom practice, however, is a challenging endeavor. Some science educators’ (Bird & Weller, 1997; Flick, 1997; Keys & Kennedy, 1999; Tobin, Kahle, & Fraser, 1990) research attests that it is a difficult task for teachers to create learning environments that have scientific inquiry as a focus. Scientific inquiry requires the “conceptual identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations” if it is to be effectively used as a method for solving problems (NRC, 1996, p. 23). Tobin, Kahle, and Fraser (1990) determined that several factors influence whether or not a biology teacher uses methods of scientific inquiry for teaching. In Australia, the mandated curriculum and assessment instruments emphasize and promote lower level cognitive learning. Textbooks are written to match the mandated standards. Classroom management styles influence the way a teacher interacts with students, determine how time is utilized in teacher–student questioning and learning, and affects the extent students could effectively collaborate in group discussions. Teacher familiarity and comfort with pedagogical content knowledge is also an important dimension: Lack of knowledge decreases questions and answers requiring higher levels of thinking. Finally, syllabus time required to “cover” the breadth of content diminishes time for in-depth learning.

While working collaboratively, Keys, a science education researcher, and Kennedy, an elementary teacher, reflected that each had personal constructions of “inquiry” (Keys & Kennedy, 1999). Kennedy found it difficult to translate NRC Standard’s ideas of reform into usable classroom practice. Her approach to inquiry was “serendipitous,” and she found a fundamental conflict between trying to model national standards versus teaching mandated state objectives. Inquiry was viewed as a time consuming endeavor and incompatible with the amount of content she was required to present.

Computer simulations now offer a new dimension through which teachers and students can collaboratively experience inquiry-based problem solving, construct knowledge, and

engage in modeling as a part of science learning (Penner, 2001). Our research goal is twofold: (a) to explore what strategies and conceptual knowledge geneticists use in attempting to solve combinations of simple Mendelian and complex non-Mendelian transmission genetics problems; and (b) to investigate problem exploration that leads to hypothesis generation, testing, assessment of evidence, and hypothesis revision as the geneticists attempted to solve the problems. The underlying reason for conducting this research was to determine whether geneticists, representing different research areas (molecular, transmission, and population genetics), utilize similar, contrasting, or idiosyncratic strategies and conceptual knowledge in solving transmission genetic problems. The issue was that if methods for solving complex problems tended to be idiosyncratic and ungeneralizable, it would constrain teachers' implementation of inquiry-based science lessons. On the other hand, if complex problem solving can be performed using a framework, it makes it more likely that a teacher can provide students, with effective instructional guidance, more meaningful opportunities to experience and understand the nature of scientific inquiry. In our study, data sources were six geneticists solving problems using dynamic computer simulations generated in Genetics Construction Kit (GCK), see BioQUEST (1998–99).

One might query how the work of geneticists can assist us in introducing transmission genetics problem solving to teachers and the novice. We have first found that, regardless of problem type or complexity, the geneticists solved all parts of the problems using the same pathway of exploration and hypothesis construction. And, secondly, the geneticists emphasized how important qualitative data is in making inferences and searching for answers to the problems. Their methods and our framework for problem solving can provide teachers opportunities to design inquiry-based science instruction for students. GCK has attributes that allow a learner to start investigating transmission genetics problem solving at a basic level (e.g. middle school), and the geneticists' general strategies used, were independent of problem complexity. For example, whether a phenotype appears, disappears or reappears, or doesn't change as a cross outcome allows one to make decisions including dominance, recessiveness, or interaction allelic relationships. GCK has option levels for offspring data display, that is, as numbers, bar graphs, or use of Chi square statistics for making estimates of ratios.

We consider our research important for science educators because (a) it contributes to an understanding of the nature of scientific inquiry during problem solving in the context of transmission genetics, (b) it demonstrates the interconnection of strategy and knowledge geneticists' use while solving problems in transmission genetics that, (c) has led to a framework for problem exploration, hypothesis generation, testing, and revision. In addition, for life science/biology teachers we (d) offer a new perspective on problem solving in transmission genetics using computer simulations that are available for their own use, and from which, they can design instruction for incorporation into classroom practice (consistent with the NRC themes and students' needs for inquiry-based learning). Finally, (e) transmission genetics is the (global) starting point for learning inheritance and serves as the foundation for understanding biological diversity and evolution (Dobzhansky, 1973; Mayr, 1982; National Academy of Sciences, 1998). Our research offers a new way students can be introduced to learning problem solving in transmission genetics.

METHODS

Selection of Geneticists

Six geneticists (identified as G1–G6) representing three traditional areas of genetics (transmission, molecular, and population) participated in the study. They described their interests as

- G1 addresses questions at the interface of ecology, evolution, systematics, and biogeography with special interest in regional plant adaptive radiation, selection, and speciation with respect to morphological and physiological traits.
- G2 is interested in developing mathematical models and explanations for evolutionary population biology and has written a “genetics notes” book that is an international text translated into several languages.
- G3 is a maize geneticist with special interest in transposon-mediated chromosome rearrangements and gene duplications in the formation of the maize R-r complex.
- G4 is a potato geneticist who has been instrumental in the development of new commercial pathogen resistant varieties and has special interests in maintaining international potato seed germ plasm.
- G5 is a research specialist in *Drosophila* genetics and is interested in the development, breeding, maintenance, and genetic mapping of mutant lines.
- G6 is a microbial (viral and prokaryotic) molecular geneticist with special interest in bacteriophage T4 and has contributed as an author on microbial genetics for several text and reference books.

Generation and Selection of Genetics Problems

An infinite number of genetics problems for a variety of selected genetic parameters can be generated using GCK. An early version of GCK beta 1.1B3 (BioQUEST, 1998–99) was used in this study. The GCK version included variables for traits that “affect viability, sterility, or recombination frequencies as well as single traits controlled by multiple loci (genetic interactions) and multiple traits controlled by a single locus (pleiotropy), and traits which are differentially expressed in different sexes” (see BioQuest, 1998–99, p. 39).

The problems used in this study were, in a general sense, of two kinds: nonepistatic with linkage and epistatic. Other genetic relationships (e.g. dominance, partial dominance) were embedded differentially within particular problems resulting in four problem types (Table 1). A detailed description of one Type 3 problem is provided in Table 2. To reduce potential geneticist bias or familiarity with a particular organism, in all problems, color was the trait. All colors (15 were possible) were randomly programmed as phenotype variations with neither attention to a particular species, structure, or function, nor regard to colors that might be expected for a typical spectral transition suggesting a natural form of partial dominance. The problems were also designed to randomly disrupt a geneticist’s predicted outcomes during problem solving, including numbers of individuals collected in the initial sample and inconsistent numbers of progeny (atypical of living organisms), that could result in anomalous or stochastic cross outcomes suggesting differential viability or mortality.

TABLE 1
The Four Problem Types Selected for this Research Had the Following Parameters

Type 1	Single trait, dominance, codominance, sex linkage; 5 phenotypes (5 for females and 3 for males), 1 locus, and 3 alleles.
Type 2	Single trait, codominance, sex linkage; 3 phenotypes (3 for females and 2 for males), 1 locus, and 2 alleles.
Type 3	Single trait and epistasis; 3 phenotypes, 2 loci, and 4 alleles, (2 alleles/gene); identified by a unique F_2 ratio of 12:3:1.
Type 4	Single trait and epistasis; 4 phenotypes, 2 loci, and 4 alleles (2 alleles/gene); identified by a unique F_2 ratio of 9:3:3:1.

TABLE 2
A Summary Overview of a Type 3 Problem

A Type 3 problem had a single trait with epistasis, and had 3 phenotypes, 2 loci, and 4 alleles (2 alleles/gene) and is identified by a unique F_2 ratio of 12:3:1.

Females and males can have one of 3 distinct phenotypes (Red, Jade, Nile) and one of 9 distinct genotypes: Red = A-B- (4 genotypes) or A-bb (2 genotypes), Jade = aaB- (2 genotypes), and Nile = aabb (1 genotype). Within the problem there are 45 distinct possible crosses of which 24 can be crosses of like phenotypes and 21 of unlike phenotypes. Twenty-eight crosses result in 1 phenotype, 11 crosses result in 2 phenotypes, and 6 crosses result in 3 phenotypes.

The epistasis problem is known by its characteristic F_2 12:3:1 phenotype resulting from an F_1 cross of double heterozygotes. The problem begins with a field collection of 29 individuals and 2 phenotypes, Jade (9 females, 8 males) and Red (8 females, 4 males), a 6:4 phenotypic ratio.

GCK problems permit solvers to begin with the same field collection of organisms. GCK allows a researcher or teacher to choose the genetic variables to be included in a problem type ranging from simple (1 gene, 2 alleles) to extremely complex problems. For a complex problem, GCK randomly selects the genetic variables to be included in the generated problem and they remain unknown to the solver. Complex problems have an infinite number of outcomes (for example, randomized numbers of loci, alleles, or linked genes with complex relationships including epistasis and penetrance). The problem's solution is based on the solvers' interpretations of qualitative, correlative, and statistical evidence.

GCK includes "tools" that allow a solver to perform crosses to explore, generate, construct, test, and interpret qualitative and quantitative data. A field collection of organisms is generated and all data (field and cross generated) is evaluated by comparing numbers of individuals by trait, variation, and/or sex using numerical ratios, bar graph representations, or Chi-square statistics. All generated data is dependent upon a solver's selection of individual phenotypes to be crossed and what genotypes can be inferred through progeny phenotypes.

Data Collection

Preceding data collection, each geneticist was introduced to GCK, but they were not told the nature of the problems they were going to encounter. Each geneticist was requested to solve a minimum of four problems. However, the researcher ensured that each geneticist had one problem from each type. Four sources of data were collected and used in the analysis: (a) printouts of the geneticist's problem solving data including the field collection, parental phenotypes, individual crosses, and offspring phenotypes (numbers and sex) as recorded by the computer, (b) transcripts of the geneticist's think aloud process recorded for each problem, (c) the geneticists' informal notes entered in personal journals, and (d) information obtained from each geneticist in a post problem-solving interview. In the postinterview, they were asked to discuss strategies, if any, and knowledge they used in solving the problems.

There were three reasons why a post problem-solving interview was considered important: (a) it was anticipated that the geneticists' general strategies plus statements made during solving individual problems, would be useful in the construction of a framework for data analysis, (b) it was thought important to establish if there was consistency between what the geneticists' considered to be their general strategies, and whether and how they

actually used them during problem solving, and (c) it was important to know whether the geneticists felt that GCK presented a “realistic” simulation of their personal field or laboratory research (fact) or if GCK problems are unique to a computer environment (artifact) and thus, “Do the two environments (real and simulation) require use of similar or different problem solving strategies?”

Data Analysis

The data analysis is based on a definition of problem solving as a method of inquiry by which one constructs an explanation linking empirical observations to theory. Validity and reliability of explanatory links are supported through hypothesis testing and using qualitative, statistical, or corollary evidence. A strategy is defined as a plan or method that has a purpose or rationale for its use in problem solving. Knowledge refers to the genetics concepts the geneticists used to solve the problems.

For the data analysis, each cross was reviewed and categorized as to (a) whether it was between individuals of like or unlike phenotypes, (b) the phenotype and sex of the parental individuals, (c) the “genetic history” of the parental types, and (d) the phenotypes, sex, and numbers of progeny. Next, a comparison of cross data was made using a geneticist’s transcripts, journal notes, and interviews to analyze each geneticist’s knowledge, strategies, and pathways used. Specifically, sequential comparisons were made from the geneticists’ initial interpretations of a field collection through the point of a problem’s successful solution or unsuccessful termination. It was important to determine (a) the purpose for each selected cross, (b) the information obtained as a consequence of making the cross, and (c) what decisions were made for generation of data consistent and inconsistent with any hypotheses under consideration.

An analysis of the problem solving pathways, cross categories, geneticists’ transcripts, the post problem-solving interview, and the strategies and knowledge used, led to the design of a hierarchical framework for inquiry in which problem solving is the goal (Figure 1) and a pathway specific for solving transmission genetics problems (Figure 2). Darden’s five problem-solving stages were found to be useful analysis categories (see Figure 1). However, it became apparent that her general strategy types and subtypes used in theory change differed somewhat with respect to hypothesis construction, assessment, and revision the geneticists used in this study.

RESULTS

Development of a Framework and Pathway for Solving Transmission Genetics Problems

First, four tiers of problem-solving strategies were developed through the data analysis and were continually refined. The first tier of strategies used are (a) identifying a problem, (b) producing ideas, (c) data exploration, (d) hypothesis construction, (e) data analysis, (f) hypothesis assessment, and (g) hypothesis revision (1.1–1.7, see Figure 1). As each tier represents a higher level of resolution, each level becomes more closely tied to discipline-specific knowledge. For example, in identifying a problem, General Strategies (2.1–2.33, see Figure 1) includes “identify an unexplained state,” that is, a sample population of individuals differing in phenotype. And, corresponding Exemplars (3.1–3.19, see Figure 2) could include the number of traits by sex in the sample. The fourth tier, Examples, is the localized problem specific instance of knowledge constructed or used by a geneticist beginning with the field population and, extending through each sequential cross made,

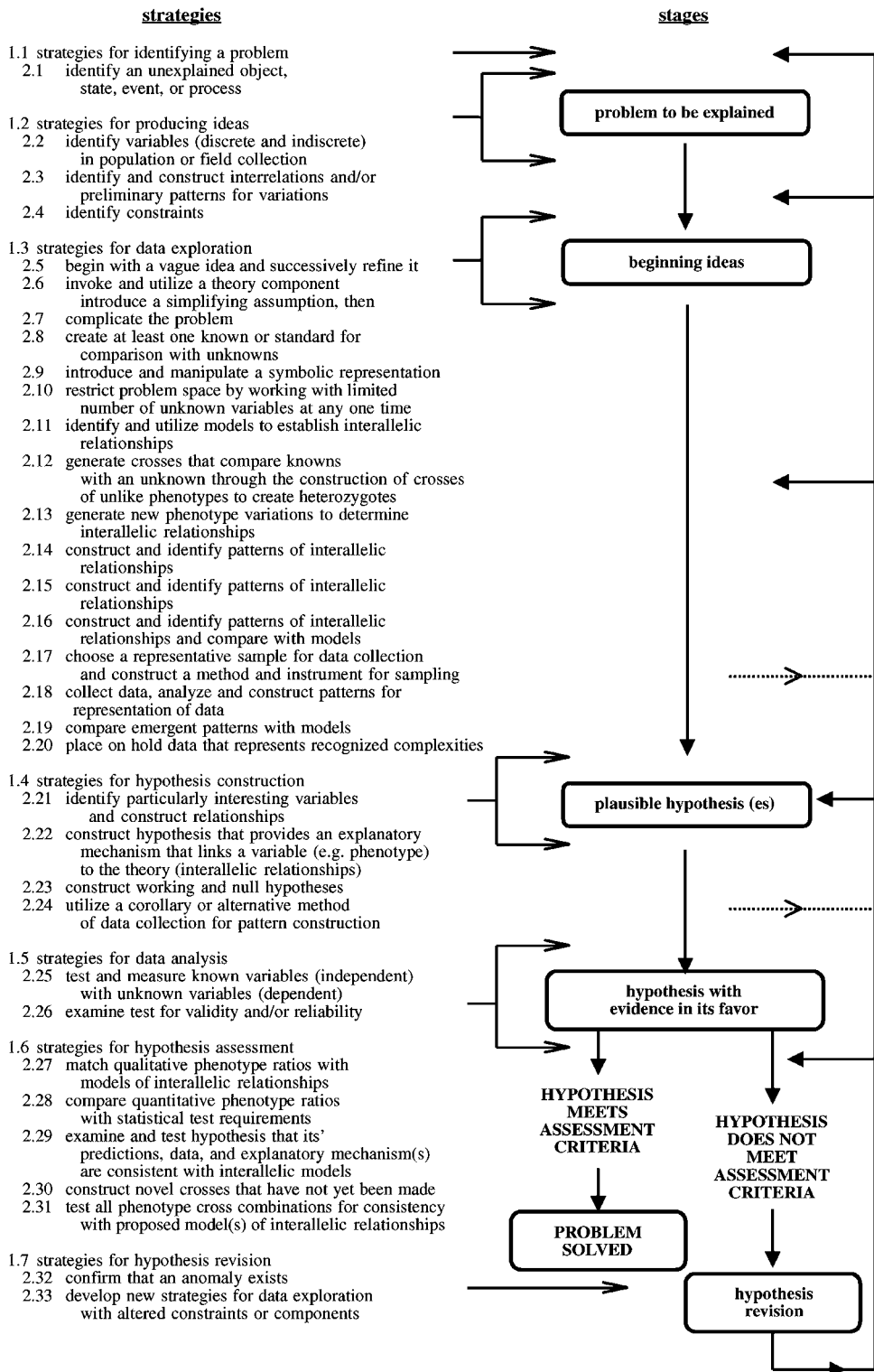


Figure 1. A general framework and pathway for problem solving that includes hypothesis construction and testing. The pathway consists of stages and strategies. Strategies are subdivided into tiers (1.1–1.7, 2.1–2.33).

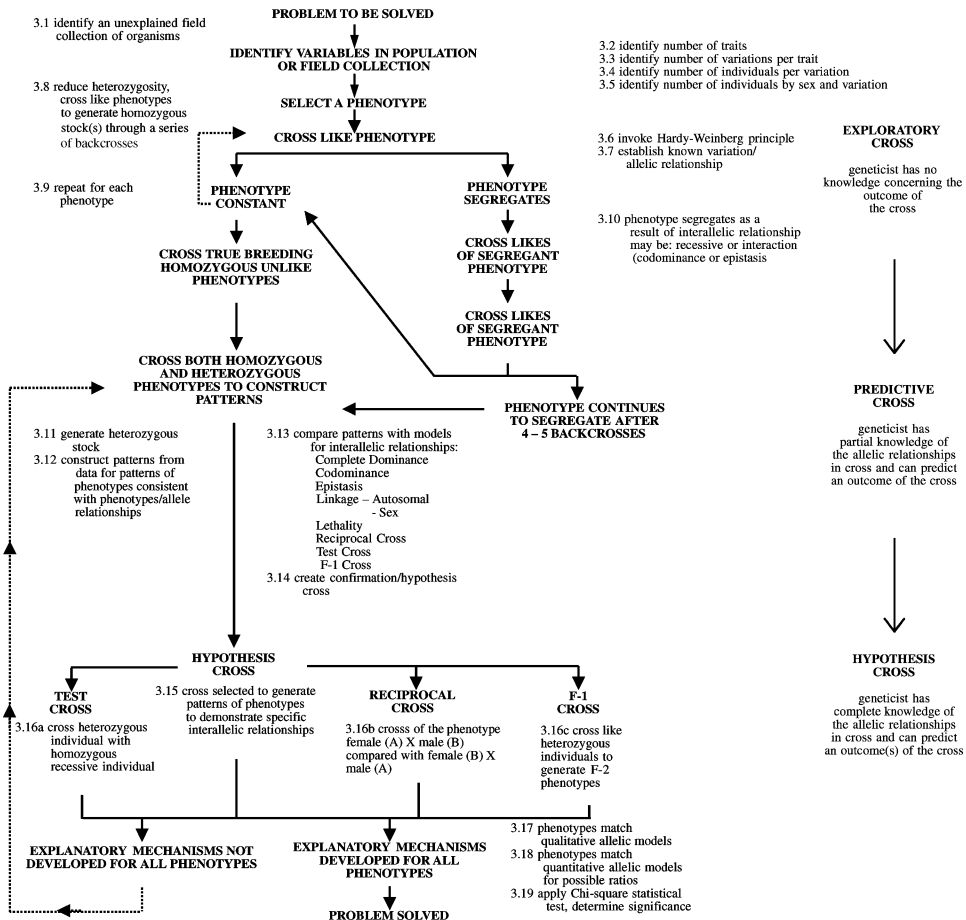


Figure 2. A pathway and strategies (3.1–3.19) used by geneticists to solve transmission genetics problems.

during solving the problem. For example, does the field population seem to fit a model of simple dominance with sex linkage? What individuals can be crossed to examine if this relationship exists and, what should be a reasonable outcome?

A pathway for solving problems was designed and crosses were categorized whether they were used for exploratory, predictive, or hypothesis inquiry (Figure 2). Finally, for each problem type, a numerical summary analysis of the crosses made by the geneticists was constructed to provide a comparison of the crosses utilized in solving a problem. Each geneticist reviewed his analyzed transcripts and the data analyses to confirm the accuracy of the researcher’s interpretations.

Geneticists’ Problem Solving Strategies and Use of Knowledge

The results are divided into two sections. First, a description of the geneticists’ general strategies and knowledge used in solving problems is presented with selected quotations. Second, an analysis of cross categories made during problem solving is presented.

In the first section, quotations from individual problems are referred as to Type (T1–T4) or post-problem solving interview (I), and each geneticist (G1–G6) is referenced to reflect a quotation’s context and source. In the text, each strategy is numerically referenced in

parenthesis (1.1–1.7, 2.1–2.33, refer to Figure 1; 3.1–3.19, refer to Figure 2). As presented, the strategies are neither numerically sequential, nor are all referenced in association with the selected quotations. Although the geneticists made many pauses during problem solving, very few are indicated (. . .) in the quotations.

Exploring the Problem

Strategies for Identifying a Problem (1.1). The geneticists consistently began problem solving by identifying and reviewing the variables and population numbers (2.2) presented in the field collection and compared it to their knowledge and experiences in order to establish what they knew and did not know about the problem (3.1):

. . . when I started a problem I would look at the population. I would assume it was characteristic of a natural population, which in general, it wasn't, heterozygotes were missing and that sort of thing, so that generated complexities. The initial crosses I would just sort of do blindly. (I, G6)

. . . if a population isn't undergoing selection, mutation, or drift, or movement-migration, then you expect, the various genotypes to be in Hardy–Weinberg proportions. (I, G3)

Strategies for Producing Ideas (1.2). Following an initial assessment of the field collection the geneticists next proceeded to identify and construct possible interrelations (2.2, 2.3) among the phenotypes. They considered the variables independently, collectively, and comparatively to what kind, if any, meaningful patterns (e.g. ratios, including linkage) might be constructed that would be consistent with known models suggestive of particular allelic relationships (3.6). Qualitative and quantitative data were considered. Such data may or may not have been useful and was also somewhat dependent upon whether there appeared to be obvious deviations from expected theoretical models for normal distributions (2.4).

Sample size also influenced the assumptions the geneticists made for interpreting an initial data set (field population) because a large sample size was considered to better represent potential for creating patterns that could be compared with known models (3.6).

If, these initial numbers were larger and, if, I were willing to assume this was from a randomly mating population . . . if I found a larger number of females than males, that would make me suspect X-linked inheritance. I would probably initially suspect that the rare classes are the recessive genes. That isn't necessarily true, but it's a good hypothesis. I would make use of population [field collection] data in those two regards. (I, G2)

Strategies for Data Exploration (1.3). In order to begin data exploration, it was important to create at least one initial comparative standard or constant as a reference point to reduce the initial problem space (2.8). The geneticists' stated that they would usually begin problem solving in an exploration phase by crossing individuals with like phenotypes to reduce heterozygosity (3.8) and establish true breeding lines (3.8). One geneticist would first make an extended series of iterative backcrosses (5–7) for each phenotype to ensure that an entire stock of homozygous individuals had been generated (3.8). The other geneticists would, more often, limit the number of backcrosses to 2 or 3 and to only one phenotype.

I would have gotten some of the earlier answers faster if I had done a little more and inbred once or twice. I'd do that for 2 or 3 generations, if I was unable to get homozygous true breeding lines . . . there would be some heterozygotes . . . I would soon discover that no amount of inbreeding would give a homozygous recessive. This breeding criterion would

reveal that faster than anything else. You might have codominance or incomplete dominance. (T1, G5)

Generally, it was assumed that individuals in the field population were heterozygous and crossing like phenotypes could result in segregation of hidden recessives or unseen phenotypes (3.10). Exposing additional unknown phenotypes could be informative in several ways. If a novel phenotypic class emerged, those offspring were then known to express an unknown allelic combination, at least relative to the particular parental phenotype being investigated. Furthermore, the emergence of a recessive phenotype could potentially represent a homozygous line that could then become the standard for testcrosses.

Once the geneticists had identified a seemingly satisfactory number of variables associated with the field population (early during solving) and created a standard, they then pursued their investigations one variable at a time (2.10). Outcomes of crosses were usually first made with comparisons to the standard they had created (2.14), and then, reviewed with respect to potential allelic relationships for other variations (2.16).

My general strategy is to try to establish strains from phenotypes that give only that phenotype. In turn, one is doing two things: you're trying to get homozygous types, but also that tends to get out the recessive homozygotes. (I, G2)

Progeny resulting from a cross were next assessed relative to the known parental phenotypes. Consequently, the geneticists utilized a great deal of time extensively reviewing the results of each cross by asking themselves "What have I learned from making this cross?" prior to constructing their next cross (2.18).

The geneticists preferred not to begin solving a problem by crossing two unlike variables (phenotypes) believing this created additional variables:

It is not useful to begin solving a problem using individuals with two unknown genotypes only to produce another group of individuals with another set of unknown genotypes. In short, why start problem solving with two unknowns by generating a third? (I, G3)

In other words, such a method was viewed as adding to the complexity of an existing problem rather than reducing or simplifying it (2.7, 2.10).

Geneticists preferred to initially work with one phenotype at a time, generating individuals with a known reliable genotype and especially in locating a recessive allele.

Usually, I just used Occam's razor. I assumed that I had a single locus until I saw evidence that there was more than one locus. If I couldn't account for the results by assuming that there was one locus I would next do crosses assuming that I had one locus and a multiple allelic series. And, then at some point, you are bound to do a cross that doesn't fit that model, if in fact, you have two loci. And, at the point I got offspring that just didn't agree with that model, well, then I would just have to work up from there. (I, G6)

If a particular component of the problem appeared to be too complex or, as a consequence of a cross several unexplained variables were generated, the geneticists often placed on hold, data that didn't seem to fit the current hypothesis. Investigations concerning novel individuals or outcomes were postponed with the intent of returning to that part of the problem at a later stage of problem solving (2.20). The geneticists much preferred to just stay with the task at hand, rather than being side tracked by new, unexplained phenomena. Good record keeping was essential for relocating problem areas or crosses that produced unusual individuals or outcomes.

While solving a problem, one geneticist revealed he had suspected an anomaly in a previous problem beginning with the field collection that continued to bother him, but he was unable to locate it. In that problem, he neither established a standard for comparison, nor came across a homozygous recessive. Ultimately, he erroneously concluded that the solution was a model of simple dominance (2.17) with two phenotypes (a 3:1 ratio) when, in fact, it involved an epistatic relationship with three phenotypes, but a relatively “rare” recessive (a 12:3:1 ratio).

I was puzzled by our last population, it didn't seem to be in a Hardy–Weinberg distribution, because we had only 2 phenotypes and that bothered me throughout that problem. (Reference to T3, while solving T4, G3.)

The geneticists took into account that each cross made, forms a reduction in variation or “bottleneck.” It was viewed as unreliable to sample only two individuals from the field population and subsequently, use only those two parents to generate information about the total population. In fact, the initial data set (field population) was considered to contain the greatest diversity of genetic material and was extensively examined for any individual's possessing hidden recessives during data exploration. Several different data sources (2.18) were usually examined before an extrapolation of an observable effect (phenotype variation) generated a causal mechanism (allele relationships), and more data was required before a predictive or hypothesis cross would be made (2.19). Consequently, the geneticists would make several crosses of a variation, randomly sampling individuals throughout the populations, for generating ideas about the total dimensions of the problem's solution. They were also interested in constructing preliminary qualitative and quantitative relationships (2.19):

Unless . . . what would be interesting would be to find out whether, like these Js, still generate this same type of pattern. Let's just do that, let's, do that to see if there are some qualitative differences here. (T3, G5)

Another role for data exploration and generation is that it allowed the geneticist to create initial patterns, which could then be compared with known models (ratios) for established genetic relationships (2.16). It was from these models that the geneticists were then able to further test for and infer interallelic relationships. In this context, knowledge of genetics history and personal experiences played a role in data exploration.

You know, this is something you might find kind of interesting, but parts of this problem take me back to a graduate seminar about 50 years ago. Landsteiner was to give a presentation on the ABO blood groups and Bernstein had proposed an explanatory model for a series of allelic relationships, it was all kind of controversial, but exciting at the same time . . . well, I guess what brings it back to mind is that, first, it turned out to be such an important discovery, you know Landsteiner later received a Nobel prize . . . and second, his hypothesis turned out to be correct! Whereas, I am not so sure about either result for me in this problem. Ha! Ha! (T2, G2)

Strategies for Hypothesis Construction (1.4). Once there appeared to be sufficient knowledge that a particular explanatory model was useful and testable the geneticists tried to develop crosses (3.11, 3.12) that were consistently predicable in their outcomes (2.22). They were in search of consistent qualitative and/or quantitative phenotypic outcomes (3.13). Individuals were crossed (3.14) from several different sources (field population

and subsequent crosses) prior to making final judgments concerning allelic relations (2.24, 3.13).

It has to do with the expectation . . . certain kinds of segregation ratios are expected and if you don't get them, well, its conceivable that it's not a simple Mendelian genetic situation or it is just drift, chance as a basis, and through replication one can see if that's the case. Now, the problem is that you can make some false assumption like I found out in a 1:1 ratio. What I should do, is do that again and see if I get it again. (T4, G1)

Strategies for Data Analysis (1.5). Data was interpreted and compared to standard phenotypic ratios or Chi-square statistical tests (2.26). Qualitative data was as important as quantitative data in constructing problem solutions (2.25). In some instances, the qualitative data was most important, "Did one or two phenotypes appear?" One geneticist did Chi-square analyses and calculations from memory. However, the geneticists generally chose not to use formal statistical Chi-square analyses to support their hypotheses, but rather relied on their "eyeball statistics." However, they did state if their data were to be used as a record of formal research they would do more crosses and extensive statistical testing (2.26).

Ideally, hypothesis testing allowed the geneticists to make novel crosses (3.16a, b, c) which were able to examine multifaceted components of the problem, the possible limits or constraints of the hypothesis, and the usefulness of the proposed model(s) which constituted the explanatory mechanism(s) developed for each phenotype (2.22). In some cases, multiple hypotheses might be constructed if a particular cross, selected by phenotype, might result in more than one possible outcome. The geneticists emphasized that it was important that the data be consistent with particular expected models (2.27). The geneticists usually did all possible crosses (including reciprocals) between phenotypes (2.31) and constructed a record keeping system or matrix for verification. They also continuously checked for internal consistency of the data between any relevant crosses that had been made and examined whether previous data matched the current model being explored (2.28, 2.29).

A 9:3:4 ratio is a little hard to tell from a 2:1:1 ratio with small numbers . . . I think a practicing geneticist uses a Chi-square test, of course, but almost always with any real organism there are viability effects that come in, that will obscure the true ratios. So I am more impressed by finding an all or none result if I can find one . . . if I can make a cross that doesn't have to count numbers at all. But, then if I have to count numbers I'd like to have a deviation from a 1:1:1:1 rather than a 9:3:3:1 . . . (I, G2)

Strategies for Hypothesis Assessment (1.6). Once the geneticists were able to construct predictable crosses, the next stage in solving a problem was to develop strategies for hypothesis assessment. The qualitative presence or absence of expected phenotypes (2.27, 3.17) was as important as the quantitative outcomes (2.28, 3.18). Hypotheses were expected to have explanatory adequacy, predictive adequacy, and the data and explanatory mechanisms needed to be consistent with their proposed models for interallelic relationships. They would construct hypothesis confirmation crosses in which any two individuals could be selected at random. And, given their phenotypes, the outcomes of the cross could be predicted (2.29, 2.30).

As previously mentioned, known homozygous recessive individuals can be used to make test crosses against other phenotypes. Use of recessive alleles allowed the geneticist to produce a small "testcross" (e.g. F_1 testcrosses) ratios of 1:0, 1:1, 1:2:1, or 1:1:1:1, which were preferred to large "dihybrid" (e.g. epistasis) ratios such as 3:1, 11:5, 12:4, and 12:3:1. The geneticists' argument was that it was easier to interpret smaller deviations rather than

those associated with larger numbers and ratios in which seemingly large deviations in progeny phenotypes in natural populations may or may not be significant (3.19). For example, in epistasis several close ratios are possible for the occurrence of two phenotypes, such as 12:4, 11:5, and 9:7, which depend upon quite different allelic relationships.

And, a given cross might result in disproportionate variance between observed and expected numbers of progeny that make the match between data and a hypothesis difficult to interpret. Consequently, establishing phenotype/genotype relationships based on phenotypic numbers of progeny that are highly variable is considered to be a difficult method of problem analysis and not an efficient use of other possible cross types. The geneticists all emphasized the importance of known allelic recessives that are used as established standards in genetics research.

... in a genetic analysis there's nothing so useful as to have a strain that is homozygous recessive, a test ... in essence one can use it in test crosses to reveal phenotypes ... it goes back to the fact that a test cross reveals the genetic composition of the gamete itself. And, it is complicated by dominance so that individual for individual one can get more information. (I, G3)

The geneticists stated that in their experience of working with natural populations of organisms, environmental effect or "noise" could mask numerical data confounding their interpretations.

As a general rule, that's the philosophy I've taken ... to look for 1:1 ratios and segregation, or to do an F_2 cross ... as opposed to squinting carefully at the data to see whether I see 3:1 ratios, or 9:3:3:1 ratios, or 15:1 ratios ... it doesn't take much in the way of dominance or in terms of noise in the data to convert a 3:1 ratio to a 2:1 ratio ... I really don't trust going beyond the 1:1 or 3:1 ratios. Where there seemed to be some deviations from the expectations I did go back and reiterate the process 2 or 3 more times to see if we saw the same deviations more or whether we saw regression to the mean ... (I, G1)

Strategies for Hypothesis Revision (1.7). There were two major sources of anomalies, which would require the geneticists to revise or improve a hypothesis. One source was a consequence of the nature of the genetics problems presented, such as, the segregation of new phenotype classes or unusual changes in progeny numbers after initial hypotheses had been constructed. (As indicated below, some frustrations were expressed during problem solving).

Ok. And this gives us 23 Jade, 22 Jade, 10, N and 9 N and we're going to have figure out what the ~~→<Ⓜ!~~ [expletive deleted] ... N is ... so N equals Nile, must be Nile green, well so that's quite amusing and uh, let's try ... well, either at this point ... either there is, ah, dominance or incomplete dominance ... I am going to press on here and continue my attempt to get a true breeding Jade line to see whether that's possible, maybe if Jade is inherently a heterozygous type it will be impossible for me to do that and that will become painfully obvious very rapidly. (T3, G1)

Alternatively, a geneticist might construct a hypothesis prior to identifying all of the components of a problem. New or anomalous data was interpreted as a challenge to the current hypothesis under test. Or, inconsistencies could arise because of a problem's design for our research, which included combinations of factors such as atypical or biased field collections, sampling error, that were seen as contrary to Hardy–Weinberg theory (2.33). In problems that seemed to include multiple explanations, the geneticists would first try to

reproduce the anomalous data to determine whether an anomaly was real, or if the outcome was a stochastic occurrence (2.32).

Alternatively, if an anomaly appeared to be a major problem the geneticists had to decide whether it was necessary to reanalyze the data on a large scale in order to localize the problem (for example, a geneticist would review his causal models or hypotheses).

You know, once you have invested a lot of time and collected an inordinate amount of data with one hypothesis in mind, it's very difficult to give it up and simply replace it with another. In light of any new hypothesis you have to go back and review all previous data before moving ahead. You just don't create data without something in mind, and all data tells you something (I, G2)

Well, let's see, I got hung up on that one problem to see if a three class segregation fit into a 1-gene incomplete dominance model. Let me see, I had to give up on that when the numbers just didn't fit. I've forgotten what . . . exactly what the cross predicted . . . 2:1:1, I just didn't get enough numbers that statistically could convince me that that was a good fit. (I, G5)

In such cases, the geneticists would need to develop new strategies for the exploration of data and revise the genetic relationships that might exist within the problem (2.33). The geneticist would review previous data and then perform additional crosses to expand data exploration leading to the development of inclusion of a new explanatory mechanism and that would be in conjunction with hypothesis modification.

. . . ummh, I guess I'll try C with L, I'm not really testing any particular hypothesis here . . . well, we have a new class coming out, M. That helps a little . . . (T1, G2)

One other major source of difficulty in problem solving was the geneticist's own knowledge and models concerning what data the geneticist chose to interpret and construct for a problem solution. For example, a geneticist might choose to focus on one set of variables and would be unable to view the data in any other way. Or, a geneticist could use correct, but incomplete models, and the further generation of data would not support the hypothesis. In fact, the outcome of a cross might totally contradict a current hypothesis under consideration. The geneticist would then be unable to locate the anomaly and consequently, the problem would be considered unsolvable.

I think I've reached an impasse with this hypothesis, I think I can reject this hypothesis, because I . . . this works fine, the first couple [of crosses] work fine, but this one won't go here, but as soon as I get to R, I am in trouble . . . it isn't apparent to me that that's the answer . . . I am not happy with any of them now, and . . . give me another sheet of paper . . . let's go to another problem (T1, G6)

The geneticists were reluctant to generate any data without a specific purpose stating that additional data would only add to the complexity of an unsolved problem. They stated that reviewing existing data with respect to hypothesis revision was time consuming and frustrating.

Evaluating a Solved Problem

After a geneticist was fairly confident that a possible explanatory mechanism had been developed for a particular set of phenotypes, novel hypothesis crosses were carefully constructed to confirm/disconfirm that predictable patterns would emerge for specified allelic relations. As one geneticist stated

... so, we'll pretend for the moment that we are dealing with one locus ... Ok, so we are crossing an L female by an L male and we got a different L female from the first one and a different L male from the first one, but we still get out only L offspring. ... Uh, I think the hypothesis is right. Conclusion. Sex-linked, 1-locus, 3-alleles. Bingo! (T1, G4)

Explanatory mechanisms were supported through qualitative and quantitative/statistical analyses and, if there was reasonable consistency between hypothesis and data, a problem was considered solved. If, confirmation cross data was inconsistent with the postulated mechanism(s), the geneticist developed more exploratory crosses to generate more data in search for different patterns from which revised hypotheses and interallelic relationships could be constructed.

In the genetics problems we faced here, I think that it's relatively easy to disprove hypotheses, but its much harder to prove hypotheses ... sometimes when the ratios aren't quite what you expect, its best to iterate the experiment a number of times with the same individuals, with different individuals, to see whether there is something, in fact, whether the deviation is real. As far as holding onto a hypothesis, as I have said proving a hypothesis is nearly impossible, but making it likely is what you are really interested in. (I, G1)

Analysis of Cross Categories Made During Problem Solving

The six geneticists chose to work on a total of 29 problems of which 27 were deemed by the geneticists to be successfully solved. Two geneticists each decided he had not successfully solved one problem (a Type 1 and a Type 4) and a third geneticist erroneously deemed a solution to be correct (a Type 3). Time used by the geneticists to solve each problem ranged from 1 to 4 h, so the total number of problem solving sessions varied among the geneticists. The number of crosses used per problem, across all problems worked ranged from 8 to 47.

As stated earlier, the problems used in our research were, in a general sense of two kinds: nonepistatic with linkage (Type 1 and 2), and eipistatic (Type 3 and 4). However, despite differences in type, there were commonalities with respect to the kinds of crosses the geneticists used during problem solving. In a comparison of number of crosses made between like (47%) versus unlike (53%) phenotypes, the proportions are nearly equal (Table 3). However, like crosses were most often used to establish true breeding lines (Table 4), test for segregants or interaction, and occasionally, used to generate F₂ ratios. Crosses of unlikes were almost exclusively used in testcrosses. In contrast, with respect to purpose, most of the crosses made were used in data exploration (72%) compared to those used for hypothesis testing (28%). The proportions correspond with the time, emphasis, and importance the geneticists placed on data exploration prior to any hypothesis testing.

TABLE 3
Numerical Summary Analysis of Crosses Made by the Geneticists in Problem Solving Across Problem Types (T1–T4)

	T1	%	T2	%	T3	%	T4	%	%
Number of Crosses	64		40		94		65		
Crosses of Like	26	41	15	38	56	60	33	50	47
Crosses of Unlike	38	59	25	62	38	40	32	50	53
Purpose of Cross									
Exploratory	30	47	17	42	41	44	18	28	40
Predictive	18	28	13	32	33	35	21	32	32
Hypothesis	16	25	10	25	20	21	26	40	28

TABLE 4
Matching Phenotypes with Possible Genotypes in a Complete Dominance Problem

No. of Phenotypes Observed in Progeny	Observed Phenotype Ratio	Variation Observed	Possible Genotypes of Progeny	Possible Parental Genotypes Crossed	Parental Cross of Like Phenotypes	Cross Number
1	1:0	Variation 1	AA	AA × AA	Variation 1 × Variation 1	[1]
1	1:0	Variation 1	AA, Aa	AA × Aa	Variation 1 × Variation 1	[2]
1	1:0	Variation 2	aa	aa × aa	Variation 2 × Variation 2	[3]
2	3:1	Variation 1 and 2	AA, Aa, Aa, aa	Aa × Aa	Variation 1 × Variation 1	[4]
Parental cross of unlike phenotypes						
1	1:0	Variation 1	Aa	AA × aa	Variation 1 × Variation 2	[5]
2	1:1	Variation 1 and 2	Aa, aa	Aa × aa	Variation 1 × Variation 2	[6]

GCK problems begin with observed phenotypes (effect) and the geneticist must infer the possible genotypes of the offspring and parents (cause), in order to construct a testable hypothesis to test (effect).

SUMMARY

Given the genetics problems presented in our research, geneticists viewed their goal as one of identifying all possible phenotypes and establishing their allelic relationships. The greatest amount of time and effort was spent in data exploration (constructing, making, and interpreting crosses) leading to hypothesis creation in comparison with hypothesis testing and assessment. It was during data exploration that integration and use of context-specific strategies and genetics knowledge was used most extensively. Thus, the construction of a meaningful hypothesis was the more time consuming task, whereas, hypothesis testing was treated as a routine procedure. Hypothesis testing usually involved constructing several alternative crosses all of which were expected to result in a predictable pattern consistent with a genetic model. The geneticists emphasized that use and the status of a hypothesis was explicitly limited to experimental qualitative or quantitative statistical testing. What some scientists propose to call research hypotheses or best guesses (see Sattler, 1986), the geneticists called exploratory ideas. And, during hypothesis testing as one geneticist stated, "I prefer to construct a cross that will provide answers to multiple hypotheses, such as testing an allele for sex-linkage and dominance."

Variability in initial populations and stochastic changes in progeny numbers were noted by the geneticists, but as one geneticist stated, "I viewed it more as 'environmental noise' than a real problem, unless the data began to suggest something important or posing a more difficult problem." However, even the initial population was scrutinized using a theoretical model, the Hardy-Weinberg law. Whenever possible, the geneticists attempted to numerically reduce data into discrete variables such as 1:0 or 1:1 ratios so they could more easily detect whether or not "noise" was a real variable for which they would need to become concerned. Throughout problem solving, the geneticists were constantly cautious about their inferences.

Although general strategies were useful in problem solving, a repertoire of conceptual knowledge and models was needed to construct crosses to produce meaningful patterns. However, not even a full complement of these factors (strategies and knowledge) could guarantee a successful solution. Darden's research used hypothesis construction and revision as a focus for theory change in science (Darden, 1991). Her findings, albeit based on historical accounts, seem to suggest that in the processes of theory change, scientists place considerably more attention on the validity and falsification of existing hypotheses. Perhaps, this is to be expected given that hypotheses are the means by which theories are tested whereas our problems involved data comparison with hypotheses within accepted theories.

In our research, all six geneticists first gave a preliminary review of the observable trait (by variation, numbers, and sex) and phenotypes presented in the field collection. But, they then consistently tried to create a homozygous recessive genotype that could serve as a "standard of reference" for use throughout the rest of the problem. In particular, during data exploration, they chose one potential variation (usually based on the lower numbers of individuals found in the field collection) to reduce heterozygosity through self-fertilization. For example, one geneticist would consistently make 5–7 crosses to develop a true breeding line for each variation, whereas others were satisfied with 1 or 3 crosses. Mendel spent 2 years establishing true breeding lines.

Initial crosses between likes were considered helpful in locating segregants. If a segregant appeared, that was interpreted to be a recessive character. And, the geneticists would immediately begin making crosses between those individuals. A standard was considered useful for reducing problem space and essential for hypothesis construction and testing. Mendel (1866), at least partially recognized the importance of this problem-solving strategy and one geneticist stated, "either Mendel knew much more than is revealed in his paper or he was damn lucky" (the geneticist preferred his first guess).

If a true breeding line could not be established for a specific phenotype, the geneticists would infer that it suggested a genotype exhibiting partial dominance, a more complex part of the problem requiring that phenotype to be placed “on-hold” and investigated after solving easier parts of the problem. This situation was most apparent in problems involving epistasis in which gene interaction was involved. For example, if a cross of two like phenotypic individuals resulted in a new phenotype and the problem involved more than simple codominance, the geneticists seemingly used a nonlinear exploratory pathway to try and locate a constant. The Type 3 problems were viewed as the most difficult to work with because a ratio of 12:3:1 made it difficult to locate a homozygous recessive.

Though individuals selected for the first cross from the field collection were randomly chosen (within a phenotype), geneticists usually began to refine their ideas and develop more knowledge about the problem (2.5) with each set of generated data. That is, each cross was carefully chosen, but more importantly, subsequently considered to be productive in providing some kind of information concerning the nature of the problem (2.11). Thus, the results of crosses were judged in detail with respect to their usefulness in helping to solve the problem. No cross was considered uninformative. Rather, a “negative” or unexpected outcome was considered to be as important as a “positive” outcome. However, this did not deny that some crosses were more informative or productive than others. For example, almost every cross made, provided evidence for either constructing homozygosity (a true breeding line) or revealing heterozygosity and segregants (a recessive phenotype/allelic combinations).

IMPLICATIONS

Simon (1962) has posited that “scientific discovery is a form of problem solving, and . . . the processes whereby science is carried on can be explained in the terms that have been used to explain the processes of problem solving” (as cited in Bechtel & Richardson, 1993, p. 1). The question to be asked, then, Are the problem-solving strategies presented in this paper transferable or applicable to other areas of inquiry in science? To that question our answer must be, “We do not know!” At lower levels of resolution the general strategies in the framework (Tiers: 1.1–1.7 and 2.1–2.33) developed in our work are consistent with Darden’s findings with respect to her stages and general strategies used in theory change. However, problem solving in science is ultimately tightly linked to conceptual knowledge within a discipline and ultimately strategies become problem specific for example, the strategies in Tier 3.1–3.19 and the individual crosses (Examples) geneticists made. Thus, the foundation of scientific inquiry is based on the structures and meanings of discipline specific knowledge. Meaningful inquiry, then, can only occur in the context or space of a specific problem (see Stewart & Rudolph, 2001). In addition, time is required for in-depth problem exploration to create data patterns that allow interpretation and comparison with models through hypothesis testing.

GCK allows teachers and students to experience four of the five NSE unifying themes that are considered consistent with the practice of science and inquiry based learning: (a) organization; (b) evidence, models, and explanation; (c) change, constancy, and measurement; and (d) equilibrium. Further, GCK can support a learning environment where students’ knowledge construction allows them to “instantiate their thought processes . . . using models that can be examined, evaluated, discussed, and reflected upon” (see Penner, 2001, p. 29). In inquiry, models provide the lens through which data can be interpreted and patterns constructed. Students need problems, and time, to determine and examine how variables might be related prior to hypothesis construction. As the geneticists stated, GCK generated genetics problems provide reasonable simulations of “real world” phenomena and differ substantively and substantially in comparison with textbook problems. GCK

provides a dynamic problem-solving environment in which populations of organisms have randomized phenotype numbers, in contrast with static populations and stereotyped phenotype numbers typically presented in textbooks. And, because GCK problems are based on probability, hypothesis assessment and revision must be based on qualitative and statistical outcomes generated by the solver.

Most classroom laboratory activities, experiments, and textbook problems engage students in a one-hypothesis test. When the outcome of an experiment fails, often time is not provided to reflect upon the perceived failure or to examine “What have I (we) learned?” The geneticists used failure to confirm a hypothesis, as a positive experience. First, they could no longer necessarily accept the particular postulated cause, but they could then move forward to test for data that might support an alternative model. The geneticists in our research spent most of their time in data exploration scrutinizing variables that could lead to hypothesis construction. In contrast, most science classroom activities and experiments (and published science research papers) typically begin with stated and unquestioned variables that limit students to hypothesis testing and confirmation, rather than providing them with opportunities to explore the ways variables may be related (see Kantorovich, 1993).

When the geneticists were asked whether students should be told by a teacher that creating homozygous lines through inbreeding could quickly lead to solving a problem, i.e. invoke a problem solving algorithm, there was a unanimous and emphatic, “By all means, NO!” Rather, they suggested that a teacher should encourage students to critically read Mendel’s paper and question how he might have been able to construct true-breeding lines (Corcos & Monaghan, 1993; Mendel, 1866; Olby, 1985) and examine their ideas using GCK. Students need to develop an understanding of how “standards” are established and used in scientific research. In at least one high school classroom, GCK is effectively, extensively, and successfully being used (see Stewart et al., 1993; Stewart & Hafner, 1994; Stewart & Rudolph, 2001). And, one author (Thomson, unpublished data) has used GCK to successfully introduce middle school students and in-service elementary and middle school teachers to inquiry-based genetics problem solving. What has been challenging is to encourage “Punnett square” thinkers to reconceptualize genetics problem solving. Transfer of static textbook-based problem solving to dynamic simulations is not a one-to-one phenomenon. Thomson and Chepyator-Thomson (2002) have found that Kenyan students more easily go from GCK problems to textbook problems than vice versa.

GCK models Mendelian and non-Mendelian genetic systems that can serve as an important foundation for understanding both simple and complex mechanisms for inheritance. GCK is also able to provide a problem-solving environment through which geneticists’ research strategies can be experienced. GCK problems invite the solver to conduct “data exploration” and “self-discovery” that is consistent with the nature of scientific inquiry (NRC, 1996, 2000). Finally, as with Darden’s work, science education researchers are encouraged to examine the usefulness of the framework developed in our research to other areas of scientific inquiry. And, as advocated by Keys and Bryan (2001), we hope that science education researchers can use our framework of problem solving strategies as a referent model for examining science teachers’ beliefs and knowledge concerning scientific inquiry in the context of inquiry-based instruction.

REFERENCES

- Allen, T., & Starr, T. (1982). *Hierarchy: Perspectives for ecological complexity*. Chicago: University of Chicago Press.
- American Association for the Advancement of Science (1993). *Benchmarks for science literacy, Project 2061*. New York: Oxford University Press.

- Bechtel, W., & Richardson, R. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.
- BioQUEST. (1998–99). *BioQUEST Library (Vol. V)*. New York: Academic Press.
- Bird, M., & Weller, H. (March, 1997). *Cultures in conflict: The role of teacher knowledge in inquiry science lessons*. Paper presented at the meeting of the National Association of Research in Science Teaching, Chicago, IL.
- Corcos, A., & Monaghan, F. (1993). *Gregor Mendel's experiments on plant hybrids*. New Brunswick, NJ: Rutgers University Press.
- Darden, L. (1991). *Theory change in science: Strategies from Mendelian genetics*. New York: Oxford University Press.
- Darwin, C. (1866). *The origin of species*. London: Dent & Sons.
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, 35, 125–129.
- Flick, L. (March, 1997). *Focusing research on teaching practices in support of inquiry*. Paper presented at the meeting of the National Association of Research in Science Teaching, Chicago, IL.
- Futuyma, D. (1998). *Evolutionary biology*. Sunderland, England: Sinauer.
- Gell-Mann, M. (1994). *The quark and the jaguar: Adventures in the simple and the complex*. New York: Freeman.
- Gleick, J. (1987). *Chaos: Making a new science*. New York: Viking.
- Kantorovich, A. (1993). *Scientific discovery: Logic and tinkering*. Albany, NY: State University of New York Press.
- Keys, C., & Bryan, L. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching*, 38, 631–645.
- Keys, C., & Kennedy, V. (1999). Understanding inquiry science teaching in context: A case study of an elementary teacher. *Journal of Science Teacher Education*, 10, 315–333.
- Mayr, E. (1982). *The growth of biological thought*. Cambridge, MA: Harvard University Press.
- Mendel, G. (1866). *Versuche uber pflanzen-hybriden*. *Verhandlungen des naturforschenden vereines in Brunn*. *Abhandlunger*, 4, 3–47.
- National Academy of Sciences (1998). *Teaching about evolution and the nature of science*. Washington, DC: National Academy Press.
- National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.
- National Research Council (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academy Press.
- Odum, E. (1971). *Fundamentals of ecology*. Philadelphia: W.B. Saunders.
- Olby, R. (1985). *Origins of Mendelism*. Chicago: University of Chicago Press.
- Penner, D. (2001). Cognition, computers, and synthetic science: Building knowledge and meaning through modeling. In W. Secada (Ed.), *Review of research in education* (pp. 1–35). Washington, DC: American Educational Research Association.
- Peixoto, J., & Oort, A. (1992). *Physics of climate*. New York: American Institute of Physics.
- Sattler, R. (1986). *Biophilosophy: Analytic and holistic perspectives*. Berlin: Springer-Verlag.
- Simon, H. (1962). Scientific discovery and the psychology of problem solving. In R. Colodny (Ed.), *Mind and cosmos*. Pittsburgh: University of Pittsburgh Press.
- Stewart, J., & Hafner, R. (1994). Research on problem solving: Genetics. In D. Gable (Ed.), *Handbook of research on science teaching and learning*. Riverside, NJ: MacMillan.
- Stewart, J., Hafner, R., Johnson, S., & Finkel, E. (1993). Science as model building: Computers and high school genetics. *Educational Psychologist*, 27, 147–336.
- Stewart, J., & Rudolph, J. (2001). Considering the nature of scientific problems when designing science curricula. *Science Education*, 3, 207–222.
- Tobin, K., Kahle, J., & Fraser, B. (1990). *Windows into science classrooms: Problems associated with higher-level cognitive learning*. New York: Falmer.
- Thomson, N., & Chepyator-Thomson, R. (2002). Keiyo cattle raiding, *Kechui*, mathematics and science education: What do they have in common? *Interchange*, 33, 49–83.