A PRELIMINARY LOOK AT TIMING OF FEEDBACK IN TUTORING SYSTEMS

by

ADRIENNE Y. LEE

Department of Psychology
University of Colorado
1989

ICS Technical Report #89-10

ABSTRACT

During the learning process, students typically are given some task to do and then given feedback as to whether or not they correctly performed the task. Feedback timing and content is therefore an important issue in learning. Timing of feedback can be either immediately after an error is made (immediate) or after the student completes the problem (delayed). Numerous non-empirical and amorphous arguments have been given for delayed feedback. Although there have been qualitative changes in learning theories from Thorndike to Anderson, these learning theories have generally supported immediate feedback after errors.

The early learning theories were based on simple tasks, while modern learning theories are based on very complex tasks. These tasks range from learning micro-processor based test instruments (Lee, Polson, & Bailey, 1989) to full classroom courses (Anderson geometry tutors). The most ambitious, controlled learning studies have been done on intelligent tutoring systems (Anderson, 1987). Of these systems, Anderson alone has applied a unified theory of learning and skill acquisition to the development of principles which could be used in tutoring system creation. The study described in this paper attempted to use Anderson's principles to develop a tutoring system for genetics that would test immediate versus delayed feedback.

This paper will start with a review of the relevant historical literature on feedback issues starting with theories based on animal learning to those based on human learning. This review will be followed by a description of the development of a tutoring system for pedigree trees. The empirical study and results are summarized and a discussion of these results is provided.

Results of the experiment indicate that immediate condition subjects learn more quickly than delayed condition subjects, but performance of the delayed condition subjects was qualitatively different on a post-test the following day. Thus, despite consistent arguments by learning theorists that immediate feedback is better than delayed, under some conditions delayed feedback may have some advantages.

CONTENTS

CHAPTER I

INTRODUCTION

This paper presents research on a genetics tutoring system. The issue is whether immediate or delayed feedback after error is better. The major portion of the historical issues as well as the problem space analyses and production system can be found in Lee (1989).

Chapter 1 gives a brief overview which defines terms, sets the context for the issues, and presents the historical background (a review of Anderson's research, other tutoring systems) and the main hypotheses that were tested. Chapter 2 contains a detailed description of the tutoring system and the methods section for the evaluative experiment. Chapter 3 contains the main results of the experiment. Finally, Chapter 4 contains the discussion and conclusion.

Overview

Tutoring systems are educational computer software that help a student learn a topic by providing basic information, practice questions which can be presented in a pre-determined sequence or an individualized sequence, and individualized help when problems arise. Although computer systems probably will never replace a human teacher or tutor, they provide some benefits. A computer can be patient, can remember all of the student's actions, and can provide individualized help (Collins, 1988). In addition, the computer can provide different perspectives on material and maintain multiple hypotheses about student responses to questions (Wenger, 1987). These systems can provide a context in which students explore a topic productively and an environment in which they can easily form and test hypotheses (Collins, 1988). To build a tutoring system, not a simple task, not only do we need to understand how computers work and build an acceptable interface, we also must understand the learning process itself and how to integrate the practical and theoretical to form an effective system.

The process of learning has been studied extensively in psychology. From 1898 to almost 1930, Thorndike's learning theories dominated all others (Thorndike, 1949). Thorndike was interested in the "bond" or "connection" between sense impressions and impulses to action. His theories are known as the original Stimulus Response (S-R) psychology of learning. His Law of Effect states that reward strengthens the stimulus-response association (Thorndike, 1911,1913). In a sense, if a response is rewarded, then the connection will be learned. For learning studies and tutoring systems, feedback timing often is defined as the time between when an answer is given by a student and the time that the tutoring system responds to that answer. An implication of Thorndike's theory for tutoring systems would then be that a tutoring system should reward a student's correct response immediately after it occurs.

Although many other psychologists worked on Thorndike's theories by expanded them and creating new learning theories, Skinner (1954) was the first to apply these theories to teaching machines. Skinner promoted errorless discrimination learning. He argued that learning would be better when students did not elicit an incorrect response (make an incorrect answer). His machines therefore were designed to elicit the correct answer and a student using his machines would not receive any feedback until the correct answer was given (Skinner, 1958).

Skinner's ideas on feedback timing were very similar to the implications for feedback timing derived from Thorndike's Law of Effect. Skinner created a teaching machine which provided immediate feedback for correct responses. However, given the technology at the time, Skinner's teaching machines could not keep track of more complicated problems with many parts or multiple responses given by a student. Therefore, Skinner did not test his ideas against contradictory ideas about feedback timing on teaching machines.

These early learning theories were based on simple tasks; but, recent work in learning has focused on more complex tasks. The tasks can be as complicated as learning how to use a micro-processor based test instrument (Lee, Polson, & Bailey, 1989) or learning a full course in some classroom material (Anderson, Boyle, & Yost, 1985b). This recent work reflects the opinion that learning theories should be examined by means of more difficult tasks and with learning defined as knowledge that needs to be acquired in situations where it will be used (Brown, Collins, & Duguid, 1988). For example, learning by doing (Anzai & Simon, 1979) long has been seen as the best way of acquiring knowledge; however, many of the early studies and even some recent studies have not examined students while they actually were learning a realistic, complicated task. Anderson (1987) makes the point that tutoring systems can provide the vehicle for examining learning while the student actually is learning some academic topic. In addition, the tutoring system can provide an experimenter with the ability to add experimental manipulations in a realistic context.

A tutoring system not only can provide a learning atmosphere in which the student uses the skill in context but it also can record what a student does during the learning process (Collins, 1987). Thus, a tutoring system constantly can monitor a student's behavior and provide immediate feedback after an error has occurred.

However, in real-life situations, people may be required to make several attempts at a problem before being able to find assistance. The trial and error aspect of a problem may be a learning experience in itself. Further, finding the correct answer is always preferable to being told. (See the generation effect reviewed in Crutcher & Healy, 1989. Also, see Foss, 1987a, 1987b.) The implication of the advancement of technology is that computers have a greater capacity not only to remember a single answer given by a student to a simple multiple choice question, but also can monitor a student through the solution to a complicated algebra problem.

However, computer-aided instruction systems generally have been developed independent of current learning theory. Most computer-aided instruction began with linear programs which were based on the programmed learning of the 1940s and Skinner's work (1954,1958). In the mid-1960s, tutoring focused on generative computer aided instruction (Uhr, 1969; Sleeman & Brown, 1982). These systems were adaptive with student models based on behavior rather than summaries of knowledge (Suppes, 1967; Woods & Harley, 1971; Sleeman & Brown, 1982).

Other systems successfully implemented an "environment for exploration" but failed to have an explicit notion of student goal states, performance or state. Also, there were no built-in capabilities to direct, to rescue or to motivate the student (Kimball, 1982). By the 1970s, systems were developed that could generate teaching materials (Yazdani, 1987). Systems had not been developed yet which took psychological learning theories into account. Many researchers have been developing systems based on what they thought was best, but not based on empirical data. The focus changed from the development of teaching or tutoring systems when learning theories were applied to the development of tutoring systems themselves.

There is a definite need for the issues in computer tutoring systems to be considered in a systematic way (Kimball, 1982; Wenger, 1987). In addition, there is a difficulty in going from the basically descriptive nature of cognitive theories to more prescriptive theory (Singley, in press). John Anderson and his colleagues have been working on this problem and from the ACT* theory of skill acquisition, have developed some guidelines for tutoring system design (Anderson, Boyle, Farrell, & Reiser, 1984). Although these principles have been used to develop several different tutoring systems, most have not been directly validated, extended, or refined (Singley, in press). Anderson and his colleagues have been trying to remedy this situation (Singley, 1987; McKendree,1986; Lewis, in press).

The research described in here uses Anderson's principles to develop a tutoring system in an attempt to validate one of Anderson's principles that a student must receive feedback immediately after an error has occurred. Anderson addressed the question of when a system should fix a student's misconception. His hypothesis was that the misconception would persist unless fixed immediately otherwise the student would become lost or confused. This idea is similar to both Thorndike's and Skinner's ideas, even though the learning theory on which it is based is very different.

The development of increasingly complex computer systems has facilitated the development of tutoring systems with much greater sophistication. Such systems have the capability of presenting vast amounts of information to the student. They also can maintain all student responses and give complete feedback after the student has completed the problem. Feedback can be given after a longer sequence of actions and can include a summary of behavior exhibited or it can be given after every step.

If tutoring systems provide more than just "right" or "wrong" responses to questions, feedback timing can become more closely related to the issue of information saturation. That is, with variations in feedback timing, students are exposed to varying amounts of information with certain amounts of time between presentations of that information. Although students may still become lost or confused when feedback is delayed following an error, they also may become oversaturated with the system constantly interrupting them to give them more information. The question then becomes how much information should be provided to a student at any one time, or when should the system interrupt with information.

The research described in this paper addresses this question: Should students receive feedback immediately after an error is made? The forum for this issue of feedback intervention or feedback timing was tutoring systems and in particular a tutoring system which taught genetics.

The remainder of this chapter covers in greater detail the issues brought up in the preceding paragraphs. First, a summary of the historical issues is presented. This is followed by a description of Anderson's theory and feedback timing, tutoring systems, and empirical studies. Other computer systems and their feedback will then be discussed.

Summary of historical issues

Feedback timing has been studied as a consequence of the development and testing of learning theories. Early literature presents studies with simple tasks and uniform trials. Many of these dealt with non-humans rather than humans. In the instrumental learning studies, a response was followed by a reinforcement which was either pleasant such as food or unpleasant such as a shock. The animal did not receive much more information than this. With more complicated experiments, however, more information was available. In particular, humans could be told information such as how short or how long a line is compared to the target. Thus, rather than thinking of reinforcement, one needed to think of correction or informative feedback.

It is possible that differences found in studies on simple skills may be irrelevant for complex skills. In addition, a delay in feedback may actually change the content of that feedback; however, timing and content were typically confounded in most studies. In studies on humans, delay has not necessarily been found to be detrimental to learning (Bilodeau & Bilodeau, 1958; Bourne & Bunderson, 1963; Kintsch & McCoy, 1964; Kulhavy & Anderson, 1972; Roper, 1977; Rankin & Trepper, 1978; Bardwell, 1981; Gaynor, 1981). Even though these studies find some positive or neutral attributes to delaying feedback, they have not been incorporated into a learning theory or even into theories on the development of tutoring systems.

Although some researchers did attempt to tie their research on timing with their learning theory (Thorndike, 1935; Skinner, 1954), the major modern work on learning has been done by John Anderson. He has attempted not only to bring together his ideas into a coherent learning theory, but also to apply his learning theory to the development of tutoring systems. Anderson has indicated that the development of theories of skill acquisition and learning should be relevant directly to complex skills taught in the classroom (Anderson, et al. 1984). Further, he makes a forceful argument that the only way to test learning theories is by the development of tutoring systems (Anderson, 1987). These systems have the capability of teaching whole academic subjects which provide a richer experimental learning paradigm. Finally, Anderson has studied empirically the learning process through the development of such tutoring systems. For these reasons, the research described in this paper is based primarily on John Anderson's ideas which will be described in the next section.

Research of John Anderson

Anderson's major statement about learning occurs in his monograph, *The Architecture of Cognition* (1983). One of his ideas is that instruction must appear in the context of problem solving. Another is that the student must generate as much of each solution as possible. Finally, the goal structure of the problem should be represented for the student. In fact, these ideas can also be found in earlier papers by James (1890), Skinner (1958), Estes (1962), Anderson (1976), Gagne and Rohwere (1969), and Tulving and Thompson (1973). The difference is that Anderson attempted to fit these ideas into a comprehensive theory of cognition.

ACT*, Theory of learning and feedback timing

From Anderson's theory, specific predictions about feedback timing can be made. In a learning situation, a person is given often a problem to test his understanding. The person must rely on some external source for feedback as to the correctness of his or her response (such as being told by an instructor) or on some internal source (noticing an inconsistency). Errors can occur for several reasons. First, an error could occur if the production created by a weak method could be incorrect. Second, the information (declarative knowledge) could have been encoded incorrectly. Finally, information could have been lost because working memory is limited. Capacity can increase, however, with experience or practice in a domain. In general, people have difficulty maintaining the problem, problem structure, solution path, and final solution in memory due to capacity limits. Therefore, the longer the delay between the feedback and the production firing the longer the learning time due to an inability to make a discrimination (Anderson, 1983). The implication for tutoring systems is that any delay in feedback would be detrimental to the learner.

ACT* and tutoring systems

In an attempt to test the ACT* Theory (Anderson, Boyle, Corbett, & Lewis, 1986), Anderson and his colleagues built a LISP tutor (Reiser, Anderson, & Farrell, 1985; Farrell, Anderson, & Reiser, 1985; Anderson, et al., 1984), a geometry tutor (Anderson, et al. 1985b), and an algebra tutor (Lewis, Milson, & Anderson, 1987). Anderson's ideas about tutors were originally set out in Anderson, et al. (1984). These principles are presented in Table 2.

Table 2. Anderson's principles for tutoring systems.

1. Identify the goal structure of the problem space.
2. Provide instruction in the problem-solving context.
3. Provide immediate feedback on errors.
4. Minimize working memory load.
5. Represent the student as a production set.
6. Adjust the grain size of instruction according to learning principles.
7. Enable the student to approach the target skill by successive approximation.
8. Promote [the] use of general problem-solving rules over analogy.

Based on these principles, Anderson and his colleagues have successfully built tutoring systems which have been evaluated and these tutoring systems will be described in the next section.

Anderson computer tutors

Geometry and LISP were chosen because they are both well-defined skills (Anderson, et al. 1986). Several common points between the two systems developed can be defined. First, only one concept is given at time and short instructional material is provided. Second, the current goal and goal structure are displayed for the student at all times. Third, a Help system is provided. In all cases, immediate feedback is used over delayed feedback.

These features are based on a few principles. Anderson believes that the critical issue for learning is the correct interpretation of instruction. Students should be told what the critical features of the domain are. Misunderstandings and slips often can cause similar behavior. Finally, the student should not have difficulties with the interface. Some of Anderson's latest ideas on interface improvements can be seen in the new interface proposed in Lewis, et al. (1987).

Empirical evaluations

Many of Anderson's principles have not been directly tested to see what modifications, if any, need to be made to them. However, several of Anderson's students have attempted to test pieces of the theory. This section will review the empirical work in the context of the principles Anderson developed. The principles are listed in Table 2. Challenges to Anderson's ideas will be interleaved with empirical evaluations.

General problem solving, analogy, and the problem solving context

Principle eight states that the tutoring system should emphasize general problem solving skills over analogy. Anderson (1981) and Anderson, Farrell, and

Sauers (1984) indicated that students use analogies to solve problems; however, Anderson et al. (1984) suggested that students did not understand how they solved the problems. Therefore, Anderson claimed that general problem solving should be encouraged. In his later work, Anderson suggested that a student should be introduced to a topic by examples (Anderson, et al., 1986).

Instruction in the problem-solving context

The second principle states that instruction should be provided in the problem-solving context. Information learned in context will be better remembered (Tulving, 1983; Tulving & Thompson, 1973). But, should concepts be presented initially or should concepts be presented after the problems are presented? Research has shown that students provided only with concepts do not necessarily attain that concept (Klausmeier, 1976; Tennyson, 1973). Several studies have shown that students who receive concepts and examples perform better on classification tasks, multiple choice questions, and solving problems than students who received the examples alone (Johnson & Stratton, 1966; Anderson & Kulhavy, 1972; Tennyson, 1971). Clearly, the presentation of concepts to students aids in learning and therefore teaching concepts may include presenting the concept, examples and counter-examples, and problems to solve.

Representation as a production set

Principle five states that the student should be represented as a production set. Newell and Simon (1972), as well as Anderson (1983), indicated that human problem solving behavior can be modelled as a set of productions. Further, other researchers have used production systems to model student behavior (Brown & Van Lehn, 1980; O'Shea, 1979; Sleeman, 1982).

Goal structure

The first principle that Anderson set forth (Anderson et al., 1984) was that the developer must identify the goal structure of the problem space for the student. Singley (1987; in press) developed a tutoring system for related rates word problems in order to examine directly this first principle. Students in the goal posting condition performed better on the post tests than students who did not. Posting goals helped subjects to make fewer moves and fewer illegal moves. This result indicates that working memory load may be a key to problem solving. If the system can perform some of the work of maintaining working memory, students will be able to learn more quickly. Singley points out that the major problem is to establish a common language with which to express goals.

Working memory load

Anderson's fourth principle stated that the system should attempt to minimize working memory load. The seventh principle stated that tutoring systems should enable to the student to approach the target skill by successive approximation. Skinner stated the same idea in his principles (see second principle listed in Table 1.) In addition, research by Jeffries and Anderson which analyzed LISP errors made by novices found that more errors occurred as the complexity of the problem. These errors were due to processing overload rather than to misunderstanding the material (Anderson & Jeffries, 1985).

Immediate feedback

The third principle states that the system must provide immediate feedback after errors. In Anderson, et al. (1984), immediate feedback is said to be better than delayed because little can be accomplished while a student makes random attempts at a solution to a problem. In Reiser, et al. (1985), another reason is given for immediate feedback. Immediate feedback can prevent students from correcting answers which are correct and also it can prevent students from remembering incorrect rules. In Anderson, Boyle and Reiser (1985), immediate feedback is said to allow the student to integrate the feedback with the problem solution. This idea is very similar to the original 1983 formulation.

Empirical study in gaming environment. An empirical study by Lewis and Anderson (1985) provided some evidence both supporting and refuting immediate feedback. Three experiments were reported, but only the last two experiments addressed the feedback timing issue. The experiments used a complicated gaming environment in which subjects moved from room to room. Each room had things you had to do and rooms you could move to. For the first experiment, no difference in performance was found, but the subjects in the delayed condition took twice as long to complete the learning phase. For the second experiment, a post-test was given on the second day. Again, students in the delayed condition took longer to learn than those in the immediate condition; but, the students in the delayed condition could identify erroneous paths.

Updated comments by Anderson. Recently, in Anderson, et al. (1986), Anderson summarized his reasons for immediate feedback, but he also suggests that immediate feedback may have some real drawbacks. Students may rely on just making errors and having the system hand-hold them through problems. Self-correction is always preferable to system corrected problems (Foss, 1987a, 1987b; Crutcher & Healy, 1989). In addition, students may find the immediate feedback annoying. Finally, the system may have difficulty in identifying exactly why an error was made at an early point in the same way that delayed feedback may have trouble isolating a single error.

In answer to his own thoughts, the LISP tutor presently does not correct a student until a full expression has been completed. A future system may be implemented that would focus on a certain amount of code having been written before giving feedback. The key is to minimize the the distance between error and feedback without the detriments of too short a delay. Thus, Anderson has proposed changes to his own prescription.

Further research on immediate versus delayed feedback. Lewis (in press) presents a comparison between the results of experiments performed on the geometry tutor and algebra tutor. Immediate feedback was the original prescription by Anderson, et.al. (1984). The LISP tutor and the geometry tutor used immediate feedback after an error. On the other hand, the algebra tutor performed effectively even though feedback was delayed using a two-strike rule (correction after two errors occurred). Lewis' explanation was that algebra is less structured than geometry and being able to attack the problem from different directions and not necessarily on a single pre-determined path may be more beneficial. Short cuts appear to play an important part in algebra solutions. Thus, immediate feedback may only be useful for domains such as geometry where specific paths are required. In Lewis' analysis, he therefore presents a clear challenge to one of Anderson's principles.

Criticism in the literature. Anderson's ideas on immediate feedback have not been without criticism. Wenger (1987) suggests that immediate feedback may only really be useful for the early learning stages and relatively short tasks. Also, the skill of debugging in LISP could not be learned because students never learn to identify their own bugs (see Lewis & Anderson, 1985 on identifying erroneous paths.) The Lewis and Anderson (1985) study used a gaming environment which is quite different from geometry or LISP and may be considered unnatural. Also, although their results indicated that immediate feedback after an error may be the optimal situation, immediate feedback does not allow the user to see the consequences of an error (Carroll & McKendree, 1987).

The problem of not being able to identify a wrong path should be considered. The ability to inspect a path of thought, reification (Brown, 1983a,1985; Collins & Brown, 1987), allows one to look at the process and not just the product of reasoning (Brown, 1983b). In a study evaluating ALGEBRALAND (Foss, 1987a), students learned how to identify wrong paths and recover from them. In another study by Foss (1987b), students were better able to remember solutions that they had generated themselves.

This conclusion is not surprising given research by Tulving (1982) indicating that recognition may be based on the reinstatement of the original item's environment. Solving a problem may allow for encoding of states. In this sense, the longer times to learn and other problems of discovery learning may be worth the trouble if students can actually remember more of the solutions (Foss, 1987a, 1987b; Wenger, 1987). Ability to detect errors is only one of many factors that are involved in feedback timing.

Summary of empirical evidence

In summary, although most of the principles that Anderson developed have not been empirically studied, some of the principles are supported by previous research and research performed by McKendree, Singley, and Lewis. In addition, the studies of McKendree, Singley, and Lewis showed the benefits of directly using a tutoring system with the intention of extending or testing Anderson's principles for the development of such tutoring systems. However, considering the contradictory empirical support against Anderson's third principle, these principles clearly need to be further tested and refined.

Other computer systems and feedback timing

Although many computer aided instruction machines and tutoring systems have been developed, only a few will be reviewed here. Each one addresses the issue of feedback timing a little differently and thus provides an interesting parallel to Anderson's work.

Skinner's linear programs

An early system by Skinner (1958) was basically a linear program in which students went through a set order of steps. Although Skinner never clearly described what constitutes a proper step, he did specify that a step must be large enough to advance the student's learning and small enough that the student can achieve it. Presumably a step was a piece of the learning material to be presented. Problems appeared occasionally to evaluate how well the learning material was understood. His system provided a student with immediate feedback when the student gave a correct answer. A student had to continue guessing until the right answer was achieved. The system's feedback was based on a learning theory that encouraged immediate reinforcement of correct responses. Skinner did not perform a controlled study comparing immediate feedback after correct responses with immediate feedback after incorrect responses or delayed feedback after either type of response.

SOPHIE

In Burton and Brown (1979) and Brown, Burton, and deKleer (1982), a coaching system is described which was built into the tutoring system. The idea was to provide friendly feedback when a student made an error and to allow the student to correct his own mistake. Their systems, SOPHIE I, II, and III, provided a limited coach that performed various checks and attempted to not overburden the student with criticism. Although no empirical tests were cited to compare the coach with an immediate feedback system, students who were exposed to the system responded to it favorably.

ALGEBRALAND

ALGEBRALAND, developed by Brown (1983a,1983b), was a problem solving environment for students learning algebra. The computer screen provided the means for a student to take the problem and explore different paths to a solution. Algebra operators were selected by the student but implemented by the system. Thus, students could not make an operation error. Another feature of ALGEBRALAND was that it allowed the student to explore alternate routes to solution without penalty and it allows the student to compare the different routes. Thus, along with supporting empirical evidence (Foss, 1987a,1987b), ALGEBRALAND provided the same minimal intervention that SOPHIE does.

Summary of other systems

These three systems provide a flavor of the research that has been done in this field but is by no means an exhaustive set. The purpose of the description has been to provide background material, to demonstrate that other psychological research in this field has been done, and to set the stage for the summary of the issues in the next section.

CHAPTER II

EXPERIMENTAL DESIGN: DELAYED VERSUS IMMEDIATE FEEDBACK IN A GENETICS TUTORING SYSTEM

This chapter begins with a description of the philosophy and goals of the project. This is followed by a detailed description of the genetics tutoring system developed to test the issue of delayed versus immediate feedback after errors. After the description, the experimental evaluation of the system will be presented.

Philosophy and goals

This experiment was designed to explore two aspects of feedback timing. First, a description of the feedback timing issues will be given. This is followed by a description of the issues covered and how they will be tested.

Feedback timing

Feedback timing can be separated into two distinct times: post-test and post-informative feedback. Post-test feedback time is the time interval after a student gives a response; while, post-informative feedback time is the time interval after a student has received feedback concerning the student's response.

In this experiment, post-informative feedback time was held constant, because this experiment tests Anderson's theory. Anderson never addressed the issue of post-informative feedback time. This issue may need to be addressed by future experiments because Levine's feedback theory and research by many others (see historical section) indicated that the post-informative feedback interval is in some cases more important than post-test interval.

In this experiment, post-test time will be either immediate or delayed. In the immediate condition, subjects will be told immediately whether or not their response is correct whenever a response is made. In the delayed condition, subjects will be told whether or not their response is correct only after all parts of a single problem are answered. Only Anderson's research has addressed feedback timing at this small grain of analysis (see Anderson section).

The bulk of the research cited in the historical section considered the immediate condition to be giving feedback after the problem was completed. Researchers gave feedback after a certain amount of time following the student's completion of a problem (0 - 30 seconds). Therefore, in previous studies, the time between response and feedback may have been a silent period of rehearsal. Thus, comparison to previous work including the work of Thorndike and Skinner from which Anderson's ideas derive, may not be accurate. In order to test Anderson's ideas, this experiment allows no delay in feedback after an error for the immediate condition or after the problem is completed for the delayed condition.

In summary, only the post-test interval will be examined with the understanding that post-informative feedback timing may actually be the variable of interest. The feedback timing will test immediate versus delayed timing where immediate feedback timing is the time after which any part of the problem has been completed and an error has occurred. Delayed feedback timing is the time after which the whole problem is completed.

Feedback timing issues in this experiment

Learning rate

One issue of this experiment is learning rate. Students in delayed feedback conditions sometimes take longer in the learning phase than those students in immediate feedback conditions (Lewis & Anderson, 1985; historical perspective section). Because a help system is provided, students who are lost in the delayed condition need can receive help on what to do next. But, a help system solves only one of the longer learning time problems. When the student must repeat the problem after each mistake, the immediate condition students always finish the learning process faster. Presumably the student could memorize the correct answer given in immediate feedback because the interval between the mistake and repetition of the problem is very small. In the delayed condition, the interval between the mistake, correction, and repetition of the problem may be very long. Thus, from the literature, one would predict longer learning times for delayed condition subjects and possible guessing for immediate condition subjects. By collecting beginning and ending times and keystroke times, both time to learn and no guessing time can be determined.

Content

A second issue is content. One problem with studying feedback timing is that timing is confounded with content. When feedback is delayed, the content of the feedback necessarily is different. In a sense, by delaying feedback to completion of a problem, the tutor has more time to determine what the student's misconceptions are and to give content for the overall problem rather than the local problem. On the other hand, immediate feedback does fix local errors. If the tutor tutors on more than the first mistake, the information conveyed to the student will be quantitatively different from that given to the immediate feedback condition students. The solution to this problem was to restrict the feedback content of the delayed feedback condition to the first error made.

Processing

A third issue concerns whether or not subjects "thought more" about the problem. The amount of attention paid to feedback can influence the amount learned in a task (Crutcher, Gagne, Anzelc, et. al, 1985). Subjects in the delayed condition can be encouraged to find alternative ways to deal with the problems. By doing work on other similar problems, subjects may comprehend more about the original problem (Craik & Lockhart, 1972).

Thinking more about the problem may be indicated in a variety of ways. First, keystroke latencies may be longer. More keystrokes may indicate guessing. Second, for the genetics problems used in this tutorial, thinking more may be indicated by a greater use of the Punnett Squares and, in particular, the use of the Punnett Squares to find the parental genotypes. A Punnett square is a symbolic representation of the genetic and/or phenotypic constitution of the next generation resulting from crossing the parental genotypes (see Figure 5, Hartl, Freifelder, & Snyder, 1988). A subject in the immediate condition can take a guess at the parents and get immediate feedback about whether the guess was correct. On the other hand, a subject in the delayed condition could not test a guess unless she chose to use the Punnett Square to see to test the guess.

	RY	Ry	rY	ry
RY	RRYY	RRYy	RrYy	RrYy
Ry	RRYy	RRyy	RrYy	Rryy
rY	RrYy	RrYy	rrYY	RrYy
ry	RrYy	Rryy	rrYy	rryy

Figure 5. Example Punnett Square.

<u>Retention</u>

A fourth issue for feedback timing is retention. Retention can be facilitated when interference is not present (as when an incorrect response does not interfere with a correct response) or when the items are so distinct that clear associations among stimuli and correct responses can be formed. Retention can be improved when a subject has time to rehearse. In addition, spending more time with the material may improve retention (Bloom, 1981). This argument is supported by depth of processing (Craik & Lockhart, 1972) and by the idea that forcing a person to generate the response helps retention (Crutcher & Healy, 1989; Foss, 1987a,1987b).

In most of the experiments on timing, students in the delayed condition may have taken longer in the learning phase, but these same students had better recall after some amount of time had passed. This experiment examined this issue by having students return on the following day (24 hour retention interval) for a test.

<u>Self-correction and error detection</u>

Related to the issue of retention is self-correction and error-detection. Studies with word processors found that subjects became lost when they did not have immediate feedback for incorrect actions (Mack, Lewis, & Carroll, 1983). Subjects in this experiment were not able to recognize incorrect actions. However, students who can correct their own mistakes, are likely to have better retention (Crutcher & Healy, 1989). The system provides a HELP system which provided the next step if the student was uncertain what to do. The problems are constrained by having a limited size of pedigree. The problems that come throughout the tutorial start out with very small pedigrees and advance to larger ones as the student gains skills. This system can be compared to the training wheels studies (Carroll & Carrithers, 1984).

Delayed condition subjects in Lewis and Anderson (1985) were better able to identify incorrect paths. Perhaps delay allows for learning in error-detection (or debugging in programming). Because the system also records all responses that a student makes, one could test whether students change incorrect responses and do not change correct ones. These changes may give an indication of self-correction and error-detection.

In summary, this experiment addressed feedback timing by examining performance during a learning day and testing day. Feedback timing will be studied by means of a computer genetics tutor/teacher. The next section describes the methodology of the experiment performed.

Genetics tutoring system

Problem domain and rationale for selection

Anderson (1982) claims that his learning theories should be applicable to diverse fields, including language processing and geometry. Psychologists and computer scientists have chosen mathematics, programming, and applied fields such as electronics and medicine for tutoring systems (Brown, 1983a; Soloway &

Johnson, 1984; Lesgold, Bonar, Ivill, & Bowen, 1987; Williams, Hollan, & Stevens, 1981). For this experiment, genetics was chosen to test Anderson's ideas.

The topic chosen was the genotype identification of individuals in a pedigree. The genotype is the genetic constitution of an individual; whereas, the phenotype is the observable properties of an organism (Hartl, et al. 1988). A pedigree is a hierarchical diagram depicting the phenotype of a group of related individuals. Identification of genotype for pedigrees was chosen because most students have had only a very limited exposure, if any, to this topic before college.

Punnett squares, which are used to identify possible genotypes of offspring of two individuals, were provided to subjects for use in the tutoring system. In a pilot study that was performed, students reported having had exposure to Punnett Squares in high school and were able to use them in tasks (Lee, 1989).

Relationship with Anderson's systems

This tutoring system was developed to be as similar as possible to Anderson's tutoring systems. Since Anderson's tutoring systems are based on production system modeling of student/expert behavior (Anderson, et. al, 1984), the genetics problems were first analyzed and their structure extracted. The analysis was used to create the tutorial and help in the choice of genetics problems for the tutorial.

The tutorial and genetics problems were tested in a pilot study prior to this experiment. [A description of the pilot study and the problems used in that study can be found in Lee (1989).] The details of the problem analysis, structure, and difficulty of the genetics problems used in the tutorial and written exam follow.

Genetics problems

In this experiment, subjects were taught how to solve pedigree problems on the first day and tested with a written exam on the second day. More detailed descriptions of the problems, the analyses performed and the problem difficulty can be found in Lee (1989).

Problem description

The problems used in the tutoring system and the written exam were standard introductory genetics problems. The problems were taken from quizzes given to introductory genetics students were solvable by those students during a typical class period. The experimenter sometimes simplified the problems by removing levels of the pedigrees. All problem answers were confirmed by graduate students in the Environmental, Organism, and Population Biology Department.

Sample question from the tutor is shown in Figure 6. The figure shows unaffected males and females (squares not filled in) and affected males and females (filled in squares). Question marks indicate where the subject was asked to identify the genotype of the individual. Usually, questions that appear in books or exams do not include the question mark. The question mark was used in the tutoring system so that earlier, easier questions which only required two or four identifications could be used in a larger pedigree. The larger pedigree assists in the identification of type of inheritance.

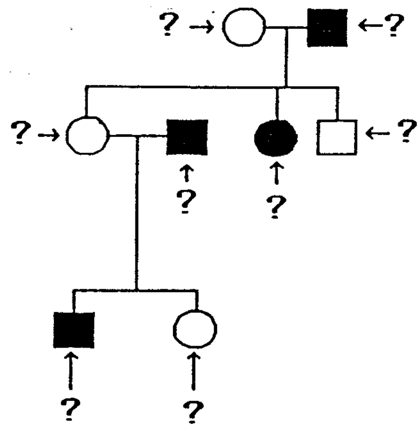


Figure 6. Example question from tutoring system, Question 4 of Problem set 4.

Figure 7 shows an example from the test. Subjects were asked to find the genotype for all of the individuals. The test question is similar to the question in the tutoring system except that question marks were not provided. In earlier questions for the test, letters were used to identify those individuals for which the subject was to find the genotype. The test included multiple choice questions which did not include pedigrees or the identification of type of inheritance or individual genotypes. The test materials will be discussed in more detail below.

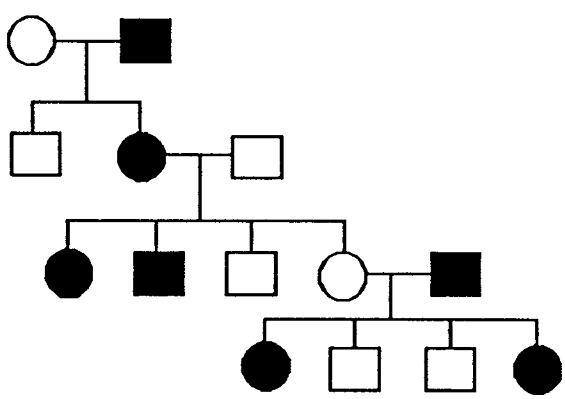


Figure 7. Example question from the exam.

Problem analysis

Anderson's tutoring systems often consist of an ideal student model and a bug catalog (Reiser, et al., 1985). An ideal student model includes domain knowledge and how to solve problems in that domain. The model does not solve problems the way an expert would solve them, because the methods that experts use often are different from those used by novices. A bug catalog is a listing of misconceptions and common mistakes made during problem solving. In order to create an ideal student model and a bug catalog, information about a domain and how students would solve problems in that domain must be collected.

A problem analysis was performed to collect data for the ideal student model and a bug catalog. This information was incorporated in the production system models. In addition, the information was used to develop the content of the feedback screens.

Structure

Anderson trained students on the structure of problems in a domain before giving the problems to the students. In addition, he provides visual information on problem structure as the students are doing the problems. In this same manner for this tutoring system, the problem structure was introduced before the actual problem solving. The diagram for a complete pedigree was provided during problem solving.

Problem difficulty in problems for tutor

Analysis showed that for a pedigree of three levels and one type of inheritance four basic steps were used to solve such problems. The tutoring system presented each step and provided problem sets which tutored on any previous steps. The largest constraint on the choice of problems for the tutor was time. The problems became progressively harder as the subject continued through the tutoring system.

Instructional materials

Material was gathered from several introductory genetics books. Several panels of introductory material were presented before the exact information on how to solve pedigree problems was given. These panels included information on Mendel and his research. The information on how to solve pedigree problems was generated through analysis of problems and procedures obtained and verified by two graduate students in biology who had previously taught introductory genetics courses.

Anderson's tutoring systems are designed to go along with a course of study and could be considered full courses in the material that they were to cover. This system does not tutor a full course in genetics.

Test materials

On the second day, subjects were given a test which consisted of a multiple choice section and problems to solve. The problems were a combination of problems taken from the tutoring system, problems taken from books, and a problem taken from an introductory genetics class. All problems were checked by a graduate student in biology except for the ones taken directly from genetics books that provided the answers.

Nine problems were given (see Appendix B). The first three problems came directly from the tutoring system. The first of these problems, the easiest problem in the test, came directly from the tutoring system problems and required the subject to perform only one step of the four possible steps in solving pedigree problems that was tutored (Appendix B). The next three problems were either taken from books, modified versions of problems from books, or simplified materials from an introductory genetics class. The next two problems were just genotype identification and did not include type of inheritance identification. The last problem came directly from a quiz given to an introductory genetics class. This problem consisted of a pedigree with two types of inheritance displayed. Subjects were asked to identify the type of inheritance and then to solve a problem about possible genotypes of a particular offspring. Thus, this problem, the most difficult problem in the test, not only required the subject to know how to do all four steps tutored, but also required the student to be able to extract the right information from a pedigree that had two types of inheritance displayed.

Two questionnaires were given to subjects on the testing day. The first questionnaire asked general information about subjects such as grade point average and numbers of math and science courses taken. The second questionnaire asked whether or not the subject found the tutoring system useful in solving problems, what the subject liked or disliked about the tutoring system, and what the subject would add.

Implementation

System description

The tutorial was developed in HyperCard for the Macintosh. Programming in HyperCard was done in HyperText, which allows the use of IF-THEN statements. A help system was provided because Anderson's tutoring systems included them. A help button appeared at the bottom of each problem screen. The subject could select that button and they would be put at the help information screen. This screen allowed the subject to select what type of help they needed. Subjects could find out the definitions for major genetics terms. Further, depending on what part of the problem the student was working on at the time of request, the help system was also able to provide the next step that the subject needed to take. The help system never provided the answer to any part of a problem.

The tutorial recorded subject number, date, time spent on each panel and responses to multiple choice questions (including whether or not the subject answered the question correctly). All answers to questions, including responses to the Punnett Square, are recorded.

Production system models. The genetics tutorial system used a production system model (IF-THEN statements) of student problem solving behavior. Anderson and colleagues (1984) and others have used this technique to model student behavior in tutoring systems (Brown & Van Lehn, 1980; O'Shea, 1979; Sleeman, 1982). In this genetics tutorial, the production system model was an ideal student model which included domain knowledge and knowledge of how students solve genetics problems. The tutor followed the student's flow of responses and identified the rule that the student was using with each action. When a student used a rule incorrectly, the system identified the problem and attempted to provide information to help the student correct the misconception.

Overview of training day

On the first day of training, subjects were seated at the tutoring system. They were asked to read each panel and answer all questions presented to them. The tutoring system presented informational text with intermittent multiple choice problems and four problem sets. Requiring students to make responses every few pages of a text facilitates learning (Hershberger, 1964; Hershberger & Terry, 1965a, 1965b). In addition to the multiple choice problems, problem sets presenting pedigree problems were also included throughout the tutor.

Selecting answers to pedigree problems. For those problems in which the subject had to identify the inheritance type depicted, a multiple choice section was provided. The members of the pedigree that the subject had to identify could be identified by looking at the arrow and question mark. When a subject selected a question mark with the mouse, a box appeared with the possible choices for that individual in the pedigree. For each part of each question, there are multiple wrong answers possible. Rather than presenting choices for which the choices could help the subject to eliminate choices in other sections of a problem, choices were generated by first listing all possible wrong answers for that problem and then selecting the choices randomly from that list. All subjects saw the same selection of choices for any one problem. After reading the possible choices, subjects were required to enter in a number. The choice then appeared in a box below or to one side of the question mark.

Repetition of problems. Throughout the experiment, if a student made a mistake, he/she was asked to repeat that problem until the problem could be answered correctly. This procedure was used to produce a baseline performance. Although, establishing a baseline has been controversial (Anderson, et al. 1986), errors have been known to persist despite feedback (Elley, 1966; Cunningham & Anderson, 1968). In addition, when Hively (1962, 1965) required subjects to repeat an exercise four times correctly in a row, improved performance was found on a post-test.

Feedback in the genetics tutoring system

Two versions of the tutorial were created. One tutorial gave feedback immediately after an error was made and the other gave feedback after one complete problem was finished. Aside from the feedback manipulation, the two tutorials were exactly the same. Feedback content was held constant.

Immediate condition feedback. Since feedback content was held constant for both groups, immediate condition subjects saw the same panels for the same errors as the delayed condition. However, immediate condition subjects saw the feedback panels immediate after an error was made and therefore had no opportunity to correct or change their answer in case of a slip. Thus, these subjects had the opportunity to see many more extra panels of feedback. Immediate condition subjects were placed back at the beginning of the problem after seeing the feedback screen, rather than being placed at the point where they made the error.

Delayed condition feedback. Subjects in the delayed condition could make errors and correct them. The system kept track of the order of errors and changes made as well, since it was tutoring on the first error made. If that error was corrected, then the subject was tutored on the second error. It is conceivable that a subject would correct all errors as they are made and the system would be required to provide no tutoring on any of those corrected errors.

In addition, for this condition the system could provide feedback on all mistakes made during execution of the problem and therefore provide the student with an explicit overall picture of their errors; however, this definition would not control for feedback content. Thus, this tutoring system would tutor the student on the first error made at the end of the problem unless that error was corrected. In that case, the system would tutor on the next error made and so on. In this way, the feedback content would be the same regardless to which condition the subject was assigned.

Method

This experiment was designed to test the tutoring system described in the preceding section. Whether the feedback was immediate or delayed was manipulated. The following sections describe the details of the evaluation.

Subjects

Twenty-two students who were enrolled in an introductory psychology course participated in the study for credit. Subjects had not previously had any genetics courses at the college level. The data from two subjects was dropped because they did not complete the first day of training. Subjects were randomly assigned to either the immediate or the delayed condition.

Procedure

The experiment was a two day experiment. The first day was the learning day in which subjects used the genetics tutoring system. Subjects returned the next day and were given a written test and two questionnaires.

Each participant was seated in front of a Macintosh Plus machine with harddrive. Subjects entered information using a keyboard and used the mouse to make selections. The system recorded time, date, keystrokes and mouse selections.

Subjects were asked whether or not they had used a Macintosh computer before. If the subject had not used a Macintosh previously, then the subject was given instructions on how to use the mouse and how to enter information.

Subjects were told that the experiment was designed to evaluate a genetics tutoring system.

Each subject saw the same basic material on genetics and the same problems. Four problem sets were shown which contained increasingly more difficult problems. The first set of problems consisted of one step in solving genetics problems; the second set contained the first and second steps for solving genetics problems; and each subsequent problem set contained the next step plus any previous steps until the full problem was given to be solved. Subjects were required to answer multiple choice questions which were interspersed throughout the text. Subjects were asked to answer correctly all multiple choice problems and problem set problems two times in a row before he/she was allowed to continue. The system automatically reset the problem by erasing previous responses before each iteration. All subjects were placed back at the beginning of the problem after reading the feedback screen.

The only difference between the two tutorials was in the timing of the feedback. For the immediate condition, subjects received feedback immediately after they typed in their responses. For the delayed condition, subjects received feedback after the problem was completed. The delay was not longer than a single problem.

On the second day, a test was given on the material learned. Subjects performed the test with paper and pencil and not on the computer. The test consisted of a multiple choice quiz, pedigree problems, a questionnaire on the tutoring system, and a questionnaire on their background.

CHAPTER III

RESULTS

This chapter will review the results of the empirical evaluation of the tutoring system as described in Chapter II. First, an overview of the variables involved will be presented. Then the data collected from the use of the tutoring systems and results of analyses performed on that data will be discussed. Finally, the chapter will end with a discussion of the test data and results.

Overview

The primary goal of this experiment was to investigate the effects of delaying feedback on errors in a tutoring system. Several factors were calculated from the information from both the training on the tutoring system and the test day that was gathered.

While subjects were using the tutoring system, the total time to complete each problem set and the whole tutorial were calculated. The number of keystrokes and the number of times that a subject saw the problem were recorded. Since questions 1, 3, and 5 of problem set 4 were not problems involving pedigree solving, they were not included in the analysis.

The learning data was analyzed to see whether subjects used Punnett squares, and whether subjects had used the Help system. For those subjects who used the Punnett square, the data were analyzed to see if the subject had used the Punnett square to find the parental genotypes. Use of the Punnett square was a possible indication that subjects were thinking about the problem.

In addition to Punnett square usage and HELP system usage, delayed condition subjects' data were scored for whether they had changed an incorrect answer to a correct one or visa versa. Since the system provided immediate subjects with feedback right after an error was made, immediate condition subjects had no opportunity to correct their responses. Changing correct answers to incorrect answers was hypothesized to be a potential problem for delay of feedback.

For the test data, a simple check for correct or incorrect was performed on each question for each subject. Each item of a pedigree that had to be identified was counted correct or incorrect. Correct responses were counted as 1 and incorrect responses were counted as 0. Since some pedigree problems have more than one part, a perfect score would be greater than 1. Question 6 was not corrected due to a typing error.

Results from the information gathered when students used the tutoring system will be discussed first. These training results will be followed by those from the test day.

Training day results

The following information will be presented in this order: timing data, number of times through each problem, keystroke data, use of HELP system, number and type of changes made, and use of Punnett Squares.

Time through tutoring system

Past studies (Lewis & Anderson, 1985; historical perspective section) have found that delaying feedback on errors produced a longer learning time for those subjects who received the delayed feedback. In this study, 80% of the subjects took two hours or less. A repeated measures ANOVA was performed on the timing data. The group by question interaction was significant, $F(9,180)=7.09$, $p<.01$. Delayed subjects took longer to complete the tutorial than immediate subjects, $F(1,19)=31.95$, $p<.01$ ($M_d = 2.12$; $M_i = 1.53$ hours)[1]. These results are consistent with the literature.

A graph of the time to do each problem in the tutoring system is shown in Figure 8. Individual T-tests were performed on the data points. The delayed condition subjects took longer on problem set 2, question 2, $F(1,19)=18.31$, $p<.01$, problem set 3, question 1, $F(1,19)=4.15$, $p<.057$, problem set 3, question 2, $F(1,19)=5.57$, $p<.03$, and problem set 4, question 2, $F(1,19)=16.84$, $p<.01$. However, no difference was found for problem set 4, question 4, $F(1,19)=1.42$, $p>.1$. Both groups of subjects showed an improvement between questions 2 and 4 of problem set 4, but the delayed condition subjects showed a greater improvement. These results indicate that the delayed condition subjects may have learned something about the problems in that time; whereas, the immediate condition subjects could have been guessing through the problem.

---

[1]M_d is mean for delayed condition; M_i is mean for immediate condition.

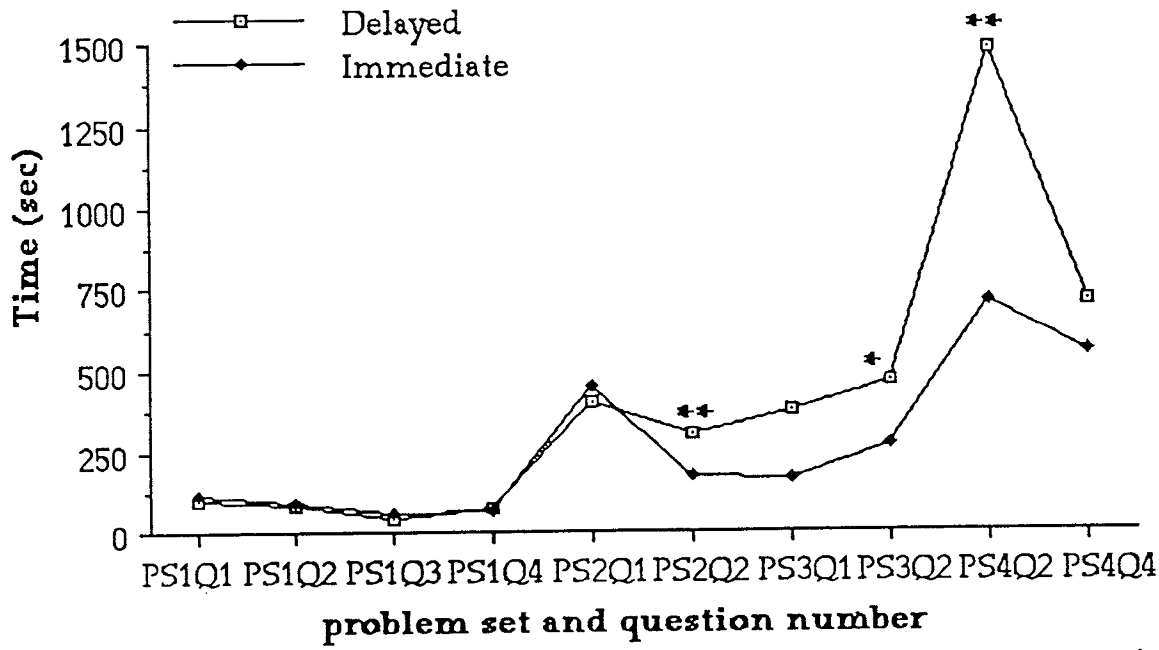


Figure 8. Time through each problem. * signifies that those two points are statistically significant at p<.05. ** signifies that those two points are statistically significant at p<.01.

Number of times through each problem

From the timing data, one hypothesis could be that subjects in the delayed condition were just doing the questions more times than the immediate condition subjects. The results from this experiment do not support this conclusion. The number of times through each question by group for the tutoring system are shown in Figure 9. A repeated measures ANOVA was performed on the cycle data. The group by question interaction was not significant, $F(9,180)=1.73$, $p<.1$. Individual T-tests were performed on the data points. The delayed condition subjects went through problem set 2, question 2 more times than the immediate condition subjects, $F(1,19)=4.07$, $p<.057$, but that is the only question that showed any difference between the two groups. Thus, test results can not be explained by saying that delayed condition subjects just saw the questions more times than the immediate condition subjects did.

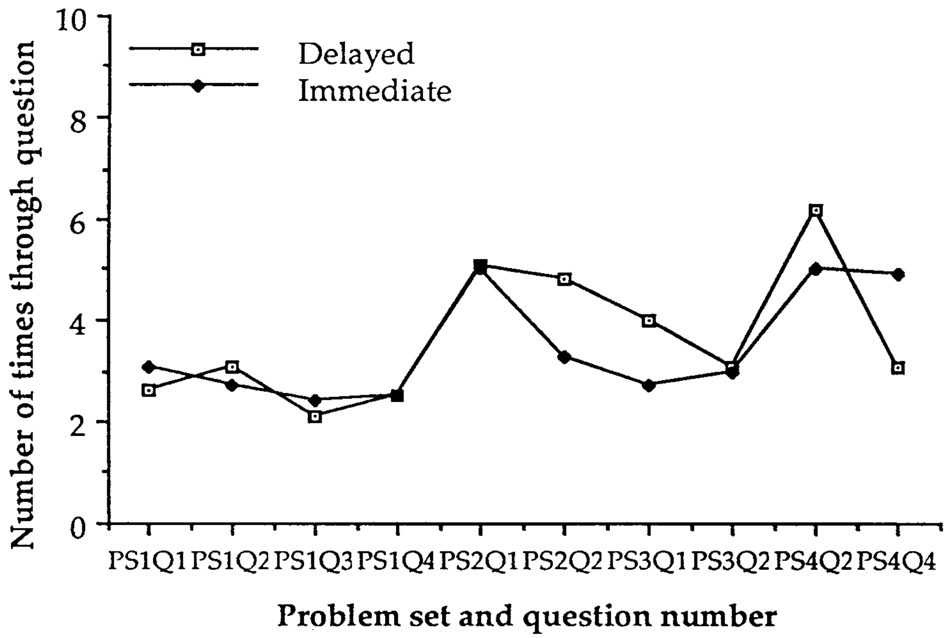


Figure 9. Number of times through each question by problem set.

Number of keystrokes

The number of keystrokes was recorded because it would give an indication as to whether or not guessing was occurring. Each entrance into a Punnett Square was counted as a keystroke. A repeated measures ANOVA was performed on the keystroke data. The group by question interaction was significant for the number of keystrokes per minute, $F(9,180)=2.35, p<.01$. Subjects in the immediate condition did have more keystrokes per minute than those in the delayed condition $F(1,19)=13.32$, $p<.01$ ($M_i=4.24$, $M_d=2.94$). A graph of the keystrokes per minute by question number is shown in Figure 10. Individual T-tests were performed on the data points. Immediate condition subjects took more keystrokes per minute for problem set 3, question 1, $F(1,19)=7.97$, and problem set 4, question 2, $F(1,19)=21.34$, $p<.01$. The total number of keystrokes by each group was not measurably different, $F(9,180)=1.10$, $p>.1$, but the total time for the immediate condition to complete the problems was faster than the delayed condition, $F(1,19)= 15.47$.

It is recognized that the amount of keystrokes per minute, the amount of time, and the number of keystrokes are not independent measures. However, given that the number of keystrokes is essentially equivalent between groups and the amount of time taken by the delayed condition is longer, the result that the immediate condition subjects had a greater amount of keystrokes per minute is not surprising. Thus, this result supports the hypothesis that the immediate condition subjects used some guessing during the problem solving.

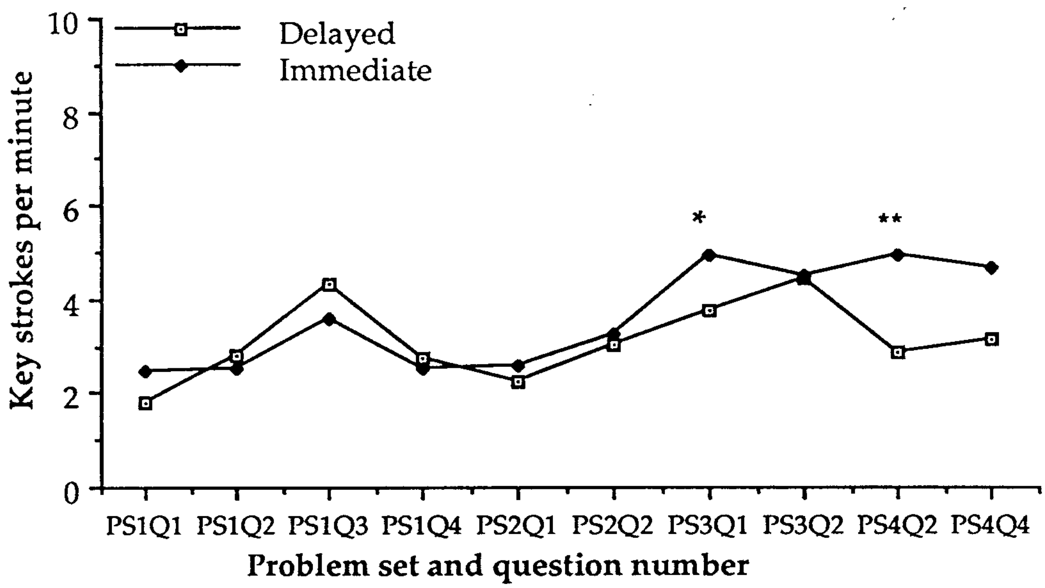


Figure 10. Keystrokes per minute by question. * signifies that those two points are statistically significant at $p<.05$. ** signifies that those two points are statistically significant at $p<.01$.

Using HELP system and Punnett squares

People in both groups used the Help system and there was no measurable difference between the groups in usage. People in both groups also used the Punnett squares, but of those subjects who used the Punnett squares, subjects in the delayed condition used the Punnett squares to find the parents more often than subjects in the immediate condition, $F(1,15)=4.47$, $p<.05$. Although both groups used the Punnett squares, these results indicate that the delayed subjects were using them for an additional function. This would lend support to the hypothesis that delayed subjects could think more about the problem.

Number and type of changes made

For the delayed subjects, a count was made as to how many times each subject changed a correct answer to an incorrect answer and an incorrect answer to a correct answer. Only delayed condition subjects' data was used because the system corrected immediate condition subjects right after an error occurred and thus immediate subjects had no opportunity to change their responses. Twenty percent of the subjects changed a correct answer to an incorrect answer and these subjects only made one such change; whereas, 70% of the subjects changed incorrect responses to correct ones and 20% of these made 4 or more such changes. Figure 11 shows a graph of the changes by question and problem set number. The number of changes of incorrect to correct responses increased steadily from problem set 3 question 1 to problem set 4 question 4; whereas, the number of changes of correct to incorrect responses stayed about the same throughout the use of the tutoring system. Thus, the anticipated problem of subjects randomly changing correct and incorrect answers was not found in this study.

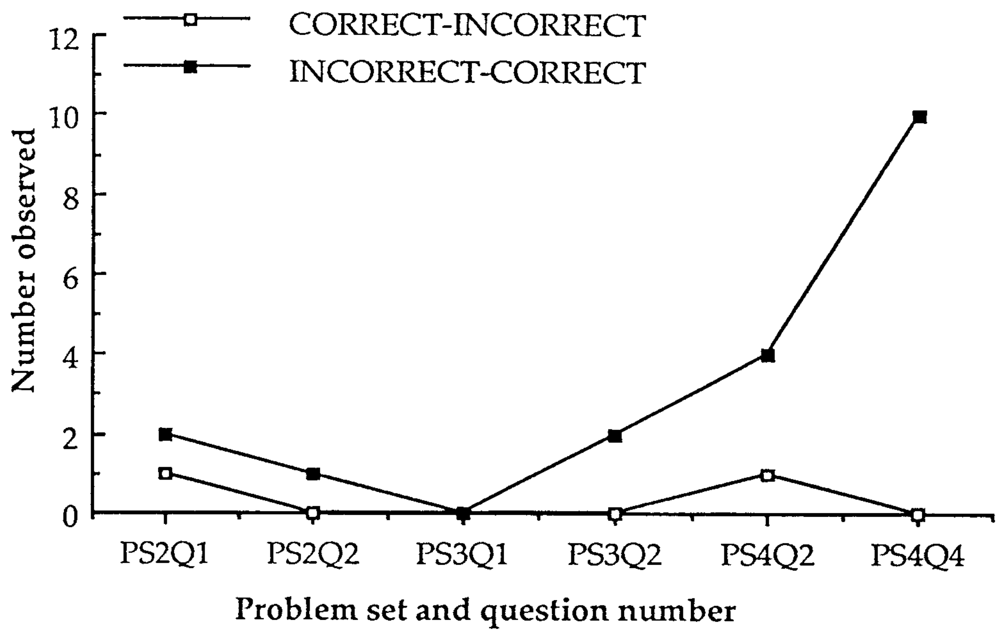


Figure 11. Number of changes of incorrect answers to correct and correct answers to incorrect by question

Test day results

The following information will be presented in this order: test results, basic information questionnaire, and questionnaire on tutoring system information.

Test results

There was no difference found in the time to complete the test (M_d = 39.6 min.; M_i = 36.8 min.). In addition, there was no difference found for the multiple choice questions and for the pedigree questions 1, 2, 4, 6, and 7. A repeated measures ANOVA was performed on the test data. No group by pedigree questioninteraction was found, $F(7,140)=1.52$, $p>.1$. However, delayed subjects gave more correct responses on the test as a whole than immediate subjects, $F(1,19)=9.13$, $p<.01$ (M_d = 60 points; M_i= 41 points). Individual T-tests were performed on the data points.

Delayed subjects performed better on question 3, $F(1,19)=8.44$, $p<.05$. For question 5 which was taken from a text book, delayed subjects did better than immediate subjects, $F(1,19)=5.3$, $p<.05$. Also, delayed subjects performed better on question 8, $F(1,19)=18.76$, $p<.01$. So, overall, the delayed condition subjects did better on the test problems than the immediate condition subjects. Figure 12 shows the percent correct as a function of test trial.

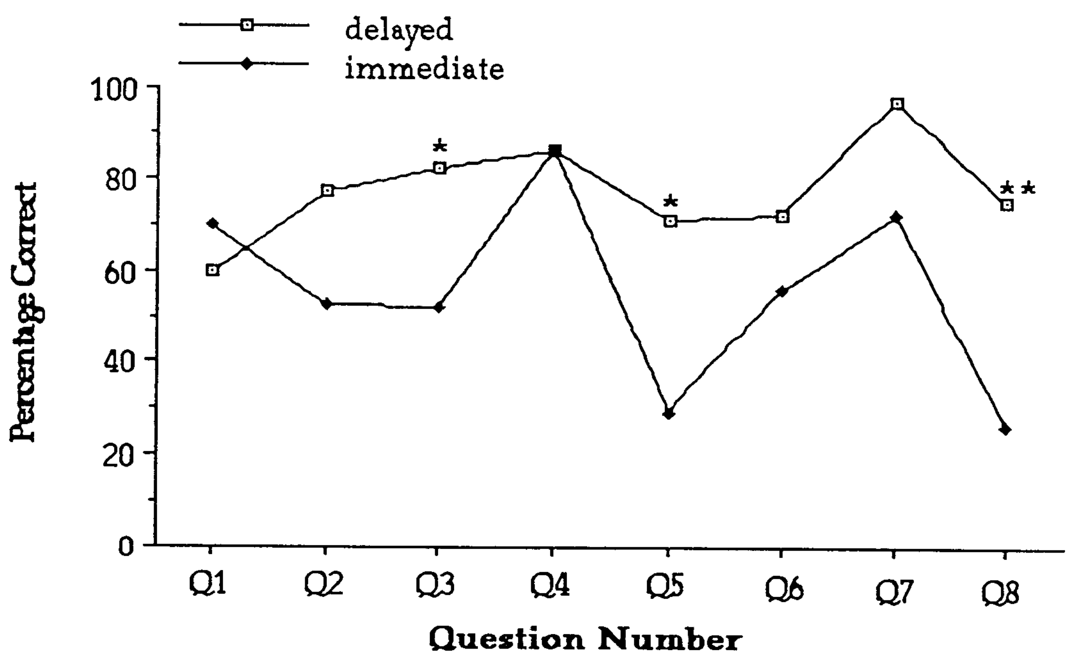


Figure 12. Test results by question and group. * signifies that those two points are statistically significant at p<.05. ** signifies that those two points are statistically significant at p<.01.

Questionnaire information

For the general questionnaire, the average number of college math classes taken was 2 with a standard deviation of 1 course; the maximum number of math classes was 5. The average number of science classes was 1 with a standard deviation of 2; the maximum number of science classes was 7. Fifty-five percent of the subjects took 1 or more of either sociology, anthropology, or archeology (social sciences). The average GPA was 2.98 with a standard deviation of .55. The GPA for the immediate condition was slightly better, $F(1,19)=3.57$, $p<.075$ ($M_i = 3.20$, $M_d= 2.76$); otherwise, there are no significant differences between conditions for number of courses taken in math, science, or social sciences. None of these factors had an affect on performance on the learning day or the test day ($F < 2.0$, $p > .2$).

In additional to the general questionnaire, a questionnaire about the tutoring system was given to each subject. For each of these questions, there were no differences in the responses given by members in each group. Generally students found the results useful, but they found having to repeat the question after having just gotten the question correct irritating. Students would have liked to have a break in the tutorial in addition to color, sound, more graphics and more examples.

CHAPTER IV

DISCUSSION

The results presented in the previous chapter and the following discussion are based on the fact that subjects were randomly assigned to groups. Further, no differences between groups could be found for grade point average, number of science courses or number of math courses.

This chapter will begin with an overview of the results. These results are then discussed in the context of the hypotheses given at the end of chapter 1. A summary of other issues, as well as the questionnaire results, will be covered. Finally, a brief summary of the issues discussed in this chapter and a conclusion section will be given.

Overview of results

Training results

The training day results indicate that the delayed condition subjects took longer on the tutorial than immediate condition subjects. With one exception (Grice, 1948), the instrumental learning studies on animals generally have shown a shorter learning time for animals in an immediate feedback situation.

On the other hand, studies with humans generally have been inconclusive. For example, for motor skill tasks of lever pulling and knob turning, no learning differences were found even when the delay varied from seconds to a week (Bilodeau & Bilodeau, 1958). But, in a discrimination learning study, Lipsitt and

Castaneda (1958) found that immediate condition subjects were more able to perform the discrimination. Lewis and Anderson (1985) presented a task to subjects which consisted of moving from room to room, finding treasure rooms, and making as many correct moves as possible. Subjects in the immediate condition learned the correct moves more quickly. Thus, the result that the immediate condition subjects took less time to learn than delayed condition subjects is consistent with the literature on animal studies and supports the hypothesis that in general human subjects take longer to learn when feedback is delayed.

Test results

The results from the test data indicate that the delayed condition subjects answered correctly more parts of more questions on the exam as a whole than immediate condition subjects. Although no differences between groups were found for questions 1, 2, 4, 6, and 7, delayed subjects performed better on questions 3, 5, and 8. Question 3 was taken directly from the tutoring system and a difference should not have been found. Question 5, on the other hand, was taken from an introductory genetics text book and question 8 was taken from an introductory genetics class quiz.

Although the results from the literature are consistent with the learning data obtained in this experiment, the results from the test data are contradictory to predictions made based on Anderson's prescription for tutoring systems that feedback should always occur immediately after a student gives an incorrect response (Anderson, et al. 1984). The content of the feedback panels can not account for the difference since subjects in the delayed condition saw the same panels that the immediate subjects did and the feedback for delayed condition subjects was restricted to the first error made.

The following sections will examine the test results in the context of the issues and hypotheses presented.

Discussion on hypotheses

Changing responses

The original hypothesis of this experiment was that subjects in the delayed condition would be able to have the opportunity to change their answers and thereby would learn how to identify errors. Results from this experiment indicated that delayed subjects do take advantage of making changes in their responses and these changes are often for the better. These results are similar to the results obtained by Lewis and Anderson (1985) in which subjects in the delayed condition were able to identify rooms which led to dead ends. Thus, subjects in this experiment probably learned how to identify errors and could also identify when they made an incorrect response on the test and correct it.

Repeating questions

A second hypothesis was that subjects in the delayed condition would need to repeat the questions more often than the immediate condition subjects and therefore, the delayed condition subjects would do better on the post test. The learning data results indicate that for only one question did the delayed condition subjects repeat the question more often. In general, there were no differences between the numbers of times that the subjects in either condition needed to repeat the questions. Thus, the number of times through questions in the tutoring system can not account for the test results.

Think more about questions

A third hypothesis was that subjects in the delayed condition would have more opportunity to think about their responses than immediate subjects. Punnett Square usage was counted to determine if subjects were thinking more about the problems. This experiment showed that there is a significant trend toward using the Punnett Squares more often and in a different way by the delayed condition subjects. However, subjects who were able to do the crosses in their heads would not get credit for such usage.

There is additional evidence to support the hypothesis. If one looks at the graph for the time to get through each problem in the tutoring system, the delayed condition subjects showed a great improvement in questions 2 and 4 of problem set 4, whereas, the immediate condition subjects did not improve very much here. Delayed condition subjects could therefore be thinking more about the problem and/or learning something about the first question that they carried over to the next one. Subjects could clearly have learned the recessive inheritance pattern in question 2 of problem set 4 and since sex-linked inheritance is often easier than autosomal (M.Y. Lee, personal communication, August 5, 1989), then question 4 should show an improvement in completion time.

In addition to the timing improvement between question 2 and 4, there is a steady increase in the number of changes of incorrect to correct responses by delayed subjects throughout the problem sets. Subjects are at least taking the time to notice their incorrect responses.

Guess more

A final hypothesis was that subjects in the immediate condition could do more guessing and memorizing of responses. The measure was number of keystrokes per minute and the results indicated that subjects in the immediate condition do take more keystrokes per minute than subjects in the delayed condition. Therefore, guessing can not be overruled.

Summary of hypotheses

In summary, the results confirmed most of the hypotheses set forth. Subjects in the delayed condition benefitted from changing incorrect responses to correct responses. Further, delayed condition subjects appeared to think more about the problems and immediate condition subjects appeared to do more guessing. It is possible that some combination of guessing and thinking more about the questions in the problem sets is taking place.

The results are consistent with the idea that the presentation of material in tutorials is extremely important. In a sense, since subjects received small steps which built toward the full problem, delayed subjects were not presented with a discovery learning context and therefore may have actually integrated their knowledge into useful structures that could be utilized when the more difficult problems were presented.

Other issues

The topic of study, may be a factor in these results. Looking at the results from several different experiments (summarized in the introduction), feedback timing appears to depend on topic of study. Considering that Lewis and Anderson

(1985) studied immediate versus delayed in the context of a gaming environment, their results may not be directly applicable to any other field except gaming. However, Lewis (in press) presents results of experiments with the algebra tutoring system which demonstrates the effectiveness of that tutoring system even with a one response delay of feedback. The analysis presented of the difference between the algebra and geometry tutoring systems indicated that immediate feedback may only be useful in domains where the student must be constrained to a path (Lewis, in press). Some future experiment comparing directly two tutoring systems in different domains, one in which a constrained path is necessary and one in which a constrained path is not, and delayed or immediate feedback may be helpful in resolving this issue.

Summary of discussion

That delayed subjects performed better on a post test was surprising. Several reasons are possible for these results. In this experiment, the presentation of the material may be the primary reason. The problems in the tutorial were presented in a carefully graded sequence. If students had been presented with a whole pedigree problem in the beginning, the delayed condition subjects might not have finished the tutorial. As with Skinner's machines, the tutorial's beginning sequence of graded material helped the students to elicit the correct responses (Skinner, 1958). This tutorial followed the guidelines set out in Anderson et al. (1984) for tutoring small steps at a time. In this case, one of the guidelines actually helped the experiment achieve results. This contradicts one of his other guidelines that immediate feedback is always preferable.

The other major reason that delayed subjects performed better on the test is that during the tutorial immediate condition subjects may have been guessing.

Feedback timing may have interacted with the desire to finish the experiment. On the other hand, the tutoring system could have been forcing subjects in the delayed condition to generate the correct responses because they do not receive feedback until the end of a question. Both the generation effect (Crutcher & Healy, 1989) and depth of processing (Craik & Lockhart, 1972) have been shown to affect learning.

Despite these results, how quickly students complete a tutorial is a factor. Students may not have the ability to do a tutorial lasting many hours. Furthermore, students with absolutely no experience with a topic and are given a complete problem in the beginning may need to be told what to do and may find it difficult to figure out how to get done. Thus, past studies and guidelines still have validity since immediate feedback does produce faster learning sessions.

This study along with others indicate that more variables may be involved. For example, feedback content may be crucial. If feedback content is too complete, the student will not have the opportunity to synthesize the correct answer. However, if one follows Anderson's guidelines, the feedback content is based on the failed step. So, the problem is that a system must provide feedback which is complete enough to be helpful and incomplete enough to force the student to generate the correct answer. Thus, more support is needed for further development of tutoring system guidelines and the study of feedback in other systems.

Conclusion

Tutoring system research should be viewed as a cycle with the testing of psychological theories feeding into the development of systems which also, through field tests, contribute to the adjustment of those theories. Although Anderson's ideas of tutoring systems state that immediate feedback should be used over delayed feedback (Anderson, et al. 1984), the results from this experiment indicate that delay of feedback may have some advantages. Given the results of this experiment and Lewis (in press), one can not say that one type of feedback timing is appropriate for all situations. The characteristics of each situation must be considered. One needs to weigh such factors as whether a student will become lost without immediate feedback with whether a student will become oversaturated by having the system continually giving feedback. Harley and Sleeman (1973) stated that evaluative studies are needed to determine when help should be given to a student. The question may have two parts: 1) when should we give a student help and 2) when should we allow the student the freedom to explore alternatives.

Acknowledgements

I would like to thank Dr. Peter Polson for his help on this project. Also thanks to the members of my committee, Dr. Jane Bock, Dr. Clayton Lewis, and Dr. Nancy Pennington.

Special thanks to Michele Lee and Peter Foltz. Without their continuous and enthusiastic help, this project would not have been successful. In addition, I'd like to thank Jan Gehring, Philip Thompson, Keith Hannon, Susan Davies, Paula Messamer, Dr. Jean McKendree, and Dr. Matthew Lewis for their help. Finally, I'd like to thank my parents for their continuing emotional support. Their encouragement and dedication to education has inspired my sister and me to strive to be better than we might have been otherwise.

CHAPTER V

REFERENCES

Anderson, J.R. (1976). Language, memory, and thought. Hillsdale, NJ: Erlbaum.

Anderson, J.R. (1981). Tuning of search of the problem space for geometry proofs. In the Proceedings of the Seventh International Joint Conference on Artificial Intelligence, 165-170.

Anderson, J.R. (1982). Acquisition of cognitive skill, Psychological Review, 89(4), 369-406.

Anderson, J.R. (1983). The Architecture of Cognition. Cambridge, MA: Harvard University Press.

Anderson, J.R. (1987). Methodologies for studying human knowledge. Behavioral and Brain Sciences, 10, 467-505.

Anderson, J.R., Boyle, C.F. & Reiser, B.J. (1985a). Intelligent tutoring systems. Science, 228(4698), 456-462.

Anderson, J.R., Boyle, C.F., Farrell, R. & Reiser, B. (1984). Cognitive principles in the design of computer tutors. In Sixth Annual Conference of the Cognitive Science Program (pp. 2-16). Boulder, CO: Institute of Cognitive Science, University of Colorado.

Anderson, J.R., Boyle, C.F., & Yost, G. (1985b). The geometry tutor. In the Proceedings of the International Joint Conference on Artificial Intelligence, 106-109.

Anderson, J.R., Boyle, C.F., Corbett, A. & Lewis, M. (1986). Cognitive Modelling and Intelligent Tutoring (Technical Report No. ONR-86-91). Pittsburgh, PA: Carnegie-Mellon University,

Anderson, J.R., Farrell, R. & Sauers, R. (1984). Learning to program in LISP. Cognitive Science, 8, 87-129.

Anderson, J.R. & Jeffries, R. (1985). Novice LISP errors: Undetected losses of information from working memory. Human-Computer Interaction, 1, 107-131.

Anderson, R.C. & Kulhavy, R.W. (1972) Learning concepts from definitions, American Educational Research Journal, 9, 385-390.

Anzai, Y. & Simon, H.A. (1979). The theory of learning by doing, Psychological Review, 86(2), 124-140.

Bardwell, R. (1981). Feedback: How does it function. Journal of Experimental Education, 50, 4-9.

Bilodeau, E.A. & Bilodeau, I McD., (1958). Variation of temporal intervals among critical events in five studies of knowledge of results. Journal of Experimental Psychology, 55, 603-612.

Bloom, B.S. (1981). All Our Children: A Primer for Parents, Teachers, and Other Educators. New York: McGraw-Hill.

Bourne, L.E. (1957). Effects of delay of informative feedback and task complexity on the identification of concepts. Journal of Experimental Psychology, 54, 201-207.

Bourne, L.E. & Bunderson, C.V. (1963). Effects of delay of informative feedback and length of postfeedback interval on concept identification. Journal of Experimental Psychology, 65,1-5.

Brown, J.S. (1983a). Learning-by-doing revisited for electronic learning environments. In M.A. White (Ed), The Future of Electronic Learning. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, J.S. (1983b). Process versus product: a perspective on tools foPr communal and informal electronic learning. In Report from the Learning Lab: Education in the Electronic Age, Educational Broadcasting Corporation. (Reprinted in Journal of Educational Computing Research, 1985, 1, 179-201.)

Brown, J.S. (1985). Idea amplifiers: new kinds of electronic learning. Educational Horizons, 63, 108-112.

Brown, J.S., Burton, R.R. & deKleer, J. (1982). Pedagogical, natural language and knowledge engineering techniques in SOPHIE I, II, and III. In D. Sleeman and J.S. Brown (Eds.), Intelligent Tutoring Systems, New York: Academic Press.

Brown, J.S., Collins, A., & Duguid, P. (1988). Situated cognition and the culture of learning (Technical Report No. 6886). Cambridge, MA: Bolt, Beranek, and Newman, Inc..

Brown, J.S. & Van Lehn, K. (1980). Repair theory: A generative theory of bugs in procedural skills. Cognitive Science, 4, 379-426.

Burton, R.R. & Brown, J.S. (1979). Toward a natural language capability for computer-assisted instruction. In H. O'Neil (Ed), Procedures for instructional Systems Development, New York: Academic Press.

Carroll, J.M. & Carrithers, C. (1984). Training wheels in a user interface, Communications of the ACM, 27 (8), 800-806.

Carroll, J.M. & McKendree, J. (1987). Interface design issues for advice-giving expert systems. Communications of the ACM, 30(1), 14-31.

Collins, A. (1987). Reformulating testing to measure learning and thinking (Technical Report No. 6898). Cambridge, MA: Bolt, Beranek, and Newman, Inc..

Collins, A. (1988). Cognitive apprenticeship and instructional technology (Technical Report No. 6899). Cambridge, MA: Bolt, Beranek, and Newman, Inc..

Collins, A. & Brown, J.S. (1987). The computer as a tool for learning through reflection. In H. Mandl and A. M. Lesgold (Eds), Learning Issues for Intelligent Tutoring Systems, New York: Springer-Verlag.

Craik, F.I.M. & Lockhart, R.S. (1975). The role of rehearsal in short-term memory. Journal of Verbal Learning and Verbal Behavior, 12, 671-684.

Crutcher, R. J., Gagne, E.D., Anzelc, J.M., Lizcano, L., Geisman, C.A., Shutz, P.A., & Hoffman, V. (1985). Students' strategies for using teacher feedback. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Crutcher, R.J. & Healy, A.F. (1989). Cognitive operations and the generation effect. Journal of Experimental Psychology: Learning, Motivation, and Cognition, 15(4), 669-675.

Cunningham, D.J. & Anderson, R.C. (1968). Effects of practice time within prompting and confirmation presentation procedures on paired-associate learning. Journal of Verbal learning and Verbal Behavior, 7, 613-616.

Estes, W.K. (1962). Learning theory. Annual Review of Psychology, 13, 107-145.

Estes, W.K. (1969). Reinforcement in human learning. In J. Trapp (Ed), Reinforcement and Behavior, New York: Academic Press.

Farrell, R.G., Anderson, J.R., & Reiser, B.J. (1985). An interactive computer-based tutorial for LISP. In the Proceedings of the International Joint Conference on Artificial Intelligence, Los Angeles, p. 106-109.

Foss, C.L. (1987a). Learning from errors in ALGEBRALAND (Technical Report IRL-87-0003). Institute for Research on Learning, Palo Alto, CA.

Foss, C.L. (1987b). Memory for self-derived solutions to problems (Technical Report IRL-87-0002). Institute for Research on Learning, Palo Alto, CA.

Fry, E. (1963). Teaching Machines and Programmed Instruction. New York: McGraw-Hill.

Gagne, R.M. & Rohwer, W.D. (1969). Instructional psychology. Annual Review of Psychology, 20, 381-418.

Gaynor, P. (1981). The effect of feedback delay on retention of computer-based mathematical material. Journal of Computer-Based Instruction, 8, 28-34.

Grice, G.R. (1948). The relation of secondary reinforcement to delayed reward in visual discrimination learning. Journal of Experimental Psychology, 38, 1-16.

Harley, J.R. & Sleeman, D.H. (1973). Towards intelligent teaching systems. International Journal of Man-Machine Studies, 5, 215-236.

Hartl, D.L., Freifelder, D., & Snyder, L.A. (1988). Basic Genetics. Boston, MA: Jones and Bartlett Publishers.

Hershberger, W. (1964). Self-evaluational responding and typographical cueing: Techniques for programming self-instructional reading materials. Journal of Educational Psychology, 55, 288-196.

Hershberger, W. & Terry, D.F. (1965a). Delay of self-testing in three types of programmed text. Journal of Educational Psychology, 56, 22-30.

Hershberger, W. & Terry, D.F. (1965b). Typographical cuing in conventional and programmed texts. Journal of Applied Psychology, 49, 55-60.

Hively, W. (1962). Programming stimuli in matching to sample. Journal of Experimental Analysis and Behavior, 5, 279-298.

Hively, W. (1965). Parametric experiments with a matching-to-sample program. In J. Holland and B.F. Skinner (Eds), An Analysis of the Behavioral Processes Involved in Self-Instruction with teaching Machines, Final Report, USOE, NDEA Title VII, Project No. 191.

James, W. (1950). The Principles of Psychology. New York: Dover (originally published in 1890).

Johnson, D.M. & Stratton, R.P. (1966). Evaluation of five methods of teaching concepts. Journal of Educational Psychology, 57, 48-53.

Kimball, R. (1982). A self-improving tutor for symbolic integration. In D. Sleeman and J.S. Brown (Eds.), Intelligent Tutoring Systems, London: Academic Press.

Kintsch, W. & McCoy, D.F. (1964). Delay of informative feedback in paired-associate learning. Journal of Experimental Psychology, 68, 372-375.

Klausmeier, H.J. (1976). Instructional design and the teaching of concepts. In J.R. Levin and V.L. Allen (Eds.), Cognitive Learning in Children. New York: Academic Press.

Kulhavy, R.W. & Anderson, R.C. (1972). Delay-retention effect with multiple-choice tests. Journal of Experimental Psychology, 63, 505-512.

Lee, A. Y. (1989). Timing of feedback in tutoring systems. Unpublished masters thesis. Department of Psychology, University of Colorado, Boulder, CO.

Lee, A. Y., Polson, P.G., & Bailey, W.B. (1989). Learning and transfer of measurement tasks. In Ken Bice and Clayton Lewis (Eds.), CHI'89 Conference Proceedings: Human Factors in Computing Systems, (pp.115-120). NY: Association for Computing Machinery.

Lesgold, A.M., Bonar, J.G., Ivill, J.M. & Bowen, A. (1987). An intelligent tutoring system for electronics troubleshooting: DC-circuit understanding. Technical Report. Learning Research and Development Center, University of Pittsburgh, Pittsburgh, P.A. (To appear in L.B. Resnick (Ed), Knowing and Learning: Issues for the Cognitive Psychology of Instruction. Hillsdale, NJ: Lawrence Erlbaum Associates.)

Lewis, M.W. (in press). Developing and evaluating the CMU algebra tutor: Tension between theoretically and pragmatically driven design.

Lewis, M.W. & Anderson, J.R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. Cognitive Psychology, 17, 26-65.

Lewis, M.W., Milson, R. & Anderson, J.R. (1987). The TEACHER'S APPRENTICE: Designing an intelligent authoring system for high school mathematics. In G. Kearsley (Ed.), Artificial Intelligence and Instruction: Applications and Methods, Reading, MA: Addison-Wesley Publishing Company.

Lipsitt, L.P. & Castaneda, A. (1958). Effects of delayed reward on choice behavior and response speeds in children. Journal of Comparative and Physiological Psychology, 51, 65-67.

McKendree, J. (1986). Impact of feedback content during complex skill acquisition, Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh.

Mack, R.L, Lewis, C.H., & Carroll, J.M. (1983). Learning to use a word processor: Problems and prospects. ACM Transactions on Office Information Systems, 1(3): 254-271.

Newell, A. & Simon, H. (1972). Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

O'Shea, T. (1979). A self-proving quadratic tutor. International Journal of Man-Machine Studies, 11(1), 97-124.

Rankin, R.J. & Trepper, T. (1978). Retention and delay of feedback in a computer-assisted-instructional task. Journal of Experimental Education, 46(4), 67-70.

Reiser, B.J., Anderson, J.R., & Farrell, R.G. (1985). Dynamic student modelling in an intelligent tutor for LISP programming. In the Proceedings of the International Joint Conference on Artificial Intelligence, Los Angeles, 106-109.

Roper, A.L. (1977). Feedback in computer-assisted instruction. Programmed Learning and Educational Technology, 14, 43-49.

Singley, K. (1987). Developing models of skill acquisition in the context of intelligent tutoring systems. Unpublished doctoral dissertation, Carnegie-Mellon University, Pittsburgh.

Singley, K. (in press) Improving operator selection through goal posting in a calculus word problem tutor. To appear in Journal of Artificial Intelligence and Education.

Skinner, B.F. (1954). The science of learning and the art of teaching. Harvard Educational Review, 24, 86-97.

Skinner, B.F. (1958). Teaching machines. Science, 128, 889-977.

Sleeman, D. (1982). Assessing aspects of competence in basic algebra. In D. Sleeman and J.S. Brown (Eds), Intelligent Tutoring Systems. New York: Academic Press.

Sleeman, D. & Brown, J.S. (1982). Intelligent Tutoring Systems. London: Academic Press.

Suppes, P. (1967). Some theoretical models for mathematics learning. Journal of Research and Development in Education, 1, 5-22.

Tennyson, R.D. (1971). Instructional variables which predict specific learner concept acquisition and errors. Paper presented at the annual meeting of the American Psychological Association.

Tennyson, R.D. (1973). Effect of negative instances in concept acquisition using a verbal learning task. Journal of Educational Psychology, 64, 247-260.

Thorndike, E.L. (1911). Animal Intelligence. New York: Macmillan.

Thorndike, E.L. (1913). Educational Psychology: The Psychology of Learning. New York: Teacher's College.

Thorndike, E.L. (1935). The Psychology of Wants, Interests and Attitudes. New York: Appleton-Century-Crofts.

Thorndike, E.L. (1949). Selected Writings From a Connectionist's Psychology. New York: Appleton-Century-Crofts.

Tulving, E. (1982). Synergistic ecphony in recall and recognition. Canadian Journal of Psychology, 36, 130-147.

Tulving, E. (1983). Elements of Episodic Memory. London: Oxford University Press.

Tulving, E. & Thompson, D.N. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

Uhr, L. (1969). Teaching machine programs that generate problems as a function of interaction with students. In the Proceedings of the 24th National ACM Conference, (pp. 125-134). New York: Association for Computing Machinery.

Wenger, E. (1987). Artificial Intelligence and Tutoring Systems. Los Altos, CA: Morgan Kaufmann Publishers, Inc.

Woods, P. & Hartley, J.R. (1971). Some learning models for arithmetic tasks and their use in computer-based learning. British Journal of Educational Psychology, 41(1), 35-48.

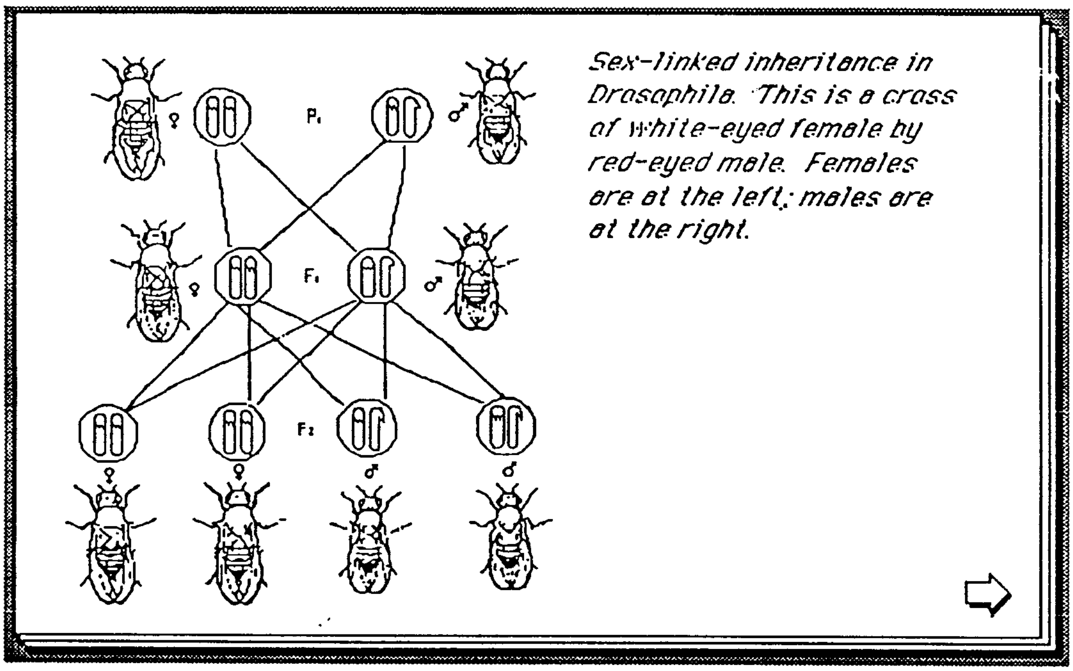
Yazdani, M. (1987). Intelligent tutoring systems: An overview. In R.W. Lawler and M. Yazdani (Eds.), Artificial Intelligence and Education, vol. 1, Norwood, NJ: Ablex Publishers.

APPENDIX A.

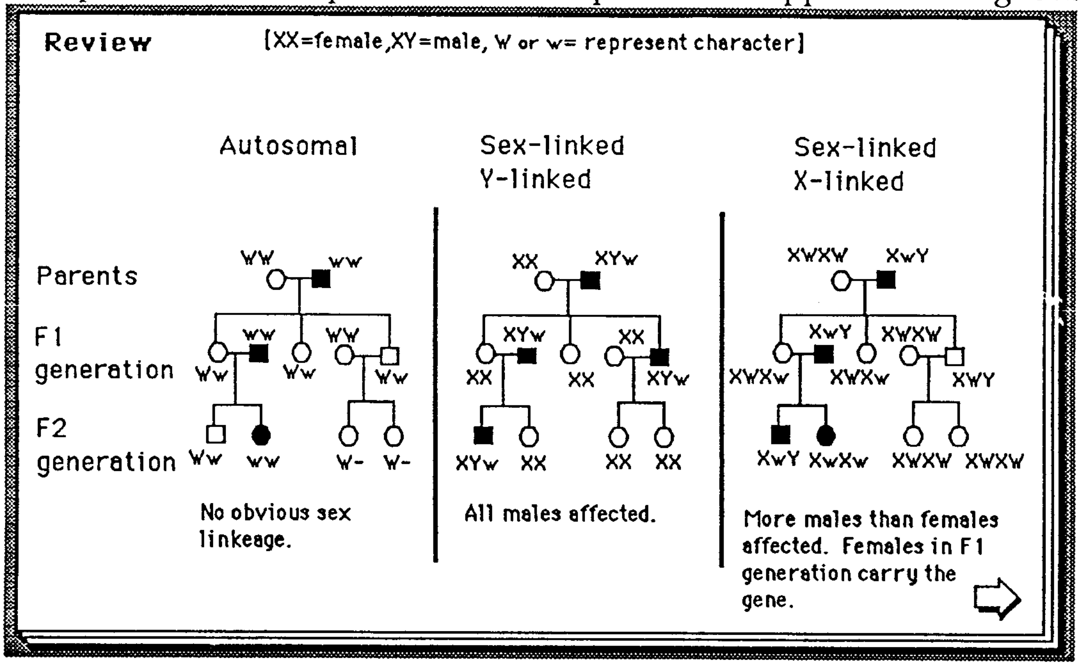
GENETICS TUTORIAL

The following pages contain excerpts from the tutorial.

This panel is an example of the type of graphics that can be used in a Hypercard environment.

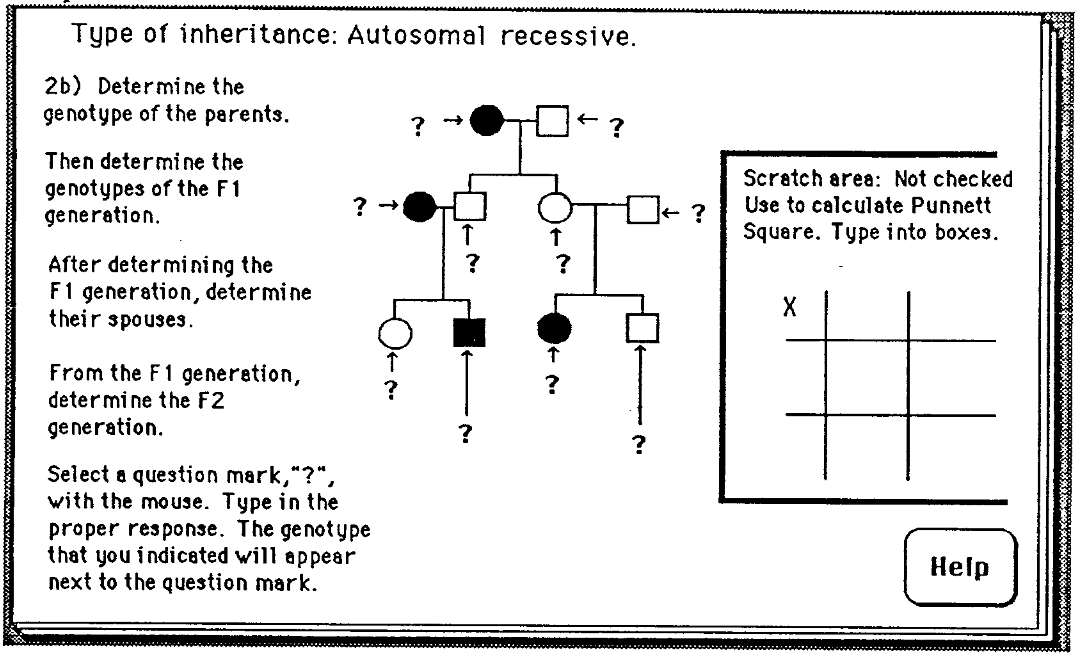


This panel is an example of the review panels that appeared throughout.



The following sections will show selected panels from the problem sets. Each question in the problem sets requires the subjects to select items and/or to type information. The system shows feedback panels after a mistake is made or tells the subject that the question was answered correctly. Subjects were asked to do correctly each question two times in a row. By the time the subjects encountered the problem sets, they had already had practice making selections with the mouse and typing in responses. Information about how to use Punnett Squares was given during the course of the tutorial, but no explicit problems were provided. Each problem set had its own HELP provided. Examples of the HELP panels and further problem set panels can be found in Lee (1989) for more details.

This panel from Problem Set 4 illustrates one of the most difficult problems from the problem sets.



APPENDIX B.

PROBLEMS FOR POST-TEST

Subject No.

Test: As you are doing the problems, use all blank spaces and even the backs of the pages as scratch paper. You can ask the experimenter if you need more scratch paper.

1. The foundations for the study of heredity were first laid by
   - a) Morgan
   - b) Mendel
   - c) Mendeleff
   - d) Muller
   - e) De Vries

2. Variation in organisms is due to
   - a) inheritance of genes
   - b) environmental pressures
   - c) both heredity and environment

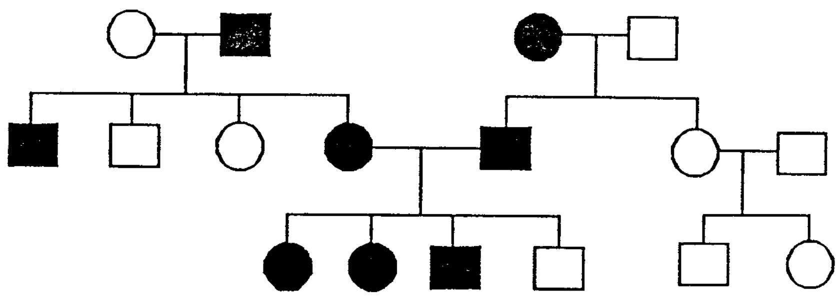
3. A man with a certain disease marries a normal woman; all of the girls have their father's disease, but none of the boys do. What type of inheritance is suggested?
   - a) autosomal recessive
   - b) autosomal dominant
   - c) y linked
   - d) x-linked dominant
   - e) x-linked recessive

4. A color-blind man marries a normal, homozygous woman. The chances of their having a color-blind son are
   - a) 25-75%
   - b) 50%
   - c) 25-50%
   - d) 100%
   - e) 0

5. A son with hemophilia will most likely result from parents represented as:
(X or Y = normal, "h" next to a letter indicates that the hemophilia gene is carried on that chromosome)

a) XX x XhY
b) XX x XhYh
c) XX x XY
d) XXh x XY

Question 1



What type of inheritance is this?

a) autosomal dominant

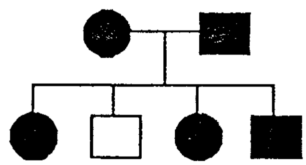
b) autosomal recessive

c) x-linked, dominant

d) x-linked, recessive

e) y-linked

Question 2



What type of inheritance is this?

a) autosomal dominant

b) autosomal recessive

c) x-linked, dominant

d) x-linked, recessive

e) y-linked

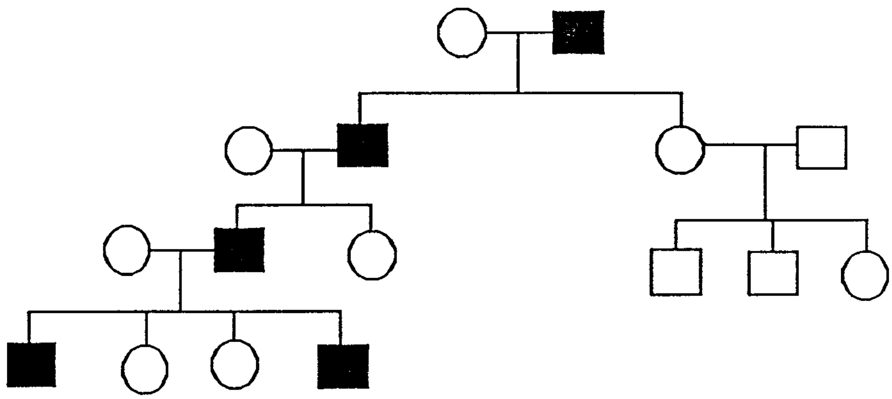
What are the genotypes of the lettered individuals?

The pedigree represents a family of gnomes with the disease. Gnomes either have "base" or they have "arches". Arches is the recessive;therefore, the disease is represented by an "a". Use "a" in your answers the questions about the pedigree.

a)

b)

Question 3



What type of inheritance is this?

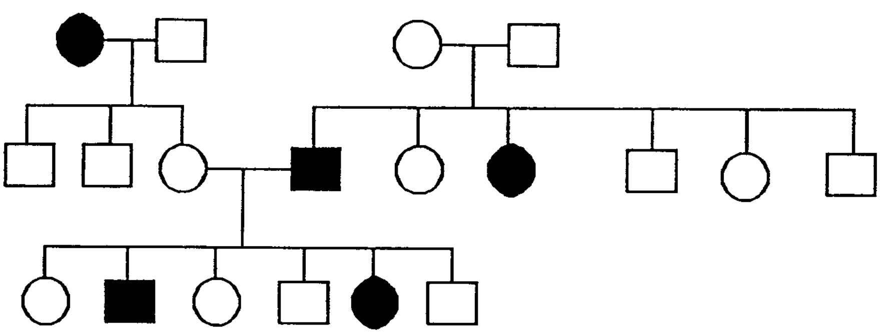
a) autosomal dominant
b) autosomal recessive
c) x-linked, dominant
d) x-linked, recessive
e) y-linked

What are the genotypes of the lettered individuals?

The pedigree represents a family of with a certain type of color-blindness. This type of color-blindness is represented by a "c". Use "c" in your answers this question about the pedigree.

a)
b)
c)
d)

Question 4



What type of inheritance is this?

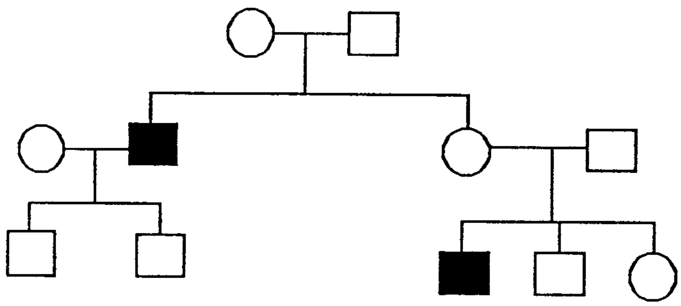
a) autosomal dominant
b) autosomal recessive
c) x-linked, dominant
d) x-linked, recessive
e) y-linked

What are the genotypes of the lettered individuals?

The pedigree represents a family of with the disease, hemophilia. Assuming that Hemophilia is the recessive expression, then the disease is represented by an "h". Use "h" in your answers this question about the pedigree.

a)
b)
c)
d)
e)
f)
g)
h)

Question 5



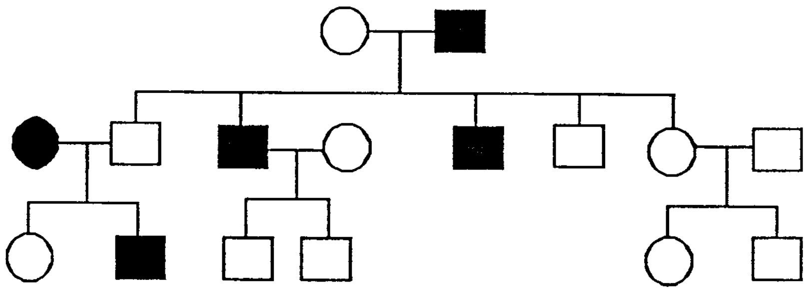
What type of inheritance is this?

a) autosomal dominant
b) autosomal recessive
c) x-linked, dominant
d) x-linked, recessive
e) y-linked

What are the genotypes of **ALL** individuals?

Write your answers directly onto the pedigree above.

Question 6



What type of inheritance is this?

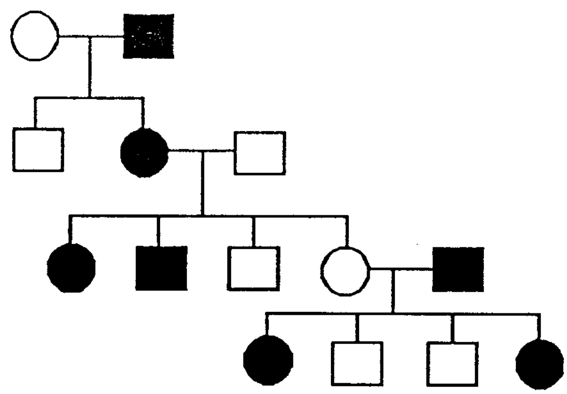
a) autosomal dominant
b) autosomal recessive
c) x-linked, dominant
d) x-linked, recessive
e) y-linked

What are the genotypes of ALL individuals?

Write your answers directly onto the pedigree above.

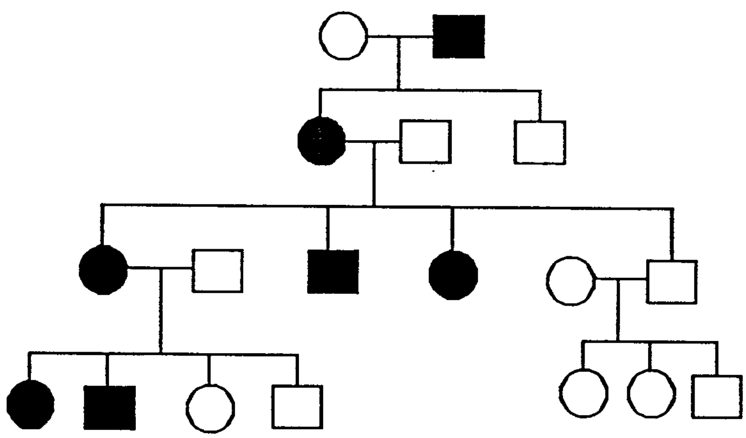
Question 7

The following pedigree represents a dweeble family. This dweeble family has a disease called "Lockets". Lockets is x-linked, dominant disease. The recessive expression is "Hoots". Fill in the genotypes of ALL individuals. Write your answers directly onto the pedigree below. Please us "h" to represent the disease in your answers.



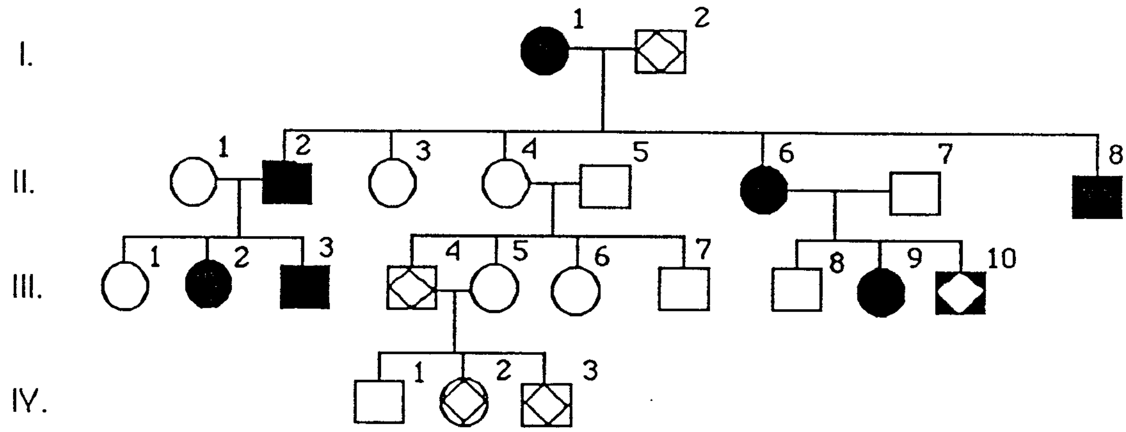
Question 8

The following pedigree represents a zork family. This zork family has a disease called "Quiggles". Quiggles is autosomal recessive disease Fill in the genotypes of **ALL** individuals. Write your answers directly onto the pedigree below. Please use "Q" to represent the disease in your answers.



Question 9

The ewoks are inhabitants of one of the moons of the planet ENDOR. They are forest-dwellers with a highly evolved (but rather incestuous) social structure; among other things they take family history very seriously and record every birth or death meticulously. The following pedigree is that of an influential ewok family:



A shaded symbol represents the occurrence of an extra toe and a diamond represents the occurrence of the much-admired square-shaped ears; both are rare traits.

Assuming that ewoks have a similar sex-determined mechanism to that of earthly mammals, and given the outsiders II1, II5, II7 are normal individuals,

a. What can you tell about the inheritance of an extra toe? Is it an autosomal or sex-linked, recessive or dominant trait?

b. What can you tell about the inheritance of square-shaped ears?

c. What are the probable genotypes of ewoks:

I1?
I2?
II3?
II7?

d. If ewok II3 and II8 decide to have children, what proportion of them would you expect to have an extra toe? square-shaped ears?

1. Was the tutorial useful in helping you to answer the questions given you today?

2. What did you like about the tutorial?

3. What did you dislike about the tutorial?

4. If you could change something or add something to the tutorial, what would it be?

1. Please list the high school math classes you have taken.

2. Please list the college math classes you have taken.

3. Please list the high school science classes you have taken.

4. Please list the college science classes you have taken.

5. Have you taken: (circle choice)

a) sociology	high school	college	none
b) anthropology	high school	college	none
c) archeology	high school	college	none

6. What is your major?

7. Did you take any Advanced Placement exams in high school? yes or no

If yes, which ones? (physics, calculus, biology, language, etc)

Symbols for use in solving the problems

AA

Aa

a a

XCXC

XCXc

XcXc

XCY

XcY

XX

XYc