Appendices A and B contain material to supplement Chapters 2 and 3, respectively.

Appendix A1: A Brief History of ISM-Diagrams

The four parts of Figure 10 show a few of the many forms taken by "the ISM diagram."

Figure 10 (and more!) is on the "Brief History of ISM" page (a link to it is on the BONUS-page).

The current ISM-diagram is Diagram 10-D. Unlike Figure 1, which has been modified to meet dissertation requirements for the University of Wisconsin-Madison by enlarging most of the type-fonts, and by moving elements closer together, 10-D has not been modified except for a photoreduction in order to comply with the university's requirements for wide margins.

****

Appendix A2: Controversies about Scientific Method

In the field of science studies there is controversy about the process of science, and how this process affects the content of science and the reliability of scientific knowledge. In Appendix A2, as earlier in Sections 2.44-2.45 and 4.23, I will not try to provide "answers" acceptable to everyone (which is impossible), but will simply discuss issues, and express opinions. And, as earlier, similar disclaimers apply, regarding a limited depth-and-scope, and the distinction between the ISM framework and my personal views about science, which are clearly stated (and are clearly not neutral) in the sections that follow.

The major issue is the reliability of scientific knowledge, with variations that involve skepticism about logic (discussed in Section B21), claims about unobservables and reality (in B22 and B23) and the effects of culture (in B24 and B25).

A21: Logical Skepticism

This section does not contain a sustained argument such as that in Section 2.44. Instead, there are simply a few comments and quotations.

A: Hypothetico-Deductive Logic

Wimsatt (1987, p. 8) describes an essential feature of a useful model; it should be "structured in such a way that we can localize its errors," and then offers this critique:

> There is a mythology among philosophers (the so-called 'Quine-Duhem' thesis') that this [localization of error] cannot be done, that a theory or model meets its experimental tests wholesale, and must be taken or rejected as a whole. Not only science, but also technology ...would be impossible if this were true in this and logically similar cases. That this thesis is false is demonstrated daily by scientists in their labs and studies, who modify experimental designs, models, and theories piecemeal, by electrical engineers...and by automechanics and pathologists, who diagnose what is wrong in specific parts...and correct them. ... Wimsatt (1981) gives a general analysis of how this is done, and a description of the revised view of our scientific methodology which results. (Wimsatt, 1987, p. 8)

B: Theory-Influenced Observations

Because modern science uses precisely defined standards, many observations are not subject to the whim of an observer. For example, a meter was originally defined by the distance between fine lines scratched on a bar preserved in a controlled environment in Paris; recently it was defined as 1,650,763.73 wavelengths of the red-orange light emitted from a Krypton-86 lamp, and now it is officially the distance traveled in 1/299,792,458 second by light in a vacuum.

And assuming a stable physiology (and memory for language terms referring to colors), a person observing a certain color, produced by light with a wavelength of 640 nm, may call it red, orangeish-red, reddish-orange, or orange, but probably not green or blue. Or if the light is observed by a spectrometer, the answer to a color question will simply be "640 nm."

Experimental design (and the theoretical orientations that guide it) determines the "what and how" of data collection, and also influences data interpretation — for example, by deciding how to categorize it during the process of observation. But there is a limit to this influence: "design determines the *type* of data, but nature determines the data. (Section 2.44_A)" Or, as described by

Darden (1991):

> The collection of Mendelian data was based on the unit-character concept. ... But the use of that concept, or category, in data collection, did not determine what the characters were, nor what their numerical ratios were. Categories reflect an advance resolution to individuate, group, and separate data along certain lines. They do not, however, determine any particular assignment of particular data points to those categories." (pp. 266-267)

C. Extreme Solutions for Pseudo-Problems

As described in Section 2.44_D, one danger of extreme skepticism is a tendency to be unable (or unwilling) to distinguish between potential problems and actual problems. Thus, there can be "cures for which there is no adequate disease. (Fodor, 1986)" Much of what I call "silly skepticism" is characterized by the proposing of extreme solutions for problems that don't seem to be problems:

> It is a curiosity of the philosophical temperament, this passion for radical solutions. Do you feel a little twinge in your epistemology? Absolute skepticism is the thing to try. Has the logic of confirmation got you down? Probably physics is a fiction. Worried about individual objects? Don't let anything in but sets. Nobody has yet suggested that the only way out of the Liar Paradox is to give up talking, but I expect it's only a matter of time. Apparently the rule is: if aspirin doesn't work, try cutting off your head. (Fodor, 1986, p. 1)

A "passion for radical solutions" also seems to motivate some modern sociologists who over-emphasize the influence of cultural-personal factors.

On the other hand, oversimplistic generalizations should be avoided; although radical skepticism seems to be an occupational hazard in 'study of science' fields, many modern philosophers and sociologists do strive for (and do achieve) a reasonably accurate understanding and description[1] of scientists' ability to convert a potential problem into a non-problem.

A22: Empiricism

One way to avoid some limitations of hypothetico-deductive (HD) logic is to avoid speculating about anything that is not observable. This *empiricist* philosophy of science was recommended by Francis Bacon in his *Novum Organum* of 1620, became the professed (but not practiced) method of

---

[1]. A "reasonable accuracy" is all I'm claiming for my own views on logical skepticism, also.

science into the mid-1700s, and experienced a revival among philosophers and behaviorist psychologists in the early 1900s.

A brief reminder of ideas discussed earlier: empiricist ≠ empirical; non-empiricist science requires HD logic, but empiricist science can be done with or without HD; and HD logic can be done for theories with observable and/or unobservable components. My negative evaluation of empiricism is briefly expressed earlier, in Section 2.24_B.

An empiricist perspective, wary of unobservable components, is part of an important debate about the proper foundations for science. Siegel (1995) describes three approaches to the epistemological justification of scientific theories — intuitionist, empiricist, and consequentialist — as exemplified historically, in the 1600s, in the work of Rene Descartes, Francis Bacon, and Isaac Newton. Descartes wanted to ground his theories on intuitive, self-evident axioms, using these as the foundation for deriving a deductive system of axioms and corollaries, and for regulating and validating the process of observation. Bacon urged the grounding of science on direct empirical evidence; in his suggested method, scientists can move from empirical particulars to generalizations, but they should not make claims that exceed their empirical knowledge. Although Newton did not explicitly describe a hypothetico-deductive consequentialist approach to justification, his theories and arguments are implicitly based on empirically testing the deductive consequences of a theory; if these consequences agree with observations, this provides indirect empirical support (but not proof) for the theory, even if the theory contains nonobservable components that surpass its empirical base.

One of the original motivations for empiricist philosophy was to build science on the firm foundation of empirical observations, and to therefore make scientific knowledge more certain. This was an important argument for empiricism in the 1600s, when an important goal of many scientists was a search for knowledge assured to be certain. But skeptics can criticize two aspects of alleged empiricist "certainty": the uncertainties of data, and uncertainties during inductive logic.

As discussed earlier, I think there are methods that, when used wisely by scientists, can minimize the conversion of potential logical difficulties into significant actual difficulties. But there is much in modern science that, in my opinion, cannot be reconciled with a rigid empiricism

such as that advocated by Berkeley and Hume:

> Berkeley continued his empiricist attack on the reality of gravitational attraction, but in addition he argued against the reality of forces more generally. ... Hume expanded Berkeley's empiricist attack to cover not just the notion of force, but the notion of causation more generally. Reality is what we see: cause is nothing over and above the constant conjunction of events, and we simply designate the earlier event the cause; it has no powers, nor is there any necessary connection between it and the effect. (Matthews, 1994, pp. 169, 170)

An important aspect of empiricism is its proposed restrictions on the freedom of science and scientists. Max Planck, responding to what he perceived as the threat (not promise) of empiricist science, describes a dangerous cognitive outcome if empiricist ideas were taken seriously by scientists:

> if the Machian principle of economy were ever to become central to the theory of knowledge, the thought processes of such leading intellects would be disturbed, the flights of their imagination would be paralyzed, and the progress of science might thus be fatally impeded. (from a public lecture of Max Planck in 1908; quoted by Toulmin, 1970, p. 26)

The Machian principle of empiricism attacked by Planck includes the claim that "gravitational attraction was not just unknowable, but there was no such thing: it was merely a human construct useful for the economy of thought and for the mathematization of particular experimental relationships. (Matthews, 1994, p. 170)" Thus we see a connection (but not one that is not logically necessary) between empiricism and instrumentalism. There also is a connection (but again, it is not logically necessary) with a particular view of metaphysics that claims to be non-metaphysical:

> For Mach, atoms were metaphysical rather than physical in character and thus, though the objects of speculation, they had no place in *science*. ... Much of Mach's philosophical activity was directed to drawing a sharp distinction...between physics and metaphysics. ... Once it was drawn, he thought, one could excise such metaphysical elements as atoms...and in this way put science on a firm and thoroughly empirical foundation. (Lambert & Brittan, 1987, pp. 5-6)

Even into the 1900s some empiricist scientists, such as Ernst Mach, criticized atomic theory for postulating the existence of 'atoms' that could not be directly observed. Similarly, the empiricist-inspired school of behaviorist psychology refused to include unobservable 'thoughts' in its theories, and harshly criticized non-behaviorist theories that did refer to the activity of thinking.

With regard to empiricist constraints prohibiting the use of unobservable entities or actions as components in theories, it can be useful to make a distinction between ontology and utility. For example, most behaviorist psychologists believed that thinking occurs; therefore, the action of 'thinking' was ontologically acceptable to them. But they did not believe it would be useful to

include thinking as a component in psychological theories, because they claimed that this would not inspire the gathering of reliable experimental data, that it would encourage theoretical speculations which could not be empirically verified or falsified, and that their own behaviorist theories were better at describing behavior.

For an empiricist, a theory should be only a way to conveniently summarize a large amount of data, or to make generalizations about observable quantities, and to make predictions. According to a leading contemporary defender of empiricism,

> To develop an empiricist account of science is to depict it as involving a search for truth only about the empirical world, about what is actual and observable. ... [This empiricist account] must involve throughout a resolute rejection of the demand for an explanation of the regularities in the observable course of nature, by means of truths concerning a reality beyond what is actual and observable, as a demand which plays no role in the scientific enterprise. (van Fraassen, 1980, pp. 202-203)

In my opinion, an "empiricist account" is a philosophical theory about what scientists *should do* — and, more important, what they should not do — rather than an empirical description of what scientists *actually do.* It is prescriptive, not descriptive. When van Fraassen states that "the demand for an explanation...plays no role in the scientific enterprise," it would be more accurate if he claimed that in his own opinion "the demand for an explanation...*should* play no role in the scientific enterprise."

Despite this, empiricism remains a credible view among philosophers, and empiricist perspectives of science did dominate two major fields, philosophy of science and psychology, during several decades in the first half of the twentieth century. But very few scientists have chosen to abandon atomic theory (which was urged by Ernst Mach, based on empiricist arguments) or to stop thinking in terms of 'forces' (which have been considered unobservable by many empiricists). The empiricist limitations of behaviorist psychology enjoyed a quarter-century reign of dominance, but recently the freedom of a less restrictive cognitive psychology has provided a liberating perspective for many scientists.

An empiricist view describes what philosophers think scientists should do. But if scientists "do what they should" they will operate at a disadvantage compared with scientists who misbehave, because most of the best modern theories are non-empiricist. Faced with this choice, to "behave as

they should" or to be effective, most scientists choose freedom, and most modern theories include unobservable entities, contrary to the demands of empiricism.

A23: Realism and Instrumentalism

One response to logical skepticism is to think of theory evaluation in terms of scientific usefulness rather than probable truth. Section 2.43 (with its concepts of status, intrinsic status and relative status, pursuit and acceptance, plausibility and utility, truth status and utility status, instrumentalism, realism, and critical realism, and variable-strength hypotheses) says most of what I want to say, by discussing a powerful arsenal of tools to use when thinking about the results of evaluation. The four subsections that follow will: describe the advantages of viewing science from the perspective of critical realism; discuss and critique the claims of instrumentalists; ponder a "million dollar wager" about scientific progress; and examine some very strange ideas about theories and reality.

A: The Flexibility of Critical Realism

Interpreting science using a flexible framework offers many advantages. Generally, flexibility improves utility, and allows a model to describe events more accurately. The ISM framework benefits from flexibility. So does the concept of *critical realism* because, rather than demanding a dichotomous choice between realism or instrumentalism, it offers the advantages of a 'best of both' eclectic approach.

Another useful approach is to recognize the broad range of possible (and actual) instrumentalist and realist positions. For example, to portray the range of 'realism', Leplin (1984) lists ten claims that a realist may or may not believe; by affirming or denying various claims, a variety of positions is possible, ranging from modest to strong. Also, the short-list of claims made by one modest realist might differ from the claims of another modest realist. This approach is useful for flexible critical thinking, because it lessens the tendency to use oversimplistic dichotomies, or to be swayed by 'strawman' arguments.

B: Pros and Cons of Instrumentalism

In my opinion, a limited form of instrumentalism can be rationally defended. Although my own position is 'critical realism' (which includes a limited instrumentalism), I respect instrumentalism as a valid position to hold. This respect contrasts with my views on strict empiricism, which I think is a very ineffective way to learn about the world; it is difficult for me to understand how anyone can seriously propose it as a way to do science in 1997. This subsection elaborates some arguments that are summarized in Section 2.45_C.

Laudan (1984) clearly expresses the two most common arguments in favor of instrumentalism. One argument is that theories in the past have contained unobservable components that once were considered to be real, but now these theories and their components have been abandoned. So why should we think that the components of our current theories will not meet this same fate? Because the history of science also provides many examples of postulated components (entities, actions, or interactions) that have survived a long time; and sometimes postulated components, originally unobservable, become observable when improved observation techniques and technologies are developed. And because this argument depends on inductive "boy who cried wolf" logic that, while it does call attention to an idea worth thinking about, is not deductively valid; after all, do you remember what happened in the wolf story? But the strongest argument for the reality of many components of modern theories is that it seems extremely improbable — if none of the theory-components (or very few of them) actually exist — that predictions based on these theories (and their components) would be accurate. In other words, the approximate truth of an empirically adequate theory (including its components) is an explanation for why this theory is able to make accurate predictions. This is no proof, of course, but it does seem like a rationally justified "good way to bet." And the claim of a realist is not that all current theories (or components) are approximately true, but that many of them probably are.

A second argument by Laudan is that a goal is "utopian" if there is no way to know if it has been achieved, and if there can never be a way to know it has been achieved. Since the truth of a theory can never be proved, realism is a utopian goal that should not be held by rational scientists. Compared with the first argument, I find this more impressive; it is a challenge that should be

given serious consideration. But I remain unconvinced, mainly due to the "best way to bet" arguments against logical skepticism described in Section 2.44. Even though there is no way to *prove* a theory is true or false, scientists can have a *rationally justified confidence* about it. And this is all that most modern scientists expect. To say that scientists *do* always think instrumentally is inaccurate, and to demand that scientists *should* never think of a theory in terms of realism is too restrictive. For reasons discussed earlier, I think that critical realism is a better way to view the actual practice of science, because it allows the flexibility to see that a scientist is thinking in terms of either realism or instrumentalism, depending on the situation. And it does not try to impose standards on how a scientist should think, because either realism or instrumentalism is acceptable within a critical realist framework.

In addition, I think that if science is not seen as a search for truth, and if there is no effort to find the truth, there can be serious ethical consequences. An instrumentalist viewpoint does not necessarily lead to a power-driven "might makes right" attitude, but if taken to an extreme this danger does exist.

C: Is there Scientific Progress?

Although to scientists the answer is obvious, in philosophy an important question is whether scientific knowledge is progressive — i.e., whether it improves over time. When judged by an instrumental standard, the progress in utility is clear. But progress in truth is impossible to verify, since none of us can be sure we know the truth, so a skeptic can challenge a realist by asking "Can you prove it?" As a critical realist, my quick answer would be "Well, no, but..." and then I would repeat the explanation that scientists have given up a quest for certainty, and are willing to settle for a high degree of plausibility. Scientists are not looking for proof, they're looking for "a good way to bet."

Consistent with this practical "best way to bet" goal of scientists, I ask you to consider a million dollar wager. Here is the scenario: Imagine that 1000 important 'state of the art' scientific theories from 1997, covering a wide range of phenomena in many fields, are being compared with the 1000 corresponding theories from 1497. You get to choose one set of theories, either 1497 or 1997, and

for each of the 1000 theories, someone who knows the truth about nature — such as an omniscient being, an alien from a scientifically advanced civilization, or whatever you want to imagine — decides which theory is closer to this truth. If your theory is more true, you win $1000, but if the other theory is more true you lose $1000. The question: Should you care which set of theories you get? Should you be satisfied if your opponent (the person who gets $1000 every time one of your theories is less true) makes the choice, and gives you the theories of 1497? According to those who claim that science does not improve with age, it should not matter which theories you get. If there is no scientific progress, the 1497-science and 1997-science have an equal chance to be closer to the truth.

With the skeptical claim that "there is no way to *prove* the 1997 theories are better," we're back at the beginning, deadlocked over an unreasonable demand for "proof" that has no place in science. On the other hand, if the evaluation criterion is "a *rationally justified confidence* that the 1997 theories are better," anyone who is not a fool (or who wants to give away a million dollars) should choose the science of 1997.

For a rough estimate of how consistently superior you think the theories of 1997 are, consider a wager with two options: you can pay $600,000 for the right to choose the 1997 theories, or you can decide not to play. If you choose to play, you lose $1,600,000 if all of the 1497 theories win, you win $400,000 if the 1997 theories always win, and you break even if there is a 20-80 split between 1497 and 1997. Are you confident that in 80% of the selected areas, science is better now than it was 500 years ago? I would eagerly take the bet, with confident assurance that I would win roughly $400,000. Would you take the bet if the entry fee was changed to $400,000 (requiring a 30-70 split to break even), or $800,000 (you need a 10-90 split), or $900,000 (at least a 5-95 split)?

There is a more important question, however, since the main goal of ISM is to describe science as it actually is: What do you think the majority of scientists would do in each of these scenarios? I think most of them would take the bet, even if they had to pay $900,000 for the chance to win $100,000. After all, these 1-to-9 odds aren't too shabby when betting on 500 years of scientific progress.

1

D: Do Scientists Create Reality?

This discussion is related to the discussion, in Section 2.45_C, about a 'consensus' definition of truth. A "consensus definition of truth" is when someone *talks* as if believing something makes it true; a theory about "creating reality" is when someone actually *believes* that believing something makes it true.

Do scientists study nature, or create nature? Somewhat amazingly,

> [Some scholars argue that] the traditional (realist) interpretation of the relationship between an object and a representation of that object should be reversed. ... Scientists have always described their work in terms of the making of discoveries; discoveries that are indeed discoveries — the unmasking of objects that exist somewhere in the world waiting to be found. Instead, Woolgar (1989) argues that scientists construct these objects through their representations of them. Objects, according to Woolgar, whether they are countries, or electrons, are socially constructed entities, and do not exist aside from this social construction. Science is therefore *not* the process of finding things that already exist, but the process of creating things that were not there to begin with. (Finkel, 1993, pp. 31-32)

In the words of Latour and Woolgar (1979, p. 64), "The bioassay is not merely a means of obtaining some independently given entity; the bioassay constitutes the construction of the substance." One application for this concept of "creating objects in the laboratory" is to discredit the credibility of data:

> If objects do not exist independently of our representations of those objects, then data (defined here as what 'anyone' would see if they examined a particular object or situation) has no meaning outside of an interpretative context. ... As a result, "The study of scientific knowledge is primarily seen to involve an investigation of how scientific objects are produced in the laboratory rather than a study of how facts are preserved in scientific statements about nature. (Knorr-Cetina, 1983, p. 119)" (Finkel, 1993, p. 32)

Commenting on this type of claim — that "objects do not lie around ready made in the world but are mental constructs (Wheatley, 1991, p. 10)" — Matthews (1994, p. 152) explains a crucial distinction: "Where he...goes wrong is in failing to distinguish the theoretical objects of science, which do not lie around, from the real objects of science, which do lie around and fall on people's heads."

A description of the way scientists typically view the observation of real objects (no, it is not necessary to "create reality") is provided by a cell biologist:

> First, I assume that cells are real objects. Second, I assume that *other people can see and think about things the way that I do.* ... Others' basic experience of reality is similar to mine. If they were standing where I am standing, they would see something very similar to what I see. ...

> Scientists act as if...the observations made by one scientist could have been made by anyone and everyone. (Grinnell, 1992, p. 20)

Another excellent description of 'truth' and 'reality' is given by a prominent philosopher:

> Whether a statement is true is an entirely different question from whether you or anybody believes it. ... There can be truths that no one believes. Symmetrically, there can be beliefs that are not true. ... The expression "It's true for me" can be dangerously misleading. Sometimes saying this...means that you believe it. If that's what you want to say, just use the word 'belief' and leave truth out of it. However, there is a more radical idea that might be involved here. Someone might use the expression "true for me" to express the idea that each of us makes our own reality and that our beliefs constitute that reality. I will assume that this is a mistake. My concept of truth assumes a fundamental division between the way things really are and the way they seem to be to this or that individual mind. (Sober, 1991, pp. 15-16)

Next, Sober illustrates what he considers to be a valid meaning of "thoughts becoming reality" by describing a situation in which a person's thoughts (he thinks he won't hit the baseball) affect his actions (he swings too high) thus causing a result (he doesn't hit the baseball). By contrast,

> What I do deny is that the mere act of thinking, unconnected with action or some other causal pathway, can make statements true. I'm rejecting the idea that the world is arranged so that it spontaneously conforms to the ideas we may happen to entertain. (Sober, 1991, p. 16)

These quotations, from a scientist and a philosopher, cover the most important concepts in "Reality 101" so I'll just close this section with an example from science: Anyone who believes that "beliefs create reality" should be eager to explain how the motions of all planets in the solar system changed from earth-centered orbits in 1500 (when this was believed by almost everyone) to sun-centered orbits in 1700 (when most people, at least in the scientific community, believed this).

A24: Relativism

This section is a continuation of Section 2.45_B, which discusses the basic concepts of relativism.

A. Motives for Relativism

What are the motivations for relativistic views? The most obvious source is critical thinking. When the process of science is examined — whether the focus is on the limits of logic, or is broadened to include other aspects of science — there are reasons to question the reliability of

scientific theories. Another source is professional rivalry. If philosophers study mainly logical relationships, and sociologists study mainly social relationships, neither group wants to believe that what it studies has a trivial effect on science. If social relationships and culture exert a strong influence, this increases the prestige of sociologists, thus producing a powerful incentive to argue for cultural influence and for relativism. And in the "battle between humanities and sciences" postulated by Snow (1959), relativist historians can attack the rationality and credibility of scientists.

Of course, either of the motivating factors discussed above, logic or pride, can also lead to non-relativist views. Critical thinking can lend support to the credibility of science. And if logical factors are judged to be more important than cultural factors, the prestige of philosophers (relative to sociologists) increases.

Relativism in social scientists may also be encouraged by the characteristics of the areas being studied. In most areas of natural science (such as biology, chemistry, physics, or astronomy), empirical data limits the range of reasonable interpretations. This is less true in many areas of social science (such as psychology, sociology, or economics) where the systems and data are complex, and diverse interpretations can be rationally defended. Therefore, sociologists will tend to be more impressed by skeptical arguments for underdetermination, because in their own fields (and thus in their personal experience with science) the 'reality checks' are less clear and less limiting than in most fields of science.

Finally, metaphysical and ideological influences can contribute to a preference for relativism, or against it.

Just as a field of science and its participants can be analyzed in terms of cultural-personal factors and thought styles, so can the 'study of science' fields and their participants. There are major contrasts in thought styles between the different disciplines that study science. Views that are praised and rewarded in sociology will be less welcome in philosophy, and vice versa. And there are differences between fields that *study* science and fields that *do* science. When I entered a history of science graduate program with a background as a scientist, it was fascinating to compare the thought styles in science and history. While there were many similarities between these fields,

such as a concern for developing theories based on solid evidence and coherent logic, there were also important differences, the most important being the perspectives on science and scientists.

Differences also exist within each study-of-science field. For example, even though philosophers tend to emphasize the functioning of logical factors in science, some philosophers do study cultural factors, and many who do not study culture can nevertheless appreciate its importance; they just choose to specialize in studying other aspects of science. An intra-field difference in thought styles is apparent in the following passage, in which a moderate sociologist questions the motivations of radical colleagues:

> What is the point of doing sociology of scientific knowledge? Are we trying to increase the understanding of science, or are we rather engaged in a curious endeavor whereby we contribute to the *sociology* of scientific knowledge without really contributing to the understanding of science? This is not to say that the current sociology...does not "work" for its proponents. In fact, here we have a good example of a self-sustaining business. ... I want to encourage research that aims at finding out about the 'truth' of matters, since I regard this as the only possible scientific attitude. ... Of course, it may be the case that many sociologists have simply given up understanding and are now engaged in school construction and empire building. (Segerstrale, 1993, pp. 59, 76, 75)

But is this a futile plea? Empire building is consistent with a cynical view of science, especially if this is combined with an instrumental perspective in which truth is not the main goal. For a relativistic cynic, if radical theories of sociology are useful professionally (in getting publications, grants, employment, tenure, and prestige) or politically (in promoting desired social changes) then it does not matter whether these theories are true. In these circumstances, it is rational to ignore the truth in order to do what is expedient for empire building. Some radical sociologists (but not all) would agree that this is a smart strategy, in sociology or in science: "Each text, laboratory, author, and discipline strives to establish a world in which its own interpretation is made more likely by virtue of the increasing number of people from whom it extracts compliance. (Latour and Woolgar, 1986, p. 285)" Is the goal of science to "extract compliance"? Frankly, I think this is far too cynical, and that most scientists, including most social scientists (such as Segerstrale), do pay attention to the guiding constraints offered by 'reality checks', and do want to find the truth.

B. Criticisms of Radical Relativism

The logical deficiencies of extreme 'strong program' perspectives have been described by many

critics. One suggestion is that perhaps scientists should have more voice in what is said about science:

> One wonders about the justification for a sociology of scientific knowledge that either ignores scientists or whose analysis cannot be perceived as useful to practicing scientists. There is an authoritarian ring to many of the new approaches...whereby the sociologist knows best and does not need to consult with the object of his or her study. ... What is arrogantly ignored in current sociologies are scientists' own cognitive commitments and cognitive style, as well as hard-to-identify personality factors. The time has come to bring the scientist back in. ... [This] would mean that the sociologist of science might use scientists as codetectives to gain some understanding about scientific judgment. Scientists are surprisingly adept at introspection as to their underlying methodological, theoretical, and even metaphysical commitments, cognitive style, and so on, if only given a chance. (Segerstrale, 1993, pp. 57-58, 75)

Another critique of the 'strong program' comments on the program and on *Knowledge and Social Imagery* (Bloor, 1976), and *Laboratory Life* (Latour and Woolgar, 1979; 1986):

> The claims of contemporary SSK [sociology of scientific knowledge]...for the external causation of scientific beliefs are baseless. ... When distilled to its essence, the entire 'argument' [of Bloor]...is just this spurious inference from underdetermination to social construction. ... The radical, iconoclastic view of science [in *Laboratory Life*] is not merely without foundation but is an extravagant deconstructionist nihilism according to which all science is fiction and the world is said to be socially constructed by negotiation. ... The undoubtedly difficult questions concerning the nature of science have led these sociologists to conclude that there is no such thing! (Slezak, 1994, pp. 265, 283, 329, 330)

Cultural-personal factors are a part of science, so they are included in the ISM framework and elaboration. As usual, though, if a good idea is taken to extremes, in this case with exaggerated interpretations of cultural influences and rigid generalizations of the "all __ is __" type, the result will be a distorted picture of science. I think it is better to maintain a better balance, and to construct a more accurate picture of science.

A25: Tools for Analysis: Idealization and Range Diagrams

Before describing the "tools for analysis" I will briefly explain why I think these tools could serve a useful function. The characteristics of cultural-personal factors, and their effects on the process and content of science, are topics for hot debate among scholars who study science. When evaluating different viewpoints, especially the more extreme interpretations of science, such as those that claim a very high influence of "politics" in "science," it helps to have tools that

encourage flexible, critical thinking that is more precise and accurate. The following subsections discuss two analytical tools, idealization and range diagrams, that may be useful in avoiding dichotomous generalizations and in clarifying the ways that science is (and is not) influenced by cultural-personal factors.

A. Analysis by Idealization

One approach to analyzing the characteristics and influences of 'culture in science' begins with idealization. Just as Newton tried to imagine motion without friction, we can try to imagine what science would be like without cultural influences. By comparing this idealized science with actual science, we can estimate the influence exerted by various types of cultural factors, and how these affect the process and content of science, in the short-term and over longer periods of time. The logic used in this idealization process is described in Section 2.26_E; when a simplified model that is known to be false because it omits a known component (in this case, cultural-personal factors) serves "as a 'control' for determining the effects of components that have been omitted or oversimplified."

In doing an idealization, great care should be taken with the selection of cases to study. Because the amounts and types of cultural-personal influence vary from one situation to another, it is important to ask whether the selected cases accurately represent a larger population of situations (for science as a whole, or for specialized fields) and how they fit into the 'big picture' of science and scientists.

In certain cases, with certain interpretations of science, idealization may seem easy. But with other cases or interpretations, the conclusion may be that some aspects of science — especially some foundational 'thought style' beliefs — are so intimately intertwined with some aspects of culture that it is difficult to separate the science and culture from each other. Either way, the process of attempted idealization could be instructive.

Perhaps one of the most instructive aspects of this process will be the attempt to idealize. How does one try to imagine what science would be like, in a certain field, without the thought style that operates in this field? When a component of a thought style is essential for doing science, how

does one define what this component — and the resulting science process and content — would be "without cultural influence"? Doing this thought-experiment skillfully requires a great deal of careful thinking, and this thinking may be the most important part of the process. One strategy for characterizing a thought style is to imagine several models of science, each with a different type of thought style, and then compare the predictions of these models with each other, and with what really happened in the situation being studied. { In Section 2.26_E this strategy is described using damping force, rather than thought style, as the component that is being varied from one model to another. }

The goal of this 'analysis by idealization' would be a deeper understanding of the characteristics and functioning of cultural-personal factors and thought styles, and how these affect the process and content of science. Of course, this understanding would also be informed by the scholarly work of historians, philosophers, sociologists and psychologists, and by models such as the interpretive framework described in Section 2.32.

One of the most valuable functions of analysis (whether it is done by idealization, by making range diagrams, or with any other technique such as those typically used by study-of-science scholars) is to encourage a study of specific cases. Hopefully, these 'reality checks' will reduce a human tendency toward simplistic generalizations and dichotomous stereotyping. With overly simplistic generalizing, pronouncements about "science" can ignore important differences. For example, the influence of societal politics spans a wide range, from a psychologist studying sociobiology (this can be very politicized and controversial) to a chemist studying the kinetic effects of groups attached to a benzene ring (there is not much 'politics' evident here).

B. Analysis using Range Diagrams

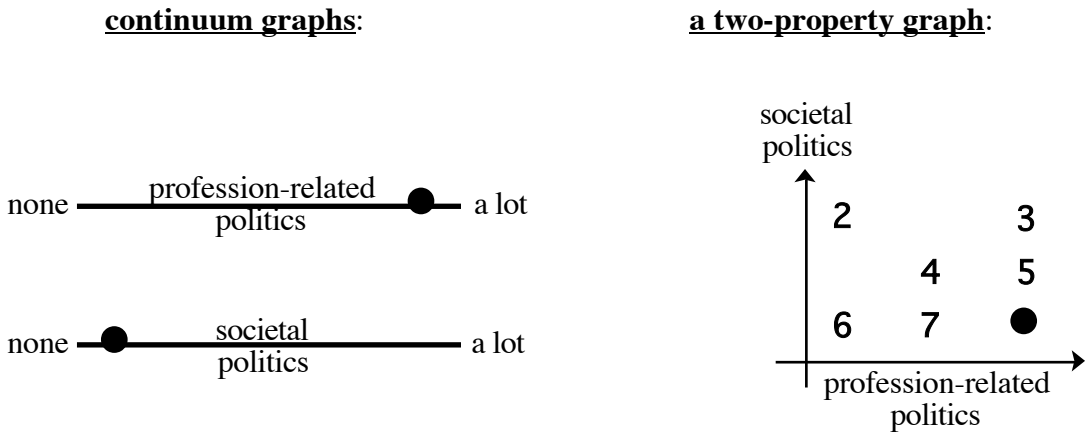
One of the most useful parts of the ISM framework is its flexible language, carefully designed to avoid the need for rigid dichotomous thinking. For example, Section 2.43 describes a set of terms (including truth status and utility status) designed to encourage flexible thinking about the goals of evaluative conclusions. Similarly, this subsection describes a set of diagrams designed to encourage flexible thinking about the influence of cultural-personal factors in the process and

content of science.

The left side of Figure 11 shows the simplest type of range diagram, a *continuum.* In the example shown, the range is from "none" to "a lot." The top diagram shows that, for an imaginary science situation, the estimated amount of 'profession-related politics', shown by the dot, is fairly high. The dot on the bottom continuum shows that for this situation the estimated amount of 'societal politics' is low.

The right side of Figure 11 shows a second type of range diagram, a *two-property graph,* that is made by combining the estimates that are made on two continuum graphs for seven imaginary science situations. The dot shows that, for the science situation shown in the continuums on the left side, the amounts of profession-related politics and societal politics are high and low, respectively. When abbreviated using a math convention of listing the x-variable first, followed by the y-variable, the first situation is "high/low." In situation #2 the profession-related politics and societal politics are estimated to be low and high, respectively (low/high); the other five situations are, in order, high/high, medium/medium, high/medium, low/low, and medium/low.

Figure 11: Two Types of Range Diagrams for "Politics in Science" (with imaginary analysis)



These graphs can also be used "in reverse" to stimulate thought-experiments; for each of the seven dots in the two-property graph above, try to imagine a corresponding situation. This might stimulate thinking about correlations between the two types of politics. For example, while

imagining dot #2 I became aware of a potential asymmetry in causal relations. I can easily imagine situations such as #1 where profession-related politics has little to do with societal politics. For #3, I can imagine situations where both types of politics overlap and intermingle, with each helping to produce and support the other, with high degrees of both getting mixed together in a mutually supportive blend. It is more difficult to imagine #2, with societal politics active in science, but with no corresponding manifestation in profession-related politics. This theory about #2 may be wrong, but it seems to be an idea worthy of pursuit, and I had never formulated it explicitly until I drew Figure 11 and thought about the meaning of each dot.

The preceding paragraph should serve as a reminder that ideas can serve many functions. Range graphs could be used as initially described, by starting with empirical data and using this data to construct one or more types of graphs. Or, as in the thought-experiments just described, a range diagram can be used to stimulate ideas. Or a diagram could be used as a pedagogical tool, to help students or colleagues understand an idea, or to persuade them of its plausibility or utility. The diagrams described in this section should be considered only a starting point for developing this type of analysis, not a final formula for a polished technique.

One variation is to expand a two-property graph to three properties. There are a number of ways this could be done. If the third property is easily 'split' — for example, by splitting scientists into those who study sociobiology, benzene chemistry,... — a two-property graph (as in Figure 12) can be drawn for each, to show the similarities and differences between different fields in science. Or, if situations can be split into those having a low, medium or high value for the third property, a separate two-property graph could be drawn for the situations that fall into each of these categories. Or just draw a 3-dimensional graph for the three properties. If the analysis gets too complicated (such as a graph with 4 dimensions), the principles learned from do-able range graphs can be extrapolated into more complex analyses where a literal drawing is not feasible.

Another possibility is to split Figure 11, for politics in science, into two parts — for the effects of politics on science process and on science content. Or perhaps these could be further split into effects that are short-term and long-term.

Or, moving in the direction of simplification, a two-property graph can be converted into a 2x2 matrix with 4 cells, for the dots at each of the four corners; in Figure 11, these are 2, 3, 6, and 1.

Or, if each of the properties is split into three categories (for low, medium, and high) this will make a 3x3 matrix with 9 cells.

Figure 12 shows a third type of range diagram, a *frequency graph.* The first graph shows estimates based on an imaginary analysis that agrees with my own views for "science as a whole." This graph shows, for imaginary selected situations, that the frequencies for profession-related politics range from none (at the left end of the graph) to high (at the right end of the graph); the frequency peaks at a medium-high amount of societal politics, and decreases on either side. In the second graph, the frequency peaks at a low amount of politics. A comparison of these two graphs shows that the imaginary person doing the estimates thinks (in agreement with myself) that profession-related politics is more frequently a significant factor in science, compared with societal politics.

Figure 12: Frequency Graphs for "Politics in Science" (for imaginary analysis)

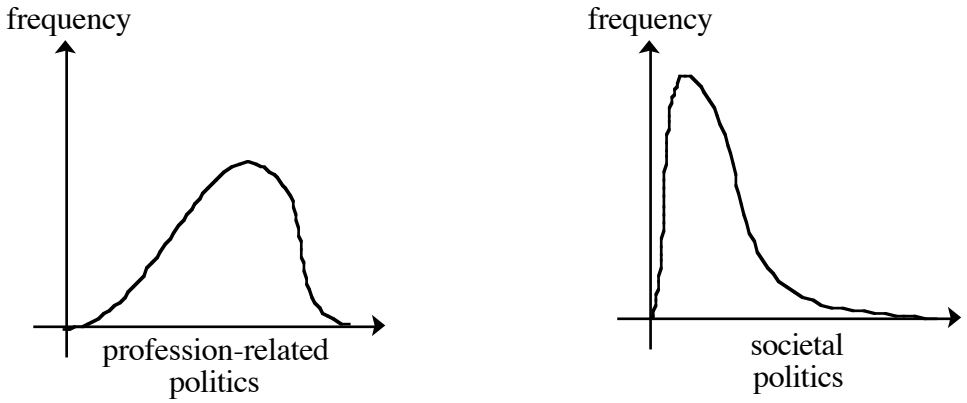
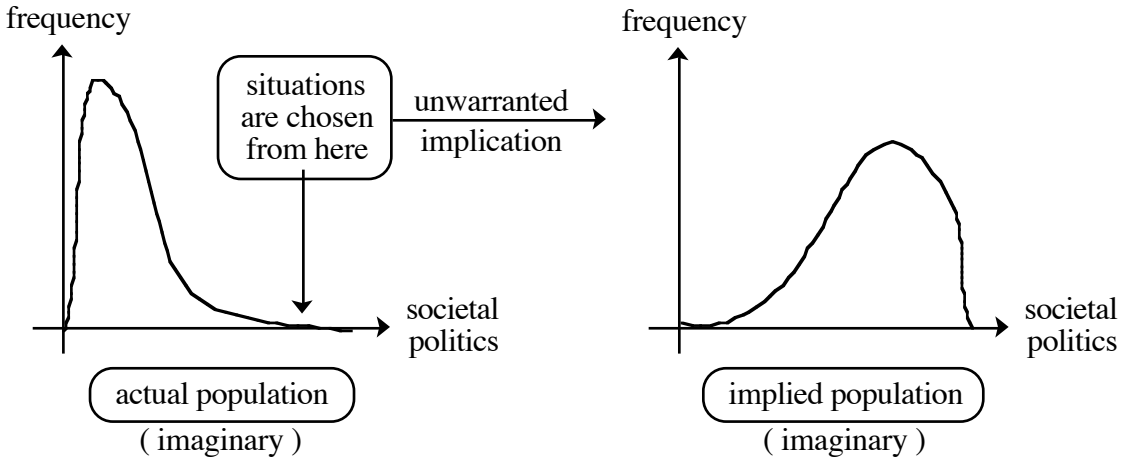


Figure 13 shows another way to summarize frequency-information. Instead of using a frequency graph (as in Figure 13), Figure 14 begins with a continuum, and then conveys a quick-and-rough idea of relative frequencies by shading the parts of the graph where frequency is highest. For interpreting this graph, it may help to imagine that the most heavily shaded areas have more 'dots' of the type used in Figure 11.

Figure 13: A Method for showing Approximate Frequencies (for imaginary analysis)

Figure 14 shows the unwarranted implications that occur when the situations selected for analysis are not an accurate representation of the entire population. For example, if the only situations that are studied involve a high degree of societal politics, this biased sample does not accurately represent the entire population which, as shown in the left-side frequency graph for the "actual population," contains many situations in which societal politics is low. But the "implied population" contains a much higher frequency of high-politics situations, because it has been inductively generalized from a sample that is biased in favor of high-politics situations.

Figure 14: Unwarranted Implications occur if Examples do not accurately represent Population (for imaginary analysis)



As with idealization, in using range diagrams the goal would be a deeper understanding of cultural-personal factors and their functioning in science. For example, careful examination might show — as I suspect it would — that profession-related politics occurs more frequently (and more obviously) than societal politics. And while some components of thought styles, such as the choice of metaphysical assumptions, can have enduring long-term effects, many cultural factors will exert

influence that is only short-term. For example, preferences for posing certain types of problems may delay the investigation of non-preferred domains or the development of theories in these areas, but it probably won't shut scientists out of these domains forever.

****

Appendix B1: Prediction Overviews and Potential Problem-Solving Actions

Appendix B1 performs two main functions. First, it explains a newly developed representation for genetics, using illustrations for six inheritance patterns. Second, by examining the strategies that *could be* used by students when they solve effect-to-cause problems that require the revising of genetics models, it provides a foundation for understanding the problem-solving strategies that *are* used by students in the MG course, according to the findings of previous research that will be discussed in Appendix B2.

The genetics terms used in Appendices B1 and B2 are defined in Section 3.25.

All diagrams for Appendix B1 (Figures 15-21) are in the "tables.rtf" file.

B10: A New Type of Representation: Prediction Overviews

Prediction Overviews provide overviews for the process of predicting the results of genetics crosses. Section B10 begins with an explanation of symbolism. Next, a Prediction Overview (PO) for the model of dominance (in Figure 15) is described and is compared with Mendel's Bible (Figure 7); a brief summary of this comparison is that Figure 15 shows how to generate, in a systematic way, the "Cross Possibilities" in Mendel's Bible. Finally, there is a brief discussion of the utility of Prediction Overviews.

A. A System of Symbols

In OView 0-A, the first diagram of Figure 15, the capital letters (**A** and **B**) represent two phenotypic variations of a particular trait for a fruitfly. For example, if the trait is eye color, A and B could be white eyes and red eyes, respectively. The small letters (**a** and **b**) represent genotypic alleles. This symbolism is logically meaningful in two ways. First, small letters represent the microscopic level of alleles (which are segments of DNA molecules) that are 'unobservable' in GCK, and capital letters are used for the macroscopic level of trait-variations that are 'observable' in GCK. Second, there is a cause-effect relationship between the small and capital letters; **a** tends to cause **A**, and **b** tends to cause **B**.

Although this use of symbols differs superficially from the symbolism in Mendel's Bible, which uses numbers (1 and 2) to represent alleles, these two symbol-systems are actually quite compatible. But the symbolism in Figure 15 (and in Mendel's Bible) does contrast with one type of genetics symbolism where small and capital letters are used for the recessive and dominant alleles, respectively. Section C16 discusses the relative advantages of several symbol systems, including letters and numbers, and the small/capital system.

B. A Prediction Overview for a Model of Dominance

Figure 15, which includes four Prediction Overviews (OViews 0-A to 0-D, for the "Round Zero" of model revising), shows a systematic way to generate all of the six "Cross Possibility" predictions in Mendel's Bible, based on a theory of simple dominance.

In OView 0-A the three cells in the left column show all possible genotypes (aa, ab, bb) of a mother fly, and the corresponding phenotypes (A, A, B) if the genotype-to-phenotype mapping is 'dominance' and the dominant allele is 'a'. The top row shows all possible genotypes of a father fly, and the corresponding phenotypes. In the body of the 3-by-3 matrix, each of the 9 cells contains a miniature Punnett Square (in the top half) and the corresponding phenotypes (in the bottom half). For example, the middle cell of the bottom row shows the prediction when a bb-mother (phenotype B) mates with an ab-father (phenotype A); the four possible genotype

combinations are “ba, bb, ba, bb” which, with dominance, produce “A, B, A, B” phenotypes, respectively. Notice that the bold lines divide the 3-by-3 matrix into four areas (with 4, 2, 2, and 1 cells), one area for each of the four possible phenotypic crosses: AxA, AxB, BxA, and BxB.

OView 0-B is derived from OView 0-A by assuming that, in a particular cross, each of the four possible genotypes for progeny is equally likely. As in OView 0-A, each area (enclosed by bold lines) shows all genotype crosses that can occur when two phenotypes are crossed.

OView 0-C is the same as OView 0-B, but 0-C has shadings to show the ‘duplications’ that occur due to the fact that in GCK (always) and in nature (usually) it doesn't matter whether an allele comes from the mother or father. Each shaded box contains an explanation; for example, the first shaded cell explains that the result of this cross (ab x aa) is the "same as aa x ab". The purpose of the shading is to make it easier to see that the 6 non-duplicate cells of OView 0-D match the 6 "cross possibilities" in Mendel's Bible. Thus, Figure 15 does not contain any information that is not already in Mendel's Bible, but it does offer a different perspective for viewing the information.

OView 0-D summarizes the data that is predicted for each of the four phenotypic crosses. It is derived by combining the data from all possible genotype crosses within one area of OView 0-B (or 0-C) into one cell of OView 0-D. For example, the AxA area of OView 0-B contains four results, but only two different results; these are summarized in OView 0-D as "100% A or 75% A, 25% B." In three cells (AxA, AxB, BxA) there can be two different results; the result that actually is observed depends on which parents are selected for the cross. But with BxB the same result always occurs, no matter which parents are chosen.

Figure 15 contains two distinctive types of representations. First, OView 0-A is a “set of miniaturized Punnett Squares with genotype-to-phenotype mappings.” Second, the data from 0-A is “mathematized” in 0-B and 0-D. The following discussion will not distinguish between a “super-Punnett” and a “data summary,” and will refer to either as an "OView."

C. Utility — Scientific, Instructional, and Analytical

Figure 15 shows, in a systematic step-by-step process, how POs can be used to generate predictions for all possible crosses. And it shows the cause-effect relationships in genotype-

phenotype mappings, how "all possible results viewed at the level of genotypes" cause "all possible results viewed at the level of phenotypes." POs are especially useful for developing a deep, sophisticated understanding of genotype-phenotype relationships in newly developed models. During the simulated science of the MG classroom, this understanding is essential for the two essential steps (anomaly recognition and anomaly resolution) in solving problems that combine effect-to-cause reasoning with model revising; and these problems are the focal point of the course. The connection between genetics knowledge and problem-solving strategies will become evident in the detailed analysis of student actions in Section B22.

Perhaps POs could be useful for instruction, by helping students improve their problem-solving skills and their conceptual understanding of genetics. If (as recommended in Section 3.52_C) POs were used in the Monona Grove classroom, they would supplement Mendel's Bible (MB), not replace it. These two representations are complementary; each is useful for different purposes, or (if they are used for the same purpose) each provides a view from a different perspective. The main functions of POs are to facilitate the processes of prediction and processing, to make interpretations for anomaly recognition, model evaluation, retroductive inference, and experimental design. On the other hand, MB is a simpler yet more complete description for the model of dominance, which is useful as a "template for tinkering" during the process of inventing a new model by revising an old model.

During the analysis in Section C2, the main function of POs is to facilitate an understanding of the problem-solving strategies (especially for making predictions) that students either have used, or could use.

B11: A Model for Round 1 — Codominance

A. Anomaly Recognition

Figure 16 begins with OView 1-A which shows the available data, because this is where students begin their problem solving. In Round 1 the phenotype symbols are changed, from 'A and B' to 'C, D, E', to emphasize that a different trait — with different variations, and a different

inheritance pattern — is being studied. Similar symbol changes will be used in describing Rounds 2-4. In the table the shaded cells indicate anomalies — when data is contrary to predictions made by the existing theory of dominance. First, and most obviously, there are three variations, not two. Second, the 25-50-25 mix (from DxD) could never occur with dominance because it produces 3 variations. Third, with DxD, in addition to this unexpected mix that *does occur,* there is also an expected result that *does not occur*; this 'missing result' is that DxD never produces 100% D, even though in all other 'alike parent' crosses the students have done (AxA and BxB with dominance in Figure 15, and CxC and ExE in this round) sometimes 100% of the progeny have the same phenotype as their parents. A fourth anomaly is more subtle because, unlike DxD, it involves crosses that initially appear normal because they produce an expected result (CxC → 100% C, or ExE → 100% E); but there is anomaly because this cross never produces another expected result (the 75-25 mix that sometimes occurs for AxA with dominance). Fifth, a particularly striking cross is CxE (or ExC) because the result is 100% D, unlike either parent; this never happens with dominance.

B. A General Problem-Solving Strategy

In a simplified version of problem solving, a strategy is to "find what is wrong and fix it." In science this strategy becomes "recognize the anomaly and resolve it." In the problem for Round 1, "what is wrong" is the anomalies that have been recognized. As described above, these could include the 3 variations and/or the results of DxD, CxC (or ExE), and CxE. In science the usual way to "fix it" is to invent a theory that can explain the puzzling data, thus resolving the anomaly. This two-step process — anomaly recognition followed by anomaly resolution — will be the organizing theme for the analysis in Section B22.

The general process of theory invention (which is one aspect of anomaly resolution) was discussed in Sections 2.52-2.53. And later, Section B22 will examine the specific processes of invention — especially by the revision of existing models — that are used in this genetics course. But in Sections B11-B14, with the exception of B12, it will simply be assumed that a revised model has somehow been invented, without asking how this happened.

C. Anomaly Resolution

OView 1-B shows a theory of 'codominance' that — because it can explain all the data in OView 1-A, including anomalies — must be evaluated as empirically adequate. In OView 1-B, for all crosses the progeny genotypes are the same as in dominance, but there is a difference in mapping genotypes to phenotypes. With the mapping in dominance, if c is dominant over e, then two genotypes (cc and ce) produce the same phenotype C. But with codominance there is a one-to-one match between genotypes and phenotypes, with each genotype mapping to a unique phenotype. Neither allele is recessive, because c and e are both active in producing D, which differs from the phenotype produced by only c (cc → C) or by only e (ee → E); or we might consider both to be dominant, as implied by the name 'codominance'.

OView 1-C summarizes the data in OView 1-B, making it easier to see that the predictions of codominance theory are identical to the data in OView 1-A. Of course, with the data generated in an actual GCK problem the match would be approximate rather than perfect.

D. Model Revising

In proposing changes to Mendel's Bible (MB) for each round of model revising, the extent of changes will be minimized. For example, in the "model revising" description that follows, instead of changing the variations from A and B (as in MB) to C, D and E (as in Figure 16), which would be 3 changes, only 1 change will be made — to A, B and F. For the same reason, the allele names '1 and 2' will be retained from MB.

One major revision converts the MB model for dominance into a model of codominance: in "Genotype/Phenotype Combinations" the mapping of "1,2 Variation A" is changed to "1,2 Variation F". This major change produces a need for minor "mopping up" changes in two sections. The "# of Variations/Trait" changes from 2 to 3. Then, in the "Cross Possibilities" table, all "Genotypes" that involve "1,2" change; the details of these changes (in the columns labeled "Phenotypes" and "Probable Results") can be worked out by using OView 1-B.

B12: A Model for Round 2 — Multiple Alleles

A. Anomaly Recognition

In Round 2, students can check the data for anomalies with respect to either of their existing models, dominance and codominance, by comparing OView 1-C with OView 1-A or 1-B, respectively. With respect to either existing model, the most obvious anomaly is that now there are 4 variations. This immediately eliminates both simple dominance and simple codominance as credible theories.

B. Anomaly Resolution

To make it easier to compare data, Tables 2-A, 2-B and 2-C are on the same page. For visual simplicity, in OView 2-C the 'duplicate' crosses are blank.

In OView 2-C the shaded areas show crosses that provide "clues" about the inheritance pattern. At the beginning of experimentation it is difficult to characterize crosses more precisely than whether or not both parents have the same phenotypic variation, so this will be our starting point. Using this strategy, for any same-variation cross of alike parents (such as RxR, OxO, SxS, or TxT) the data in OView 2-C can be compared with the predictions (in Tables 2-A and 2-B) for same-variation crosses made with dominance (AxA, BxB) and codominance (CxC, DxD, ExE). Similarly, the results for different-variation crosses (of RxO, RxS, RxT, OxO, OxT, or SxT) can be compared with the different-variation crosses for dominance (AxB) and codominance (CxD, CxE, DxE).

The shaded experimental results are useful for retroductive reasoning. First, notice that "RxS → 100% **O**" is analogous to "CxE → 100% **D**"; since **D** is a codominant variation produced by a **ce** genotype, maybe (by analogy) **O** is produced by **rs**. Next, notice that the same-variation cross of RxR producing "100% R or 75% R, 25% T" is analogous to AxA producing "100% A or 75% A, 25% B"; **a** is dominant over the recessive **b**, so maybe (by analogy) **r** is dominant over **t**. By similar reasoning, maybe **s** is dominant over **t**. Using the results of these three phenotype-crosses

and analogical reasoning, we can construct a model with 3 alleles (r, s, t) in the population; there are 3 homozygous genotypes (rr → R, ss → S, tt → T), and for the 3 heterozygous genotypes there is 1 codominant relationship (rs → O) and 2 dominant relationships (rt → R, st → S). This model is shown in OView 2-D (for the genotypes) and OView 2-E (for the corresponding phenotypes).

The remainder of the shaded crosses also support this model. OxO always produces a 25-50-25 mixture, analogous to DxD. TxT produces only 100% T, analogous to BxB. RxT and SxT produce "100-0 and 50-50" mixtures, analogous to AxB.

It is more difficult to see how the unshaded crosses are related to the same-variation or different-variation crosses in dominance or codominance. In fact, some of the results in OView 2-C have never been seen before. For example, a different-variation cross can produce 3 variations (with RxO or OxS) or 4 variations (with RxS), or 50% of a non-parental variation (with RxS), or 50% of two non-parental variations (with OxT). But all of these results can be explained with the "3 alleles and 1 codominance" model.

During the process of retroduction described above, notice that a combination of "theory and experimental system" is the immediate product of inference. Retroduction leads to the educated guess that "O is a codominant variation" and that "r and s are dominant over t," not simply that "there is one codominance and two dominances."

C. Model Revising

In Round 2, converting Mendel's Bible into a model for three alleles requires one major revision: the "# of Alleles/Gene in the Population" changes from 2 to 3. This major revision causes the "# of Possible Allele Combinations" to change from 3 to 6; this, in turn, produces a change in the "# of Variations/Trait," and many changes in the section labeled "States in the Simple Dominance Model."

D. Other Sub-Patterns for the Pattern of Multiple Alleles

OView 2-F shows the 7 types of sub-patterns of inheritance that can occur for the pattern of "multiple alleles" when there are 3 alleles. The fourth column of OView 2-F, labeled "4 varns

[variations], codominance, t loses" shows the subpattern that is always used in Round 2 of the MG course, with the genotype-phenotype mappings that occur between Tables 2-D and 2-E. For each genotype in the far-left border, the fourth column shows the corresponding phenotypes in bold fonts; the codominant variation is now labeled RS (rather than O) to show that it is co-influenced by the r and s alleles. But 6 other subpatterns are also possible, as shown in the other columns.

The first column shows another subpattern, hierarchical dominance, in which r is always dominant and t is always recessive, with a dominance hierarchy that can be represented as "r > s > t". Even though there are 5 other ways to do this (in addition to a rst hierarchy, there can be rts, srt, str, trs, or tsr), these are all the same type of inheritance subpattern because one allele is always dominant while another is always recessive. But in the second column there is a different type of dominance. This can be called "rock-paper-scissors" (abbreviated "r-p-s") by analogy to the game where "rock breaks scissors, scissors cut paper, and paper covers rock." In the "r-p-s" column of OView 2-E, r > s, s > t, and t > r; each allele has two relationships (one dominant, one recessive) with respect to the other alleles, so there is no hierarchy. In each of the next three columns there are 4 variations because one heterozygous genotype (rs) is codominant. For the two remaining heterozygous genotypes, rt and st, the t-allele can be dominant in both (so "t wins" if an allele is considered a winner when it succeeds in determining the phenotype), or dominant in neither ("t loses"), or in one ("t splits"). It is also possible to get 5 variations if two of the three heterozygous genotypes behave according to codominance; and 6 variations are produced if there are three codominances.[2]

If the teacher selects another subpattern (instead of "4 variations with one codominance") for Round 2, the characteristics of problem solving will change. For example, if the field population has 3 variations it will not be immediately obvious that codominance should be eliminated from consideration. With 3 variations, the possibilities for retroductive inference will also be different.[3]

---

2. For 6 of the 7 basic pattern-types, there are many ways to assign the dominances, as in the example above that describes the 6 varieties of hierarchical dominance.

3. The teacher's options are somewhat limited because GCK refuses to simulate 3 of the 7 sub-patterns. For example, if "3 alleles and 3 variations" is selected, GCK always simulates hierarchical dominance, never rock-paper-scissors dominance. And if "3 alleles and 4 variations" is selected, GCK always uses the "t loses" sub-pattern.

No matter which subpattern is used, if there are 3 alleles in the population and two alleles per genotype (one from each parent), there will always be the same 6 genotype combinations, so each of the 7 subpatterns in OView 2-F could use the same OView 2-D, although for some patterns (such as rock-paper-scissors) the phenotypes in OView 2-D will be "split up" (instead of "grouped together") which makes the table more difficult to use. But there can be 7 different versions of OView 2-E, corresponding to the 7 subpatterns for genotype-to-phenotype mapping. By splitting the genotype combinations (in OView 2-D) and the corresponding phenotypes (in OView 2-E) rather than combining them (as in OView 0`A or 1-B), the PO is made more flexible so it can be adapted easily to any of the 7 types of inheritance patterns. Of course, this type of table-splitting could also be done for Figures 15, 16, 18, or 19, but in each of these cases I think it is simpler and more meaningful to group the genotypes and phenotypes within one cell, in the top and bottom halves.

In the context of a multiple alleles pattern, "multiple" means 'more than two', so multiple alleles could mean 3 alleles in a population (as in Figure 17), or 4 alleles, or more. With more than 3 alleles, there will be corresponding changes in Tables 2-A to 2-F. With 4 alleles, for example, all of the tables will be much larger (OView 2-D has 100 cells!), and the number of trait-variations can vary from 4 (with all dominances) to 10 (if each of the 6 heterozygous genotypes is codominant) with a corresponding increase in the number of subpatterns.

B13: A Model for Round 3 — X-linkage

A. Anomaly Recognition

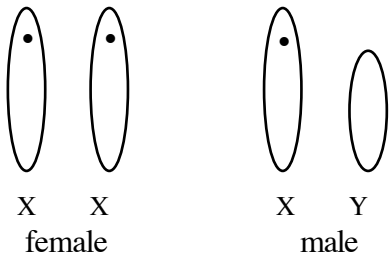
In Round 3 the students observe only two variations, but the models of codominance and multiple alleles each predict 3 or more variations. Therefore, these models need not be considered, and the only serious candidate for 'theory selection' is dominance. As shown in OView 3-A, the only anomaly is ZxH (with Z-mother and H-father), which never produces one of the expected results: 100% H. It is also surprising that, in contrast to the usual situation for duplicate crosses, HxZ and ZxH can produce different results.

When the data is split into females and males, other anomalies become apparent, as shown in OView 3-B, where H_f and Z_f represent females with variations H and Z, while H and Z males are H_m and H_f. First, although ZxH superficially looks like the familiar 50-50 mix that is sometimes seen with dominance in a cross of not-alike parents, now this 50-50 mix is produced in a new and unusual way; every female is H, and every male is Z. This first anomaly — observing a different phenotype ratio for females and males — has never been seen with dominance, codominance, or multiple alleles. Second, a related anomaly (or perhaps it is just another way of interpreting the same anomaly) is that with ZxH there is never a 50-50 mix of Z and H in either female or male offspring, as there would be with dominance. This second difference is less surprising, because we already have seen offspring ratios that are not 50-50 in Rounds 1 and 2 of model revising. In the same ways that HxZ is unusual, so is ZxZ. First, the familiar 75-25 ratio is produced in a strange way, with ratios that differ for females and males. Second, a cross of alike parents never produces a 75-25 mix of H and Z, for either females or males, as it would with dominance.

B. Anomaly Resolution

OView 3-C shows the predictions, for genotypes and phenotypes, that are made by a model of dominance, assuming that h is dominant over z. It was made by duplicating OView 0-B, except that the letters have been changed from a, b, A and B to h, z, H and Z. This new version (of the same old table) will make it easier to compare the predictions made by dominance (in OView 3-C) and by sex-linkage (in OView 3-D) because these tables are side-by-side, and because both tables now contains the same letters: h, H, z, and Z.

OView 3-D explains the anomalies described in OView 3-B. The essential explanatory concept is that a difference in one pair of homologous chromosomes determines whether a fly is female or male. As shown in the picture below, if a fruitfly has two large X chromosomes, it becomes a female; but if a fly has one large X and one small Y, it becomes a male.



The lower part of X and Y contain the same genes. But if a gene, such as • in the picture above, is located on a part of the chromosome that is present on the large X but is missing on the smaller Y, then each female will have two copies of the gene, and each male has only one copy. In OView 3-C the absence of a gene, in males, is represented by a –. Females, with two genes, can be either heterozygous (hz) or homozygous (hh or zz). For this gene the genotypes of males are called *hemizygous* because they have only one copy of the gene, which is half the usual number. An H phenotype can be produced (in females) by hz or hh, or (in males) by h-; and Z is produced (in females) by zz or (in males) by z-.

The shapes of Tables 3-C and 3-D are different because with X-linkage there are only two types of fathers; more specifically, there is only one type of H-father. This is one reason why anomalies occur with H-fathers but not Z-fathers. Another reason for a lack of Z-father anomalies is that, with or without X-linkage, with dominance the allele contributed by a Z-father (whether it is z or -) does not affect the progeny phenotype, which is totally determined by whether the mother contributes an h or z. But with an H-father, there are no hh-males so the entire left column of 3-C (with an hh-father) is missing from 3-D. And because it doesn't matter whether a father contributes a z or -, the hz column (in 3-C) and h- column (in 3-D) produce the same phenotypes (HHHH, HHHZ, HZHZ) except that in the h-column the table entries are segregated by sex; all progeny on the left side of the column are females, while all progeny on the right side are males, so for two of the three possible genotypic matings the ratios within each sex-group (shown in OView 3-B) are anomalous even though the overall ratios (in OView 3-A) are not anomalous.[4] Narrowing our focus to ZxH, it produces anomalous results for two reasons: because there is only one genotypic

---

4. The 'expected data' ratios calculated in Table 25-A assume that a father is equally likely to contribute an X or Y chromosome, so there will be an equal number of females and males in the progeny.

mating (compared with two for HxZ), and because with an h- father the selection of an allele determines both the phenotype (whether a progeny is H or Z) and sex (female or male) so there are only H-females and Z-males.

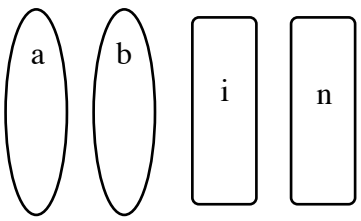
C. Model Revising

For a model of X-linkage there is one major change, which shows up directly in three places in Mendel's Bible: the "# of Alleles/Gene in an Individual" is still "2" for females but now it is “1” for males, because the "allele combinations" (which are unchanged for females) are now “1,- or 2,-” for males, which causes two additions to the "Genotype/Phenotype Combinations." Due to this major change the "Cross Possibilities" must be revised, using OView 3-B (derived from 3-D) as a guide.

B14: A Model for Round 4 — Autosomal linkage

For simplicity the ‘autosomal linkage’ problem will be called "Round 4" even though, as discussed in Section 3.25_C, the ‘X-linkage’ and ‘autosomal linkage’ problems are not always used in a “1, 2, 3, 4” sequence in the MG classroom.

In Round 4, two traits are being studied, so the genotypes (and the analysis) are more complex. In the following explanation, one trait (with variations A and B) is controlled by one gene (with alleles a and b), while another trait (with variations N and I) is controlled by another gene (with alleles n and i). If these genes are located on two different chromosomes, as shown below for a doubly heterozygous “abin” mother, in mitosis if allele **a** is selected there is an equal chance that either **i** or **n** will also be selected; this is called ‘independent assortment’. And if **b** is selected there is an equal chance that **i** or **n** will be selected. Therefore, all four possible genotypes (ai, an, bi, bn) are equally likely to be passed on to the progeny.



The right side of OView 4-A shows the resulting genotypes and phenotypes for a cross of AIxAI that involves two 'abin' parents. As described above, four gametes are possible, so there is a 4x4 'Punnett Overview' (a Punnett Square plus phenotypes) with 16 possible genotypes and the 16 corresponding phenotypes: 9 are AI, 3 are AN, 3 are BI, and 1 is BN. As indicated below the right-side table, the phenotype ratios for AI:AN:BI:BN are 9:3:3:1. The left-side table shows another AIxAI cross, but with two 'aaii' parents. Since only one gamete is possible for each parent, this "Super-Punnett" (SP) representation is much simpler, and so are the resultant phenotype ratios of 1:0:0:0.

OView 4-B shows the phenotype ratios for every possible combination of parental genotypes. The ratio in each cell was found as illustrated in OView 4-A. For example, the phenotype ratio for "aaii x aaii" was found to be 1:0:0:0, which is the same as 8:0:0:0, abbreviated "8000" in the first cell. Similarly, the cell for "abin x abin" (in the fourth row, fourth column) has "9331" to indicate the 9:3:3:1 ratio found above. For each of the 81 cells, the phenotype ratio can be found by using a Punnett Table with 1, 2, 4, 8, or 16 cells.

Of the 81 cells in OView 4-B, 7 are shaded to show the anomalies that occur when there is autosomal linkage, when the genes controlling two traits are located on the same chromosome. The picture below shows that when two genes are on the same chromosome there are only two possibilities for an "abin" female:

one type of female another type of female

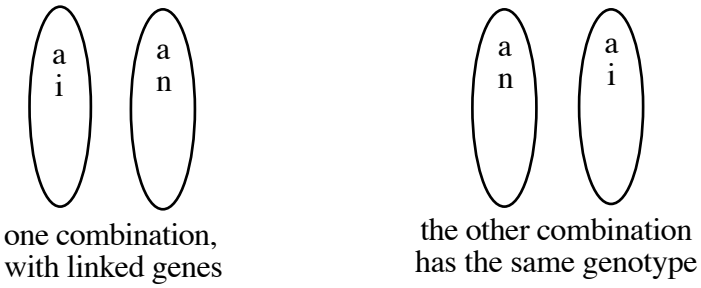
Similarly, there are two types of abin-males, with "ai bn" or "an bi" combinations. OView 4-C shows the Punnett Square possibilities for crosses of four females (abin, abnn, bbin, bbnn) with an abin-male. For the first Super-Punnett (abin x abin, AI x AI) there are four possible genotypic crosses, and four results for "AI:AN:BI:BN ratios", abbreviated as 6002, 4220, 4220, and 4220. If

each of these results occurs equally often the overall ratio will be 9331, the same as without linkage, even though a 9331 ratio never occurs in any collection of progeny. These four ratios (6002, 4220, 4220, 4220) are entered in OView 4-D in the appropriate cells (in the 4th and 5th columns for the abin-father, the 4th and 5th rows for the abin-mother); these four shaded cells correspond to the one shaded "abin x abin" cell in OView 4-B. The other three shaded cells in the abin-father column of 4-B correspond to the six shaded cells in the two abin-father columns of 4-D; for each of these six cells the ratio was found using the Punnett Square Overviews in 4-C. The three "overall ratios" (3311, 3131, 2222) match the non-linked ratios predicted in 4-B even though (as with "abin x abin") the predicted ratio never occurs in a collection of progeny for any cross of parental genotypes.

The analogous crosses for an abin-mother produces similar results; reading down the two columns for an abin-father, ratios are the same as reading rightward across the two rows for an abin-mother, which correspond to the "one type of female" and "another type of female" in the pictures above. The right side of OView 4-F contains the same data as OView 4-D; for example, the 20 cells in the AIxAI area of 4-D contain the five different ratios shown in the AIxAI area of 4-F. Similarly, the 16 cells in the AIxAI area of 4-B contain the four ratios in the AIxAI area of the "predictions" table at the left side of 4-F. The predictions that don't occur (9331) and anomalous observations that do occur (4220, 6002) are in bold-face type.

The Super-Punnett overviews in OView 4-E, for "abin x abin" without linkage (on the left, as in 4-A) and with linkage (on the right, as in 4-C) show why the overall ratios are the same in both cases. In fact, if the order of genotypes in 4-A is rearranged (from 'ai an bi bn' to 'ai bn an bi') the only difference between this "revised 4-A" and 4-C is that 4-C has bold lines dividing its 16 cells into four areas, while the revised 4-A doesn't have the bold lines.

In OView 4-D, why is only one genotype column (for abin) split into two columns that contain shaded cells? Because this is the only genotype for which the cross-results depend on whether genes are non-linked or linked. For example, with aain each of the two possible combinations produces the same genotypes, as shown below:



one combination, with linked genes

the other combination has the same genotype

A 4x4 Super-Punnett (SP) for a linked "aain x aain" would be similar to the 4x4 for the linked "abin x abin" in 4-C, except that now all four areas produce a ratio of 6200, as shown in 4-D (2nd row, 2nd column). With non-linkage the SP can be a 4x4 (as in 4-A) or a simplified 2x2 with two rows for the mother (contributing either ai or an) and two columns for the father (ai or an); genotypes in the four cells of this 2x2 are identical to the genotypes in the four cells for any of the four areas (which are all identical) in the 16-cell SP for linked genes.

OViews 4-G and 4-H show another way to calculate the predicted ratios for non-linked genes, by using phenotypic probabilities instead of genotypic Punnett Squares. The first cell of 4-G, AIxAI, shows that two results are possible for AxA (it can produce 25% B, or 0% B) and also for IxI (either 25% N, or 0% N). These possibilities can be combined in four ways, as shown in the AIxAI cell of 4-H, to produce four results (8000, 6200, 6020, 9331) that are the same as the ratios in the AIxAI cell on the left side of 4-F. For each of the 16 phenotypic crosses the probabilistic predictions (in 4-H) are the same as the Punnett predictions (in 4-F).

Model Revising? With two traits instead of one, even without linkage a Mendel's Bible looks different than Figure 8. The principles haven't changed, but there is more of everything, and it is more complicated. For example, now there are 10 non-duplicate phenotype crosses instead of 3, and 45 non-duplicate genotype crosses instead of 6. With linkage, one principle changes; now there is no "independent assortment" because the necessary condition — that "genes...are found on separate chromosome pairs" — is not satisfied.

Experimental Design? After a theory of autosomal linkage is developed, OView 4-D can help design a crucial experiment. With 4 of the 45 genotype crosses, anomalies are always observed. But with the other 41, anomalies are never seen. Interpretation requires prediction, which is

difficult with two traits, especially if all possibilities (at the levels of both genotypes and phenotypes) are considered. Although doing all 10 phenotypic crosses requires discipline it is fairly easy. But observing all possible results is very difficult; it requires patience, and a large field population to insure that each of the genotypes (9 or 10, with or without linkage) is present. Crossing a doubly recessive parent (bbnn) with a doubly heterozygous parent (abin) is easy to do (because the genotype of one parent is known from its phenotype) and interpret (because the PS is relatively simple, and the anomaly is obvious); in the MG course the teacher may suggest this experiment if students cannot recognize any anomalies.

B15: A Prediction Overview for "3 Alleles per Individual"

During Rounds 1 and 2, students sometimes propose a theory in which each individual has 3 copies of a gene, instead of the "2 alleles/gene in an individual" in Mendel's Bible. With this theory it is impossible to make predictions using a standard Punnett Square or Super-Punnett. But predictions can be made with the modified Super-Punnett shown below, for a theory that postulates "2 alleles in the population" and "3 alleles per individual" with progeny receiving two alleles from the mother, and one allele from the father. The far-left border shows that for a aaa- or ddd-mother there is only way to contribute 2 of the 3 alleles, but with aad there are two possibilities (aa, ad) for passing on two alleles, and with add there are two possibilities. Similarly, with aaa or ddd there is one way to pass on one allele, but with aad or add there are two possibilities. The phenotypes change from A to D as the proportion of d-allele increases.

B16: A Comparison of Three Symbol-Systems

Many symbol systems are used in classical genetics. In one common representation the recessive and dominant alleles are symbolized by lower-case and capital letters, respectively. Often, the allele-letter matches the first letter in the dominant variation; for example, a pea with a seed coat that is yellow is produced by alleles YY or yY, but not yy. The advantages of this system are an easy recognition of whether an allele is dominant or recessive, and the connection between alleles (y and Y) and their trait-variation (Yellow). But there is no such connection for the

recessive variation; “yy causing green (not yellow)” is not very intuitive. In this case the PO system of small/capital symbolism is more useful, with y and g (small alleles in a genotype) causing Y and G (large phenotypes).

The conventional ‘y and Y’ system is useful for describing dominance, but it runs into difficulties when the relationship between alleles is not recessive/dominant, or with more than two alleles in a population. For these situations the symbol systems used with Prediction Overviews or in Mendel's Bible are more flexible and useful. Also, it is difficult to quickly make a Super-Punnett by mentally thinking as one fills the squares, “cap-y, small-y, cap-y, small-y”; it is easier to think clearly when thinking “a, b, a, b” or “1, 2, 1, 2.”

Of course, any system faces a challenge when one gene causes (or influences) several traits, or when one trait is influenced by several genes. In these cases, neutral symbols can be used; for example, m and n might cause yellow and green, and also short and tall,... By using letters for traits but numbers for alleles, the “Mendel's Bible system” avoids the difficulties (and also the advantages) of forming mental associations between allele-letters and variation-letters.

With any system it is a challenge to cope with a new field population that has unknown characteristics. For example, if ‘1 and 2’ (or ‘a and b’ or ‘Y and y’) have been used for ‘dominant and recessive’, using these in a new situation where 1 is not dominant (or where neither allele is dominant) can be misleading. But this association can be overcome with mental flexibility, or by using new symbols such as ‘3 and 4’ or ‘c and d’. With a PO, if the eyes of flies are ruby and white, alleles (r, w) and traits (R, W) can be assigned immediately, and the relationships (dominant, recessive, codominant,...) can be figured out later. And if there is a third variation that is codominant, “rw causing purple” is similar (in its lack of a symbolic connection between alleles and variations) to “34 causing purple.”

****

Appendix B2: Actual Problem-Solving Actions

Sections B10-B15 discussed problem-solving strategies that students potentially *could use* when solving problems in the MG course.[5] Sections B21-B22 will examine the strategies and actions that students actually *do use.* The difference is thus between *potential* strategies and *actual* strategies. In order to understand, at a deep level, what students do during GCK model revising, it is necessary to know the logical possibilities; developing this understanding of the available 'logical tools' is the purpose of B10-B16. The methods used for the analysis in B21-B22 were described in Section 3.28_A. The following section just describes the data sources.

B20: Four Sources of Empirical Data for the Analysis

Figure 21 summarizes the main features of doctoral dissertations, done at the University of Wisconsin-Madison, that have studied GCK-based problem solving.

All tables for Appendix B2 (Figures 21 & 22) are in the "tables.rtf" file.

The recent work is more useful for my analysis of the Monona Grove (MG) genetics course. Collins and Thomson studied GCK-based problem solving, but not in the context of the MG course. Hafner observed students enrolled in the MG course while they solved, outside the classroom as individuals, the same types of problems used in the course. The four most recent studies — by Finkel, Wynne, Lemberger, and Johnson — observed students solving problems in their research groups in the MG classroom. Since my objective is to examine instruction in the classroom, these four studies were used as the main sources of information for the analysis that

---

5. In philosophy the analysis of a historical event by developing a scenario for the process-of-logic that potentially could have been used (even though it is not necessarily the logic that actually was used) is called a 'logical reconstruction'. This approach was used by Darden (1991), who is wary of making definite claims about unobservable cognitive processes, in her analysis of thinking strategies used by classical geneticists, ca 1900-1920.

follows. The following summaries are only a brief introduction to each study. The observations, analyses, and conclusions of each researcher are described in Section B22 whenever these contributions are relevant for the discussion.

Elizabeth Finkel (1993), using a 'sociology of science' perspective, examined social and intellectual interactions, and the problem-solving strategies used by students; this research is summarized in a paper by Finkel (1996). *Cynthia Wynne* (1995), using perspectives including Clement (1989), focused on students' use of a meiotic model. *John Lemberger* (1995), using the Conceptual Change Model (Posner, Strike, Hewson & Gertzog, 1982; Hewson, 1981), studied the ways conceptual learning is influenced by the nature of classroom instruction. *Susan Johnson* (1996) studied the similarities and differences between the problem-solving strategies used by research scientists — as described by Darden (1991) — and the strategies used by students in the classroom. The same class was studied by Finkel and Johnson, who offer two perspectives on the same events.

Two additional sources of information, which will not be cited, are my nine weeks of classroom observation in 1993 (for the same class studied by Wynne), and interviews with two of the three course developers, Sue Johnson and James Stewart.

B21: An Overview of the Analysis

This section, even though it was written after the detailed analysis in Section B22, is presented first because it provides an overview. During the analysis in B22 the major types of student actions will be highlighted in bold print. B21 organizes these major actions into the 17 categories used for the activity-and-experience grids in Chapter 3.

References to Section B22 will be enclosed in brackets. For example, "...external memory storage {F, J, K}" indicates that this action is described in Subsections F, J, and K.

Preparation: If essential content-knowledge (or process-knowledge) is not cognitively available, problem solving effectiveness will decrease. This can be due to a lack of internal memory storage { F } when students either never learn concepts, or cannot remember them. The

teacher encourages students to use external memory storage { F, J, K }, such as Mendel's Bible, by often referring to the whole bible or (as in Subsection J) to a specific part of it.

Posing: As discussed in Section 3.44, students do some posing of constraints { D }, but no posing of an area to study.

Probing: Students constantly make decisions about pursuit { D, M, N } — about what to do next and how to use time. These decisions can involve the balance between observation and interpretation { in M }, a choice of which theory to pursue { D, N }, or a mega-action that involves several component actions, such as testing a theory { M } by designing experiments, doing experiments, and evaluating theories for either pursuit { E, S } or acceptance. Invention (of possible actions) precedes evaluation (of these possible actions), which is followed by execution of a selected action. Sometimes there are serendipitous opportunities { C } that students, if they are alert, can use to their advantage.

Selection of Theory: Because the teacher tells students that they will be revising their existing models in each round, students "expect to be surprised" and the psychology of selection-versus-invention { C } is different than in most research science.

Invention of Theory: Students can invent theories by using strategies of analysis-and-revision { E } or synthesis { L }, or by resolving multiple empirical constraints one at a time { E }. When making revisions, students often protect certain components from change; depending on the situation, and whether tradition or innovation is needed { I }, protecting a component { G, I } may or may not be an effective strategy. A theory that is retroductively invented has passed the first stage of evaluation; it can then undergo 'testing' { E, M } that evaluates it more thoroughly. Retroductive inference can result in the revision of either a model or a system characterization { T }. Invention (and strategies for inventing) can result in the production of a new theory or a new idea for a pursuit-action { D, M }.

Experimental Design: A goal-oriented design of experiments { W } can have a general goal (to gather lots of data by doing all possible experiments, to use for retroductively inventing models and system-theories) or a specific goal that is inspired and guided by a specific theory { B }. A "do it all" strategy for experimental design is common because GCK experiments are quick-and-easy,

costing little time and no money, so there is not much incentive to do ‘thought experiments’ for experimental design { N, W }, in contrast to most situations in research.

Do Experiment and Make Observations: Although students never do physical experiments, GCK generates data as if an experiment had been done. Sometimes alert observation is necessary { C }, especially if, in a new situation, an important characteristic of the data has not been important in the past.

Predictions: A prediction can be made { B } by using deductive logic (based on an explanatory theory or a descriptive theory) or inductive logic (with or without a descriptive theory). When testing a newly proposed model { M }, making predictions based the new model is one of the most important, and most difficult steps.

Degree of Agreement: To recognize anomalies { B, D }, to retroductively invent a model or system-theory { T }, or to test and evaluate a model { N }, degree of agreement is the main evaluation criterion { O }. For many GCK experiments, accurately estimating a degree of agreement requires the use of statistical logic { C }.

Degree of Predictive Contrast: A ‘crucial experiment’ { W } is rarely planned in advance; more often, after it is done (and maybe after a new theory has been invented) an experiment is recognized to have discriminatory power due to a high degree of predictive contrast.

Previous Experiments: If a model has been used to make accurate predictions in previous experiments, involving a variety of systems, there may be good reason to question the validity of an apparent anomaly { P }.

Internal Characteristics of Theory: Students generally prefer an explanatory theory (which postulates a composition-and-operation mechanism to explain what is being observed) over a descriptive theory that, even though it can make accurate predictions, does not postulate a causal mechanism { U }. Other factors sometimes used as evaluation criteria include systematicity with an absence of suspiciously ad hoc components { O }, and a simplicity { Q } that avoids the use of an unusually large number of genotypes in a mechanism that is intended to causally explain a much smaller number of phenotypes. // Another factor — considered mainly by the course designers, not by students — is structural modularity { E } in making a Mendel's Bible that can be easily

revised by students.

External Relationships with Other Theories: Although degree of agreement (an empirical factor) is the major criterion used by students in evaluating models, often the consistency of components between models (a conceptual factor) is also important. { H, I, K } As discussed above, in some situations, but not others, it is beneficial for theory invention { G, I } when a component is protected from revision in order to avoid conceptual inconsistency. As discussed in Section $3.33_{B,}$ 'consistency of components' can be viewed as either internal consistency or external consistency, when viewed from different perspectives about the relationships between mega-theories, theories, sub-theories, and theory components.

Metaphysics: Students make a metaphysical assumption of 'consistency', that identical experiments should produce identical results, but in practice this criterion must be applied with caution because { S } experimental systems that appear to be identical (at the level of observable phenotypes) may not really be identical (at the level of unobservable genotypes).

Personal Factors (Psychological and Practical) and Authority: Students interact with their scientific colleagues (their fellow students) by competing and cooperating, in ways that can be either detrimental or motivational { Z }. Students develop strategies for "how to do science in the classroom" based on their ideas about both science and the classroom; these ideas can be described as a context-dependent 'thought style' that operates in the classroom and in each group, that influences students' approaches to experimenting and theorizing { I, Q }. An important part of any classroom context is the authority of the teacher { J }; in this classroom a model can also have authority, in part due to its external representation { F }.

Evaluation and Conclusion: Students can evaluate a model based on estimates of its plausibility and/or utility { I, O }, for purposes of acceptance or pursuit { E, R, S }. In addition to definite conclusions to retain, reject or revise, there can be an inconclusive 'delayed conclusion' { E, X } while more data is being collected and interpreted, or there may be a reversal of an earlier decision. Usually the recognition of anomaly is a serious threat to a theory, but sometimes { P } students are willing to retain a theory anyway, at least temporarily. In addition to evaluation of theories, students can also evaluate options for pursuit-actions, as described earlier in the "probing" paragraph.

Persuasion: Internally-oriented persuasion, inside a student or within a research group, occurs frequently during model-revising problem solving, for the purpose of evaluating both theories and options for pursuit. Externally-oriented persuasion is sometimes a goal (of the students or the teacher) during interactions between the students and the teacher.

B22: An ISM-based Analysis of Problem-Solving Actions

As explained in Section 3.28, GCK model revising is the central activity in the MG course, so it is examined in greater detail than the other activities, using empirical data instead of just educated estimates of the "probable actions" that seem necessary when students solve problems. Section 3.34_A is a condensed version of the analysis in B22, which was completed first and describes student actions in more detail.

Section B22 is a topical analysis of student problem-solving actions. It expresses my own ideas and incorporates insights, empirical evidence, and illustrative examples from the work of previous researchers. This analysis does not attempt to be exhaustive, but instead will focus on those empirical observations and theoretical claims that I consider the most interesting and important. I will also try to describe a 'cross section' of student actions that is a roughly representative sampling of what students do. The main content of B22 is summarized in B21.

In B22 the major student actions are summarized in phrases highlighted with bold print. Usually there will be a reference, enclosed in brackets, to a section in the Chapter 2 elaboration where this action is described. Although references to the ISM-elaboration are cited, all actions are also contained in the shorter ISM-framework.

My analysis assumes, as background knowledge, the Prediction Overviews (and descriptions of potential logical strategies) in Sections B10-B15. To be consistent with the use of 'population' and 'model' in the existing literature for the MG course, I will refer to 'field population' instead of 'field collection'; and the terms 'model' and 'theory' will be used interchangeably, in contrast to their distinctive meanings in Chapter 2.

A. An Overview of the Problem-Solving Process

Before examining the process of problem solving in detail, it will be useful to sketch a brief overview, beginning with the framework of Finkel (1996) who describes 3 types of knowledge used by students: knowledge of *genetics,* of *model revising,* and of *one's own actions.*

> Knowledge of genetics was initially used...as a way to recognize anomalous aspects of new field populations,...to identify problems that a new model of inheritance must resolve. ... [It] was also used as students began to develop alternative models to explain aspects of the data characterized as anomalous. (pp. 355, 357)
>
> Knowledge of model revising was used by students to structure the work that they did. ... This knowledge determined what students *did*...and provided a framework within which they developed and used the conceptual knowledge of genetics. (p. 357)
>
> Metacognitive knowledge consists of students' understanding of what they are doing as they model-revise. ... It allows students to keep track of what they have done and what they intend to do next. ... [And it] helps them make connections between knowledge of genetics and knowledge of model revising, and thus helps students develop stronger conceptual understandings of genetics and model revision. (p. 362)

As students use their knowledge to revise models, a simplified summary of their strategy is to "find what is wrong and fix it" or, more specifically, to "recognize an anomaly and resolve it." This two-step process forms a basic structure for my analysis. A brief look at anomaly recognition, in the following two subsections, is followed by a more detailed analysis of anomaly resolution.

B. Anomaly Recognition

To find an anomaly, students check for **degree of agreement** { 2.12_A } between predictions and observations. **Predictions** { 2.11_D } can be made using either deductive logic (based on a theory that is explanatory or descriptive) or inductive logic (based on an explicit descriptive theory, or on prior experience). Anomalies that occur during GCK problems used in the course are described in the "Anomaly Recognition" parts of Sections B11 to B14. This section provides an overview. There are four basic types of anomalies: an unexpected number of trait-variations in the field collection, a cross that produces unexpected ratios for the trait-variations, a cross where ratios differ in males and females, and a 'missing result' when (despite the normalcy of results that do occur) an expected cross-result never occurs.

In Round 1 the first anomaly to be recognized is always the presence of 3 variations in the field

collection. Initially, this recognition is due to the difference between observations and prior experience, rather than a conscious deductive prediction, because at the beginning of Round 1 most students will not have thought much about the reasons that Mendel's model predicts two variations, but based on experience there will be an assumption that this is what they should see. { **prediction by deduction or induction**, 2.11_D } After this initial recognition, students can examine their Mendel's Bible and see that, yes, their model does predict two variations. Then, when students begin doing crosses, CxE and DxD (using the symbolism of Section B11) are easily seen to be anomalous. Most groups recognize several anomalies in Round 1.

In Round 2, again the most obvious anomaly is the number of variations in the field population (now there are 4 instead of 2 or 3) and all groups quickly see this. Following this, a defining of anomaly depends on a group's theory-based expectations. For example, some allele combinations are governed by dominance, and others by codominance; a group's initial model usually involves dominance, so a codominant cross is seen as anomalous; but if students expected codominance a dominant cross would be viewed as anomalous.

In Round 3, when a problem involves X-linkage with dominance, students must recognize that ratios differ for males and females in all ZxH crosses (but never in HxZ) and in some (but not all) HxH crosses. Beginning in 1994 two types of problems were used in Round 3, and groups working side-by-side could be solving different problems. In the second type of problem, X-linkage with codominance, anomaly recognition is usually easy because when students try to complete the common strategy of "doing all crosses" they cannot do DxD because there are no males with the heterozygous genotype that produces D.

In Round 4, with two-trait autosomal linkage, it is tough to find anomalies, even after Sue advises students to "consider both traits together." This hint may be difficult to understand and use, and sometimes students "struggle with what the teacher meant by telling them to consider the two traits together. (Finkel, 1993, p. 245)" If necessary, the teacher may even suggest trying a specific cross, typically the 'abin x bbnn' that, although quite distinctive in producing an unexpected result, is a difficult experiment to intentionally design until after a model of autosomal linkage already has been invented and understood. { **theory-based experimental design**, $2.61_{A\text{-}C}$ — but this action is done by the teacher, not students }

C. Serendipity, Surprise, Alertness, Statistics

Anomaly recognition requires the generation, observation, and interpretation of data. If anomalous data is never generated by performing an appropriate experiment, such as the 'abin x bbnn' cross in Round 4, there can be no recognition of anomaly. Sometimes an experiment done for one purpose will be useful for a different reason. { **serendipity** } For example, with an 'X-linkage and codominance' problem in Round 3, when students use the common strategy of doing all possible crosses to search for surprising results, the big surprise is that it is impossible to do one cross because for one trait-variation there are no males! Another example of an obstructed experimenting strategy occurs in Round 1 when a student "decides to create pure strains of each variation, a strategy not often used by any of the groups. (Johnson, 1996, p. 71)" But the intended function of this strategy is thwarted when students realize that DxD always produces three variations — never the 100% result that sometimes occurs with every type of 'same variation' cross in Mendel's model — but the strategy does lead to an unplanned opportunity for anomaly recognition. This is a rare case of anomaly recognition due to a 'missing result'; usually anomaly recognition occurs due to an 'unexpected result'.

Although scientists in their normal work do not expect to be surprised, students in this class know that anomalies will occur so they expect the unexpected. { This produces a difference between student experiences and those of typical working scientists, due to a shift in the balance between expectations for **selection versus invention**, 2.51. } For example, Finkel (1993, p. 202) describes an interaction where one group member "returns to his earlier idea that this problem is 'the normal like two trait, simple dominance' and this idea is rejected vehemently by F who is sure that because this is 'model revising' the answer can't be a model that they have already worked with."

Students will miss an opportunity for anomaly recognition if an important characteristic of the available data is not "mentally observed." But the act of **alert observation** { 2.73_C } is sometimes difficult for students. For example, the two essential principles in Round 3 — that ratios can differ with sex, and that 'duplicate crosses' such as HxZ and ZxH may produce different results — have not been important in the students' previous experience with Mendel's Model and in Rounds 1 and

2. Therefore, in Round 3 students must somehow learn to pay careful attention to these newly important details when they observe data and design experiments. Similarly, in Round 4 students must internalize new strategy-principles such as learning how to "consider both traits together" and knowing why this is important.

When evaluating whether an observed ratio agrees with a predicted ratio, an understanding of statistics is helpful. { **degree of agreement using statistics**, 2.12_{A} } Lemberger (1995, p. 102) reports the insight of a group who "fortunately...realized that they wouldn't always get statistically perfect results." And Wynne (1995, p. 52) explains that

> [Some ratio-anomalies] were abandoned because they were no longer considered to be anomalous. Usually, after groups generated more data (and thus looked at a larger sample of the population), they determined that the observed ratios were closer to the expected ratios than they had previously thought. This is to be expected, as it is usually true that observed ratios are more likely to match expected ratios as a sample size increases.

D. Connecting Anomaly Recognition with Anomaly Resolution

The use and non-use of anomalies is described by Wynne (1995, pp. 51-52):

> Not all anomalies were pursued, however. ... Anomalies were considered to be abandoned when (a) groups did not mention them again, and/or (b) groups never generated a hypothesis in an attempt to explain the anomaly. ... Recognized anomalies were abandoned for different reasons, including: (1) groups found another anomaly that they chose to pursue instead, and/or (2) the anomaly was no longer considered to be anomalous.

{ **theory evaluation for pursuit**, $2.71_{A\text{-}D;}$ **degree of agreement** } There are several ways to reach a conclusion of "no longer anomalous"; one way, described above, is if observed ratios become closer to predicted ratios when additional data is collected.

After anomalies are recognized, students must decide what to do next. { **invention-and-evaluation of pursuit actions**, $2.71_{B,D}$ } Lemberger (1995, p. 204) reports that,

> Of the five groups in round one for which there is data, four recognized the anomaly without the aid of the teacher, but all five groups needed help in posing a problem [i.e., in 'constraint-posing', as discussed in Section 3.44]; ... this changed in subsequent rounds of the problem-solving sequence as the student groups became more experienced.

Students improve their "**knowledge of model revising**" (discussed in Section $B22_{A}$) as they learn the strategy of "anomaly recognition followed by anomaly resolution" that is used in the course, over and over, to guide decisions about "what to do next."

E. Anomaly Resolution by a process of Invention-and-Evaluation

The process of anomaly resolution combines invention and evaluation. To invent new models, students revise parts of their initial models, as summarized in Mendel's Bible. { **invention by analysis-and-revision**, 2.53_A } The required changes are described in the "Model Revising" parts of Sections B11-B14. A brief summary follows. For Round 1 the major change is that for the heterozygous genotype (1,2), instead of dominance (which produces the same phenotype as 1,1) there is 'codominance' to produce a third phenotype. For Round 2 the major change is from 2 alleles in the population to 3 alleles; a minor yet important change is from dominance (in Mendel's Bible) or codominance (in Round 1) to the use of both patterns. In Round 3, for females (with X and X chromosomes) the trait is influenced by two alleles, but for males (with X and Y chromosomes) there is only one allele because the trait-gene is present on the X chromosome but is missing from the smaller Y chromosome. In Round 4, instead of independent assortment (with genes on different chromosomes) there is autosomal linkage (with genes on the same chromosome).

A key insight into the process of "invention by analysis-and-revision" comes from Wimsatt (1987): for the purpose of resolving anomalies, a model is more useful for model-revising if this model and the associated methodology (the experimental and heuristic tools available for analyzing the model) are structured in a way that allows the localization of anomaly to specific components of the model, which can then be revised to produce a new model that resolves the anomaly. This characteristic — a structure that lets a user localize the anomaly and then resolve it — has been intentionally designed into Mendel's Bible. { 2.53_A, the **logical structure** of the Mendel Bible representation was developed by the course designers, not by students } The primary strategy for anomaly resolution, which the teacher encourages and students actually use, is to revise parts of Mendel's Bible, which — by serving first as a standard for defining expectations and *recognizing* anomalies, and second as a template for inventing new theories that can *resolve* these anomalies — facilitates the two main steps in model-revising problem solving.

In their effort to resolve anomalies, students often select one anomaly to serve as a focal point for the retroductive invention of a revised model that can resolve this anomaly. { **retroduction with multiple empirical constraints**, 2.52_E } Then the new model produced in this invention

stage, which already has undergone an initial narrow evaluation during retroduction, is submitted to a more comprehensive evaluation to determine if it really does resolve the original anomaly, and whether it avoids new empirical anomalies and conceptual difficulties. { **retroduction plus testing** for a more thorough evaluation, 2.52_C }

But this two-stage process of invention (inspired and constrained by evaluation) followed by evaluation (of a newly invented model) is rarely the neat, organized process implied above. In real life, anomaly resolution is usually a messy, meandering process because with a challenging problem it is rare for students to invent, in their first attempt, a theory that can be quickly and thoroughly evaluated as empirically and conceptually adequate. Instead, evaluation usually shows the first 'new model' to be inadequate. Or the process of evaluation may be slow and difficult, spanning a long period of time in which there is no definite yes-or-no conclusion. During this period of uncertainty students may propose other new models, including variations of their own previously invented models, and there will be an overlapping of the evaluation phases for the many models still being considered. { **delayed conclusion**, 2.42; also **competitive theories**, **relative status**, **evaluation of theories for pursuit** } The remainder of Section B22 examines the meandering process of anomaly resolution in detail.

F. Memory for Models

Why is it often difficult for students to see the changes that will produce anomaly resolution? Partly because students typically lack a deep understanding of existing models. For example, Johnson (1996, p. 149) observes that for the McClintock group,

> Amazingly one of the crosses that continued to trip them up was the one in which two hairless parents produced offspring that were a mix of three phenotypes, hairless, tiny and spineless. This is the classic cross of the heterozygous individual in the codominance model.

Two years later a similar response to this phenomenon is reported by Lemberger (1995, p. 143), "the appearance of a third variation 'outa nowhere' confused them"; this time the surprise occurs during Round 3 after the group has worked with codominance in both preceding rounds. { **memory**, 2.73_B } If these students had learned one of the major lessons from Round 1 by internalizing this distinctive phenomenon into their mental models of genetics, the relevant

information — which would have inspired them to propose codominance as one component of a new model — would have been easy to retrieve from memory. Instead, their conceptual knowledge is weak and this option is overlooked. In order to solve problems successfully, memory is not sufficient but it is necessary.

As discussed in Section 2.73_B, memory storage can be either internal or external. In the following excerpt the teacher encourages students to bolster their sometimes-inadequate memory of genetics knowledge by using their **external memory storage**:

> 'You guys are actually using those models already, but you're using them in your head, and so sometimes you seem to be forgetting parts of them. ... You aren't starting from scratch. You're using the ideas in your head. But have those sheets of paper [Mendel's Bible, and the Codominance Bible that was developed in Round 1] out. Those two sheets of paper out so you can be a little bit more organized about what might we change. And how might we change it.' (Lemberger, 1995, p. 234)

In offering this general suggestion the expectation is that once these external supports have enriched the students' memories with the relevant genetics concepts, the invention of new models will be less constrained by a lack of conceptual understanding. Even without explicit suggestions to "look at your bibles," Finkel (1993) points out that the mere existence of these formal representations for certain models (Mendel, and maybe codominance,...) helps to give these models higher status, thus making it more likely that students will learn the models more thoroughly and will use them in future rounds of model revising. { **authority** supporting a model, 2.31_C, due to its external representation }

G. Conceptual Constraints on Thinking

Sometimes part of the difficulty is caused by conceptual constraints acting as a 'cognitive block' that impedes the perception or acceptance of a potential option. In the example above a lack of memory about the distinctive DxD cross probably played a role in misinterpreting the data in Round 2, but perhaps another contributing factor was a reluctance to abandon the concept of dominance. { **protected components** as a hindrance to free creativity, 2.52_A, 2.72_H, 2.73_C } McClintock also had difficulty with giving up dominance in Round 1:

> When McClintock saw "CxE → 100% D] they seemed to recognize that the results were unusual, but they still didn't realize that it was an anomaly requiring a revision to the concept of dominance. Instead they continued to try figure out which of the three variations were

> dominant and which, recessive. As with other student groups the prior model of simple dominance acted as a constraint to the model revision process. ... [In all rounds, many groups had similar difficulties.] Even though all groups were now aware of the Mendel model and codominance models, most groups tried to apply Mendel to cross results, even when a particular cross was one of the two that so defines codominance by the types of offspring produced. ... As was the case with the other groups they [Crick] had a difficult time giving up the idea of dominance and continually attempted to describe the phenotypes as dominant or recessive. (Johnson, 1996, pp. 136, 170, 162)

This conceptual constraint can be either unconscious (to prevent a perception that changing the protected component is an option) or conscious (to prevent acceptance of this option). As discussed in Section $2.73_{C,}$ there is a close cognitive connection between invention and evaluation, between rejection that occurs during retroductive invention (either consciously or unconsciously) and afterward during evaluation.

H. Three Alleles Per Individual?

In Round 1 the most obvious anomaly is the presence of 3 variations. The simplest way to get 3 variations is to assign a different variation to each of the 3 genotypes. But this requires revising the 'dominance' concept of Mendel, so students tend to avoid this simple solution. Instead, they often pursue a different a different revision: "As with the other groups, they seemed more willing to modify the number of alleles in an individual, which is in conflict with both the Mendel and meiotic models, than the concept of dominance." If dominance is retained but the number of alleles is changed from 2 to 3, then 3 phenotypes (A, B, C) can be produced by the 3 homozygous genotypes (11, 22, 33), with the heterozygous genotypes (12, 13, 23) also mapping to these same 3 phenotypes. But a model in which an individual has 3 alleles often meets with conceptual resistance due to its conflict with meiosis. { **consistency of components**, in either 2.21_{D} or 2.28_{C}; relationships can be defined as either internal or external, because meiosis can be considered a separate model or a sub-model within a larger model } This conflict can be recognized at the stage of theory generation, as described above for dominance, or later during theory evaluation. For example, Finkel (1993, pp. 303-304) reports that in Round 1 the Morgan group develops a 3-allele model for awhile, then thinks about the difficulties involved in making predictions using a Punnett Square — "'This thing is gonna be huge, ... It's gonna have nine squares, ... It's gotta be three dimensional'" — and recognizes the conceptual difficulty:

M2: Wait, there's only two parents.
F: You know, you know the problem? You know what it is? You can't have it — the three allele theory is shot because
M2: you have to take two from one.'"

In another group (described by Wynne, 1995, p. 71), this time working in Round 2, when asked by a fellow group member whether it is possible to revise the part of Mendel's Bible which states that "the number of alleles/gene in an individual is 2," one student says "I'm not sure" but another objects to this change because "'each individual (parent) passes one of the two alleles to each offspring.'" But soon after this the student who objected was willing to consider a 4-allele model in which each parent passes two alleles to each offspring. And in Round 3 many groups avoid the postulate that males can receive one allele from the mother but none from the father, which does occur in the scientifically accepted model of X-linkage. Based on these observations of students who are deciding whether or not to challenge a model based on its consistency with principles of meiosis, Wynne (1995, p. 117) concludes, "This premise — that each parent contributes an equal number of chromosomes, and thus an equal number of alleles, to their offspring — was one of the most basic to the groups' understanding of meiosis." This principle is a useful generalization, but there are exceptions; students can eventually accept the "1 allele per individual" concept of X-linkage, and sometimes they challenge the concept of "4 alleles per individual" by protesting that "'it kind of violates the things doesn't it? the rules?' (Finkel, 1993, p. 187)" or that according to meiosis "'you don't get two alleles from each parent.' (Lemberger, 1995, p. 115)"

I. Protected Components

When students are unwilling to revise a particular conceptual premise, whether this is dominance or 'an equal number of alleles from each parent', this premise attains the status of a protected 'hard core' theory component (Lakatos, 1970). { **protected component**, 2.53_A } A component can be protected for three reasons: because students think it is true, because they think that changing it would make a model more difficult to use, and/or because changing it violates their conception of what 'model revising' means. { **plausibility** and **utility**, 2.43; **thought styles and methods of theorizing**, 2.72_B } Sometimes it is difficult to know which reasons occur in a student's unobservable thinking, based on observable dialogue. For example, when a student

challenges a proposed model with 3 alleles per individual — "'No, 'cuz we're using the Mendel bible, that's too far.' (Finkel, 1993, p. 274)" — is the protest due to thinking that the original component is still true, or that changing it is "too far" to still be scientific model-revising or to fit within the rules of the course? Whatever the reason(s) for protection, an important aspect of problem solving is when students "discuss their understanding of what is meant by the term model revising and make decisions about what changes they can make in the model of simple dominance based on ideas about what parts of the model can be changed and what parts cannot be changed. (Finkel, 1993, p. 270)"

Sometimes a protected component is beneficial for problem solving, and sometimes it is detrimental. { **tradition and innovation**, 2.72_H } Viewed from a realist perspective, when constructing a theory that can make accurate predictions the difference between benefit and detriment usually depends on the truth of a component; components that are accurate (i.e., those that seem to correspond to what actually occurs in nature) will usually help, and components that are inaccurate (i.e., false) will hinder. This correlation between predictive utility and truth is not 100% reliable, but I think it is a good way to bet.

For example, in Round 1 an 'equal numbers' postulate can help eliminate a '3 alleles per individual' model that doesn't work, but in Round 3 it may hinder the development of a theory (X-linkage) that works well. And if students don't realize the distinction between '3 alleles per individual' (not consistent with meiosis) and '3 alleles in a population' (consistent with meiosis), during Round 2 a valid model could be eliminated from consideration, as N (of the KLN group) attempts to do in response to the proposal that a third allele may exist in the population (Lemberger, 1995, p. 124):

> R [Lemberger]: Why don't you think you should be adding a 3?
> N: Because you just get one from each parent.'"

Eventually, however, a model involving 3 alleles in the population is accepted by KLN, in part because it is supported by consistency with a known biological system, "'K: Right. There could be, like for eye color, you mean? There's a whole bunch. There's tons of different alleles of the gene for the population, but still each persons got two.' (Lemberger, 1995, p. 125)" In this case the students find a naturally occurring precedent for their own model, and evaluation based on

conceptual criteria increases the status of a theory. { **consistency of components** } This use of conceptual evaluation to increase status contrasts with the idea-blocking, status-reducing conceptual evaluations discussed in Subsections G and H.

J. Conceptual Information from the Teacher

During instruction related to conceptual evaluation, the teacher's objective can be to increase the beneficial effects of students' concepts, to decrease the detrimental effects of their concepts, or to do both. Because students' knowledge of biological phenomena and theories is often limited, the teacher can play a valuable role by serving as a source of reliable information that can be used as a basis for evaluating the 'external conceptual consistency' of a newly proposed model. { **external memory storage** for content knowledge, 2.73_B } This instructional function is described by Lemberger (1995, p. 304):

> Often times the teacher would help students judge the plausibility of a conception by showing the similarity of the conception to another, scientifically accepted conception, or by providing an explanation of the conception at another level of biological organization, i.e., by discussing the consistency of observations at the phenomenological level of the phenotypic variations with the students' thoughts about the conceptual level of the genes.

The teacher also provides hints to help students break through a conceptual block. Sometimes all that is needed is a nudge in the right direction, just enough to get students to examine their own concepts. For example, if students are not making a distinction between the number of alleles in a population and in an individual, Sue can simply ask students to look carefully at the top section of the Mendel Bible. If this is not enough, she can point to the relevant lines and ask "Why are there two lines here? Do both lines describe the same concept? If not, can you explain the difference?"

When the teacher provides information or hints, students usually listen carefully, because they know that the teacher knows a lot about genetics and problem solving, she knows the answer to the GCK problem, she assigns the course grades, and because they like her and they know that she really does care about her students. { **authority**, 2.31_C, of the teacher }

K. An Example of Conceptual Assistance

Or the teacher's advice can be more specific and detailed. For example, although Sue often lets students "do what they want" in Round 3, occasionally during students' GCK work (and always in the conference that follows) she will discuss the special characteristics of X and Y chromosomes. { **external memory storage** for genetics knowledge } The following brief summary is expressed in direct statements, but during instruction much of the content, along with the underlying logic, is conveyed by the skillful use of questions interspersed with discussion, supplemented by looking at the relative sizes of chromosomes in a Human Genome poster on the classroom wall.

The argument: In humans there are more than 100,000 genes but only 23 pairs of chromosomes, so each chromosome pair carries approximately 5000 genes; an X chromosome is normal size but a Y chromosome is much smaller, so it seems logical that an X will carry more alleles than a Y; because of this difference, for some X-linked traits a male will have an allele on the X chromosome (received from his mother) but will not have a corresponding allele on the Y chromosome contributed by his father. The key conceptual distinction here is that even though during meiosis the gamete receives one homologous chromosome from each parent, this does not necessarily mean there will be one allele from each parent.

For many students, this distinction was not easy to accept:

> Most groups were able to construct the idea that the genes were riding on the XY chromosome pair. What was hard for all groups was making the conceptual change that the Y chromosome did not carry a gene for that trait. This meant changing the conception that there were two genes per individual for every trait. (Lemberger, 1995, p. 183) { **consistency of components** to avoid a conflict with the currently accepted component of "one allele (or equal numbers of alleles) from each parent," discussed in Subsection H }

Based on observations of their problem-solving behavior, most students seem to begin Round 3 with an implicit assumption that each allele is carried on a separate chromosome. This appears to be an unexamined assumption, rather than a conscious refusal to believe otherwise. As discussed in Section $3.43_{C,}$ this assumption is allowed by the limited meiotic model that students learn early in the course, and is consistent with the "independent assortment" component in Mendel's Bible. And during the model revising that follows these initial models, before Round 3 there is no reason

to think about the relationship between alleles and chromosomes, so most students do not think about it. During Round 3 some students overcome this unwarranted assumption by themselves; but for other students it is helpful, in order to reduce the status of this misconception, to compare the number of genes with the number of chromosomes, as described above. For X-linkage there are two essential concepts — 5000 alleles per human chromosome, and fewer alleles on Y than on X — but for autosomal linkage only the first of these concepts is needed.

L. Combining Ideas in New Combinations

In Round 2 the most obvious anomaly is the presence of 4 phenotypic variations. Students commonly propose three models that can produce 4 variations. First, if two alleles are in the population, and if the phenotype for a '12' offspring (with a 1-allele from mother, and a 2-allele from father) differs from the phenotype for '21' (2 from mother, 1 from father), there will be 4 genotypes (11, 22, 12, 21) and thus 4 phenotypes. Second, analogous to a common strategy in Round 1, if there are 4 alleles in the population, 4 variations are produced by the 4 homozygous genotypes (11, 22, 33, 44), and with dominance no new phenotypes are produced by the heterozygous genotypes. A third strategy involves 3 alleles, which produce 6 genotypes; with dominance there are 3 variations, with codominance there are 6 variations; to get 4 variations, 1 of the 3 heterozygous genotypes must involve codominance. For this third model, after Round 1 the retroductive logic should be easy: every heterozygous genotype that is codominant increases the number of variations by one.

But often it is difficult for students to propose this idea. This is partly explained by the principle that combining ideas usually requires more creativity; it is more difficult to do, and is less likely to occur. { **invention by synthesis**, $2.53_{A,C}$ } For example, the Crick group

> detected both simple dominance and codominance in round two, plus proposed a third allele, but did not easily fuse the three ideas...[into] the model currently held by geneticists. (Johnson, 1996, p. 99)

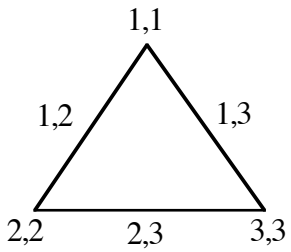
Sometimes, however, a student also has conceptual resistance:

> 'You cannot combine the Mendel model rules with the [codominance] model rules. ... You can't do that though for a system. ... You can't just say, OK, today this is dominant and tomorrow it's not. Things don't lose dominance and then regain it. ... If you take a 1 from like one parent, it's either going to be dominant or it isn't. You can't, I mean, looking at it all by

> itself and not with its partners you have to see how it relates. You can't just say, either here it's dominant and here it's not. You can't say it might be if this is something, or it might not be, it either is or it isn't on its own.' (Lemberger, 1995, pp. 246-247; discussion between the student's comments has been omitted)

Lemberger (1995, p. 240) concludes that "the greatest constraint to their problem-solving ...was the conception that all the alleles of one gene had to interact in either dominant/recessive or codominant fashion." On the other hand, members of the Morgan group (Johnson, 1996, p. 131) "were not uncomfortable" with mixing dominant and codominant relationships.

To help students understand why this mixing is logically acceptable, the teacher uses a triangle (shown below) as a graphic representation that makes it easier to perceive and believe the possibility that the three heterozygous genotypes (12, 13, 23) each have a relationship (of dominance or codominance) that is independent of the other two relationships.



Because the teacher's objective is for students to struggle with a problem for awhile and then achieve success, her guidance must be carefully controlled or the problems become too easy. For example, for each research group Sue must decide whether (and when) to show this 'triangle' graphic. The pedagogical challenge of maintaining an appropriate level of difficulty by sharing some hints and information — but not too much — is discussed in Section 3.45.

M. Key Factors in Successful Model Revising

Subsection E contains a brief summary of problem solving: "In real life, anomaly resolution is usually a messy, meandering process." The complexities of the problem-solving process, and the decisions that guide its path, are the main topics of Subsections N-Z. For decisions about how to use time during model-revising problem solving — such as whether to gather more data or shift to inventing models, or whether to abandon a model or continue developing it — there are no algorithms for success. But there are reliable heuristics, and Finkel (1993, p. 264), after comparing

the actions of relatively successful and unsuccessful problem solvers, offers a summary of the major differences: "Groups that construct final models are able to develop and use successful strategies for testing their models, and frequently work in such a way that they do not discard promising models but alter them slightly."

The following subsections explore “strategies for using time” (**pursuit invention and evaluation**) and the crucial actions highlighted by Finkel — **testing** models (by **designing experiments**, **doing experiments**, **making predictions**, and **evaluating theories**) and revising theories (**inventing theories**) — but do not attempt to draw any definitive conclusions regarding these strategies or actions.

In the following discussion there often will be excerpts from four groups studied by Finkel (1993) and Johnson (1996); Johnson characterized one pair of groups (Morgan and Pauling) as more successful than the other pair (Crick and McClintock).

N. Using Time: Observation and Interpretation

One of the most important decisions during research (real or simulated) is how to invest time in observation and interpretation. For example, after students see the field population for a problem, they must decide how to split their time between gathering data and inventing theories. Finkel (1996, p. 358) describes the “data first” strategy adopted by all students:

> Although some students began their work on the first model-revising problem by making hypotheses and testing them, this strategy was quickly replaced by the data collection strategy, and on subsequent problems, all students in all research groups began their work by using the computer to make crosses and collect additional data about the new population.

In this area, the work of students contrasts with actual research science. For scientists, an important goal of experimental design is efficiency. As described in Section 2.63_A,

> As an essential part of a cost-effective strategy, mental experiments serve as a quick-and-cheap preliminary screening process that can facilitate the effective designing and selection of physical experiments that typically require much larger investments of time and money.

But with GCK there is little incentive to be efficient. The quick-and-easy GCK experiments do not require "large investments of time and money," so an effective strategy is to do lots of experiments, with little concern about whether they are necessary or helpful. { **thought-experiments** to guide

experimental design, $2.63_{A,}$ is usually not done by students }

But if excessive observation interferes with interpretation, there can be a cost. Even after an anomaly has been recognized (as in Rounds 1 and 2 when a 'variation number' anomaly is observed in the field collection), deciding to gather additional data first, before trying to invent a new theory, is usually an effective strategy. But data collection is just a preliminary action, done in order to eventually allow the main action of model construction. If too much time is invested in preliminary data collection, this can interfere with the main action. For example,

> Their production of new ideas [was hindered by] their waiting to propose tentative models until they [Crick] felt that they had exhausted all the cross possibilities. ... They performed many crosses before proposing a model and seemed to be waiting for a pattern to 'appear' for which they could quickly build a model. ... This is in contrast with the Morgan group who proposed models early on in rounds one and three and therefore had more time for model assessment and anomaly resolution. (Johnson, 1996, p. 190, 132)

In some situations, such as this one, it seems to be a waste of time to continue gathering more data, thus avoiding the cognitive challenge of model building. But sometimes additional data does lead to recognizing a pattern that inspires an explanatory model. There is a delicate balance between investing too much time and not enough time. The characteristics of an effective balance vary from one situation to another, and there is no algorithm for deciding when to shift from data gathering to model building, or vice versa. { **action evaluation for probing**, 2.71_{D} — decisions about balancing observation and interpretation }

However, at some time (whether this occurs early or late) most groups reach a point where several new models have been tentatively proposed but have not been fully evaluated. At this point there are multiple models, which multiplies the available options, so students must decide which model(s) to pursue for testing and possible development. Before discussing these decisions about **theory evaluation for pursuit** { $2.26_{A\text{-}B}$ } in Subsection R, it will be useful to look at some evaluation criteria for this type of decision, in Subsections O-Q.

O. Theory Evaluation: Balancing Empirical and Conceptual Factors

In scientific research, and in the MG classroom, theory evaluation involves both empirical and conceptual factors. { **multiple criteria for theory evaluation**, 2.22, 2.45_{A} } According to Lemberger (1995, pp. 120, 269) during model revising the relationship between these factors is that

empirical evaluation (especially degree of agreement) is the driving force that produces conceptual change, with conceptual evaluation playing a moderating role:

> As with other groups, group 4 used fruitfulness [empirical adequacy] as the greatest indicator of status during the three rounds of the problem-solving sequence. ... In this classroom, problem probing put students in a position where they were required to make their conceptions of inheritance consistent with their observations of the data. In order to make their conceptions consistent with the empirical data they were forced to change their conceptions. These changed conceptions, however, still had to be consistent with other, related conceptions in the problem solver's conceptual ecology. These other conceptions acted as constraints on any possible changes constructed by the problem solver.

Based on my own study of student problem solving, this conclusion seems valid, and it is not contradicted by the analyses of others who have studied the classroom.

One reason to prefer empirical evaluation is that students have a limited base of knowledge about biological phenomena and theories. This limited knowledge makes it difficult to state confidently that a particular model should be rejected because it is inconsistent with known phenomena or with currently accepted scientific theories. But students can be confident about whether a model's predictions agree with the data generated by GCK.

Empirical adequacy is the 'bottom line' in determining the *result* of evaluation, but conceptual constraints can play an important role in the *process* of developing models, as discussed in Subsections G-L. And conceptual criteria can influence the ease of making a final decision based on empirical criteria. If students think a model makes sense conceptually, they may respond to empirical anomalies by revising this model instead of rejecting it. But if a model already has low status due to conceptual evaluation, at the first sight of anomaly it will be easier to say "well, we didn't like that model, anyway" and to quickly reject it. This attitude of "rejection without tears," due to the combined influence of conceptual and empirical factors, is reflected in the following excerpt (from Lemberger, 1995, p. 115):

> Researcher: Four alleles (inaudible) so if you get two alleles from each parent, and that would give you four. I mean that would be a way to figure out four different combinations.
> A: Yeah. But you don't get two alleles from each parent. (inconsistent with Womendel's conception of meiosis)
> R: But we think you might have to.
> A: But it doesn't work, anyway. (lack of fruitfulness [empirical adequacy])'

Wynne (1995) describes two reasons for students to "mentally run" a meiotic model: to generate gamete-genotypes for making predictions about expected phenotypes, and to check

whether a proposed model is consistent with the principles and process of meiosis. Thus, the model of meiosis is involved with both empirical and conceptual evaluation. It can also be involved with both plausibility and, as described in the following quotations, utility:

> In order for it to work they will have to figure out how to make predictions using Punnett squares. ... They struggle with how to make a model with three alleles per individual work, and run into some problems because they cannot decide how the parents give alleles to their offspring. ... They formulate a problem that they can't resolve: how to make a Punnett square to test a three allele model. (Finkel, 1993, pp. 76-78)

> One reason that groups protected this assumption [that both parents contribute the same number of alleles to offspring] may be because, had they not, their meiotic models may have, in their minds, "fallen apart." ... Perhaps they could not envision revising this protected assumption, without losing the ability to use their meiotic models to describe, predict, and explain. (Wynne, 1995, p. 149)

If students cannot combine their existing knowledge of meiosis with a new '3 alleles per individual' model to make predictions, there is an obvious decrease in utility. This inability to combine can also be interpreted as incompatibility between the new 3-allele model and their accepted principles of meiosis, leading to a decrease in plausibility of the new model due to conceptual evaluation. Students can thus reject a 3-allele model for instrumentalist or realist reasons — because it is difficult to work with the new model, or because they think "this is not the way it is." { **utility and plausibility**, 2.43 }

Perceived utility is important in evaluation. Even though it is possible, using Prediction Overviews, to systematically predict all possible crosses for a model with 3 alleles per individual, if students do not perceive this as a possibility then for them the practical utility of a 3-allele model is greatly reduced. This lack of perceived utility will produce a reluctance to begin applying and testing the model, with a lowering of the model's **utility status** { 2.43 }, and probably (but not necessarily) a lowering of its **plausibility status**.

> Concern for conceptual consistency was often important in evaluation, but not always:
>
> This lack of attention to consistency was evident often in rounds one and two. They [the McClintock group] worked with tentative models that were truly ad hoc, such as their 'corecessive' model, but that didn't occur to them and therefore ad hocness was not used as an assessment strategy. (Johnson, 1996, p. 160)

And another group, in Round 3, "was able to get around some prediction problems...by making an ad hoc assumption that heterozygous individuals never mated with each other. (Lemberger, 1995,

p. 266)" { **systematicity (lack of ad hocness)**, $2.21_{A\text{-}C}$ — this criterion was sometimes used by students, but not always }

Similarly, there are differences in thoroughness of testing for empirical adequacy. For example, the Pauling group, despite its success in producing satisfactory models, was a bit lax in testing them: "As with the model produced in round one, this one didn't get tested on all crosses, but that didn't seem to bother them. ... I feel uncomfortable about their lack of completeness in the assessing process. (Johnson, 1996, p. 82)"

P. Denial of Anomaly

A rare response to empirical anomaly is to challenge its validity. This might be done because there is a reason to suspect that — due to an error in making predictions, doing an experiment, or interpreting data — a result is not really anomalous. From my own experience, both personal (in solving GCK problems) and vicarious (in reading about students' actions), it is obvious that errors occur, so this cause of apparent anomaly should always be considered. This is one of the reasons that sometimes (but not always) it is productive to persevere and to retain a model despite the recognition of anomaly. { **logical underdetermination** by data, 2.44 } If a model has made accurate predictions for previous experiments, especially if these involve a wide variety of systems, students have more reason to suspect that apparent anomaly may not be actual anomaly.
{ **previous experimental hypotheses**, 2.12_C }

Another option is to deny that an anomaly is a serious difficulty for the model being considered. This can occur by simply ignoring the anomaly, as in the McClintock group who "in general were not troubled by a natural world that was not consistent, and explained crosses that were anomalous with the fact that is must be ‘coincidence’ or ‘chance’. (Johnson, 1996, p. 160)" In this case Johnson thinks the denial is an escape, to avoid a need to explain the anomaly. But denial can also be the result of careful analysis that leads to a decision about how to define the domain of phenomena for which a theory is claimed to be valid: "When this strategy was used by early geneticists it was generally done if the anomaly was assumed to be an unusual case or outside the scope of the theory. (Johnson, 1996, p. 176)"

In the classroom a productive denial strategy occurs when the Morgan group, in the final round of model revising, is given a very difficult problem involving autosomal linkage with crossover. When there is ‘crossover’ the GCK fruitflies sometimes behave as if there is autosomal linkage, and sometimes they don't. This seemingly random behavior necessitates a sophisticated modification of the basic linkage model — a modification that was probably beyond the capabilities of students in the course, and that would require experiments beyond those possible using GCK. So what happened? The Morgan group "worked with a linkage model from the first day, but were especially frustrated by the anomalous cross results attributable to cross-over events. ... They discussed localizing one part of the anomaly (crossing over) outside the domain of their model and the other part (linkage) within the model. (Johnson, 1996, p. 207, 205-206)" The productive and unproductive uses of ‘denial’ as a response to anomaly are summarized by Johnson (1996, p. 206):

> There are examples from the Darden account in which localizing an anomaly outside the domain of the theory was crucial to its resolution and the acceptance of that resolution. An example of that strategy being used by the students can be seen when the Morgan group suggested that cross-over events should not be interpreted as necessitating a negation of the linkage model. That is an example of a productive use of the strategy. In the remaining examples from the student research in which an anomaly was localized outside the domain of the model, the choice of strategy seem to be more to ignore its existence and therefore was unproductive.

But denial responses of any type are rare. Almost always, an empirical anomaly is viewed as a serious threat to a theory's validity. The usual response to anomalous data is a rejection that is either permanent or temporary: "Most of the time a tentative model was dropped at the first sign of an anomaly. In many of those cases the model was only temporarily dropped and reemerged several times during the model revision process. (Johnson, 1996, p. 199)"

Q. Evaluation based on Thought Styles and Complexity

This subsection is an extension of the more comprehensive discussion of conceptual evaluation in Subsections G-L. In the context of a high school science course, students expect their newly invented models to be consistent with other scientific theories and phenomena (as discussed earlier) and also with their own beliefs about high school education. For example, one student explains

6

why a certain model is unlikely to be a solution for the type of problem the teacher is expected to assign:

> 'It would be too hard for problem two for it to matter which allele you got from which parent. You know what I'm saying?' [This criticism of the model] has nothing to do with biology or genetics. Plausibility [conceptual evaluation] was not always based on a conception's consistency with other high status conceptions, but also on its consistency with what the students thought was an appropriate problem for this class! (Lemberger, 1995, pp. 115-116)

With a liberal interpretation, this is roughly analogous to the restrictions that a scientific **thought style** { $2.73_{B,H}$ } might impose on theorizing. Or, again with a liberal interpretation, this might be considered analogous to a metaphysics-based restriction on theory components, except that now the restriction is on the types of systems and models that are assumed to occur within the 'classroom universe'.

Another reason to reject a potential solution — either because students think it is too tough for the class (as described above) or is too difficult to use, or is unlikely to be true in nature — occurs when students "frequently reject models as too complex if they offer too many possible genotypes given the number of phenotypes observed in a population. (Finkel, 1993, p. 270)" { **simplicity** by having a reasonable number of causal factors, 2.21 }

A more effective conceptual evaluation, especially involving criteria such as "an appropriate level of difficulty for the classroom" or complexity, would have made the process of problem solving much easier for a group that worked hard but got nowhere, as described in the following paragraph.

In Round 2 the 'standard solution' is conceptually simple — instead of 2 alleles in the population, there are 3 — but the application of this solution (by figuring out genotypes, cross results,...) is difficult, and "students were often unable to organize their data and even their thoughts, because of the complexity of what they were seeing. (Lemberger, 1995, p. 110)" With the standard solution this problem is moderately complex, and there are 21 possible crosses involving different genotypes, but with a model proposed by the Crick group (4 alleles in the population and 3 alleles per individual) there are 1600 different crosses! One member of the group tells the teacher that with their model,

'You can't do a Punnett square but you can...take all the possibilities. ... It's super hard to

> figure out the genotypes...'cuz if they're heterozygous they can be a number of things.' (Finkel, 1993, p. 89)"

Describing the group's struggles in trying make this model work, Johnson (1996) says,

> They didn't state that this would be too complicated (an assessment strategy often used by the Morgan group). ... They were incredibly persistent, running through cross results and assigning genotypes. (p. 176)

The students claim to be able to predict the results of crosses, but even if this was true it would be viewed with suspicion, because with 20 genotypes and 1600 crosses there is plenty of "wiggle room" for ad hoc data-fitting maneuvers. In addition, there is a clash with meiosis due to postulating 3 alleles in each individual.

This group could have saved themselves much time and frustration if they had asked themselves, as described above, whether the teacher would assign a problem that required so many genotypes and crosses, or whether it was a good use of their time to continue developing this extraordinarily difficult-to-apply model. Or, if they had followed the advice of Occam's Razor to "seek simplicity" by beginning with the simplest models that can produce 4 variations, and if they had been able to calculate or estimate the number of different geno-types and crosses for each model (as shown in Figure 22), they would have focused their attention on the "3-and-2" model (the standard solution) and the "2-and-3" model.

R. Combining Perseverance and Flexibility

Although in the situation described above they were too persistent,

> The Crick group never fully tested a model in any of the three rounds. ... They jumped from one model to another. They also moved quickly from one population to the next if they weren't successful in the construction of a model saying 'the old one [population] doesn't work.' (Johnson, 1996, p. 191; brackets in original)

In her analysis of the factors that contribute to successful problem solving, Finkel (1993) emphasizes the importance of perseverance in staying with a promising new model and revising it in response to anomalies. Johnson (1996, p. 145) agrees, "It is frustrating to watch the groups 'flirt' with the currently accepted model and give it up too quickly." But in the example above the Crick group is tenacious in testing and revising their "4-and-3" model when they should have decided to abandon it (and its 1600 possible crosses) in order to focus their efforts on developing a simpler model.

Compared with making real-time decisions, it is much easier to judge the quality of decisions about time use in retrospect. If a group continues to develop a model that eventually succeeds, their perseverance is productive. But if they continue to develop a model that fails, they may be wasting time. I say "may be wasting time" instead of “are wasting time” because — unless the process of invention is so efficient that an inadequate model is never proposed, and this is highly improbable — testing models that are eventually rejected is a necessary part of science.

Clearly there is a tension between the virtues of tenacious hard work and the wisdom to stop wasting time on an approach that isn't working and probably never will. The contrasting yet complementary merits of perseverance and flexibility are eloquently portrayed by Robert McKim (1972) in his book, *Experiences in Visual Thinking:*

> When should you abandon one strategy to try another? When is perseverance a virtue, and when is flexibility? Sometimes dogged persistence in the use of a single strategy yields a solution. ... On the other hand, it may simply be the wrong key. ... Genius is often associated with the ability to persevere;...[but it] is also linked to the ability to be flexible. Clearly, we are facing a paradox. Perseverance and flexibility are opposites that together form an important unity. (p. 165)

When deciding how to use time during problem solving, one of the most important decisions is whether a particular model should be pursued or abandoned. { **evaluation for theory pursuit**, $2.26_{A\text{-}B}$, $2.71_{A\text{-}D}$ } According to Finkel (1993, p. 264), one secret of success is to "not discard promising models." It is difficult to dispute this good advice, but the obvious question arises, a question that is the key to coping with the paradox of perseverance and flexibility: How does one know when a model is "promising" and when it is not?

S. Observables and Unobservables, Logic and Patience

A good *working knowledge* of genetics is helpful in making wise decisions about which model to pursue for testing and development, and which models to temporarily abandon. { **preparation** by learning content-knowledge, 2.71_{F}, 2.73_{B}; **evaluation for theory pursuit** } Three researchers discuss some reasons for students abandoning models too soon:

> It is frustrating to watch the groups ‘flirt’ with the currently accepted model and give it up too quickly, in this case because they didn't assign genotypes accurately. Most groups bounce from one model to another and between models without spending much time applying genotypes to the cross results. (Johnson, 1996, p. 145)

> The greatest difficulty at this point in developing a conception was...organizing the cross data and assigning genotypes to phenotypes. This was an extremely frustrating process and...they almost abandoned the new conception because they couldn't get it to explain the crosses they were seeing. (Lemberger, 1995, pp. 251, 249)

> [In contrast to unsuccessful groups who] tend to discard models completely in favor of new, unrelated models,... groups that construct final models...frequently work in such a way that they do not discard promising models but alter them slightly (for example by re-arranging the ways in which genotypes are assigned to phenotypes) until they fit the population being studied. (Finkel, 1993, p. 264)

There is widespread agreement that in making decisions about which models to pursue, and how to revise these models, one valuable asset is fluent, accurate thinking about the connections between phenotypes and genotypes. It is especially important to know what can and cannot be justifiably inferred from data. { 2.41-2.45 } A **metaphysical assumption of consistency** { 2.32_B } produces an expectation that identical experimental systems should always produce the same results. But in GCK problems two experimental systems — i.e., two sets of parents used for crosses — may look alike (at the level of observable phenotypes) even though they actually differ (at the level of unobservable genotypes) in ways that lead to the causal production of different results. But these results would be predictable if the unobservable genotypes were known. Thus, apparently identical systems may produce results that differ in predictable ways. For example, AxB (using the 'simple dominance' symbolism of Section B10) will produce 100% A if the A-parent is homozygous, but only 50% A with a heterozygous A-parent. Conversely, identical phenotypic results do not imply that phenotypically identical experimental systems are identical at the level of unobservable genotypes. For example, the same result of 100% A will be produced by either "aa x aa" or "aa x ab" systems, both of which appear to be the same AxA experiment.

Obviously, in effect-to-cause problems a correct interpretation of experimental data requires an understanding of genetics concepts and the careful use of sophisticated logic. But it also requires patience. In GCK problems the nature of an experimental system is not immediately apparent from observable data; instead, it must be inferred by a process of retroductive logic, and sometimes the inference must be delayed until more data is collected. Hence the need for patience; ambiguity must be tolerated for awhile, to avoid the error of jumping to conclusions and landing on the wrong ones.

The following excerpt illustrates the understanding that only a limited conclusion should be

drawn from a certain cross (stubble x stubble) that produces 100% of the stubble variation:

> 'Arright, all you get is stubble, so we could say stubble is homozygous...but we don't know if it's dominant or recessive.' F asks M1 how they 'will know if it becomes dominant or recessive' and M1 responds that they will 'just have to do a lot of crosses and figure it out.' (Finkel, 1993, p. 141)

Even though M1's explanation indicates a lack of conceptual understanding,[6] he realizes that a useful type of conclusion is that "we can infer this (homozygosity) but not that (dominance)."

T. Retroductive Inference of Models and System-Theories

In retroductive logic, either of two variables (domain-theory or system-theory) can be adjusted for the purpose of obtaining a prediction that agrees with the third variable (observations): domain-theory + system-theory → observations. This 3-variable logic, using the terminology of Chapter 3, is "model + system-theory → observations." { **retroduction of model and/or system**, 2.53_D } Because predictions depend on a model and a system-theory, even if a model is correct it will produce incorrect predictions if the experimental system is not properly characterized. Therefore, as described in the preceding subsection, students can become dissatisfied with a valid model because they could not "organize the cross data and assign genotypes to phenotypes (Lemberger, 1995)" or because they tried to do this but "they didn't assign genotypes accurately. (Johnson, 1996)" If students understand the retroductive relationship of "model + system-theory (for parental genotypes) → prediction (of progeny phenotypes)," they will understand another option: "Do not discard promising models but alter them slightly (for example by re-arranging the ways in which genotypes are assigned [in system-theories] to phenotypes) until they fit the population being studied. (Finkel, 1993)"

As explained in Section 2.52_D, a scientist either can assume a model is known (in order to retroductively infer a system-theory), or can assume a system-theory (to infer a model). Whatever strategy is used, there is usually a close correlation between "knowing the genotypes" and "having

---

[6]. For a dominant/recessive pair of alleles, a same-variation cross will produce 100% of that variation for any of three genotype crosses — aa x aa, aa x ab, bb x bb — so it is impossible to draw a logically justified conclusion about either homozygosity or dominance, except that at least one parent (not "stubble" as stated by M1) is homozygous, or that if the trait is recessive then both parents have to be homozygous.

an explanatory theory" with the former leading to the latter, and vice versa. When a model is first proposed it is often only a vague outline that needs to be developed by specifying the details for a fruitfly population — such as system-theories, genotype-phenotype mappings, and which traits are dominant, recessive, or codominant — before it can be adequately tested and perhaps revised. A 'large model' can be developed incrementally by combining 'small models' about specific parents (are they homozygous or heterozygous, and with which alleles?) or about predictions that involve system-theories (for parental genotypes) and the resultant observations (of progeny phenotypes).

Effective retroduction is stimulated by a solid foundation of genetics knowledge. Conversely, a weak foundation will hinder the invention of ideas for models and/or systems. For example, Subsection F describes a situation where poor memory (and/or misconceptions) prevented the recognition of options for model revising; because students did not remember the distinctive DxD cross, they did not draw the warranted conclusion, "aha! this indicates codominance." A lack of genetics knowledge — not knowing (or not remembering) that "this type of cross result indicates this inheritance pattern" — caused students to miss a golden opportunity to recognize a key component of a successful model.

A lack of knowledge can also lead to a wrong retroductive conclusion about an experimental system, as in the Crick group (Finkel, 1993, p. 266) during Round 2:

M3: See these two cuts, see we got, look at that though. We took a cut and a cut and we got cut and fat. Then we took the cut and the cut again and we got cut and cut.
M2: I know that.
M3: See we get all cuts again, so something in cut has to be like recessive.'

Instead of correctly concluding, based on the evidence — "'we took a cut and a cut and we got cut and fat'" — that "something in cut has to be like dominant," their faulty logic, based on an inadequate working knowledge of genetics, leads them down a wrong path, at least for awhile.

U. Descriptive Theories and Explanatory Theories

A consistent characteristic of the Crick group was a failure to make the phenotype/genotype connections that, in the GCK problems being solved, are the key to converting descriptive theories into explanatory theories:

As in the first round of model revising, they recognize that they are able to predict cross results

> but not explain them, a feature of their work that they find continually frustrating. ... M2: 'We know what we get every time [we make a cross].' M3 agrees, adding: 'I know, but we have to find out a way to explain it.' (Finkel, 1993, p.81; brackets in original)

To clarify the meanings of 'predict' and 'explain' as they are used here, consider a GCK problem in which all possible crosses are done on a population with phenotypes F, G and H, and a group recognizes a pattern (at the level of observable phenotypes) but cannot invent a mechanism (at the level of unobservable genotypes and alleles) to explain the cause of the patterns they see in the data. If this group changes to a new population with phenotypes J, K and L, after a few preliminary crosses — to figure out how J, K and L are playing the same roles that F, G and H did in the original population — the group can use the empirical pattern they have recognized to predict (at the level of phenotypes) the results of the remaining crosses, even though they still cannot explain (at the level of genotypes) the mechanism of what is happening. { **descriptive theories** and **explanatory theories**, $2.11_{D,}$ 2.24, A22 }

Finkel describes a two-stage process of developing a model:

> As the class comes to an end, M2 tells the researcher that they have figured out the results of all the crosses but that they 'don't know why' they work that way. The researcher tells them that figuring out what is happening is 'the first step,' and that now they can go on to 'figure out why, and how that is possible.' (Finkel, 1993, p. 155)

> Although the Crick group often made successful predictions they were doing so only at a phenotypic level. ... [By contrast] the Morgan group consistently applied genotypes to the parents and then predicted the cross results. (Johnson, 1996, pp. 132-133)

Finkel (1993) describes the students' own standards for what constitutes a satisfactory model:

> All models produced by these groups can be described as explanatory in the sense that they are attempts to explain the sources of observed patterns. However, in many cases students recognize that they are able to predict cross outcomes without being able to explain them. In the minds of most of the students being able to predict cross results is insufficient because prediction is of secondary importance to explanation. (Note: This idea is prevalent in discussions among all groups, but its source is not clear.) (p. 259)

Some "sources" for this idea are discussed in Sections 3.34_{A} (Category 5a), 2.24, and A22.

{ Students (and scientists) prefer non-restrictive **non-empiricist methods** for doing science. }

V. Testing Models: Experimenting and Evaluating

The testing of models involves **observation** (by designing and doing **experiments**) and **interpretation** (by empirical and conceptual **evaluation**). Previous researchers who have studied

this classroom claim a high correlation between the quality of model testing and the development of successful models:

> In her study Finkel (1993) found that groups that successfully produced working models were able to develop and use strategies to apply and test those models. I have also proposed that success at model production may have been more dependent on model assessment rather than production of tentative models. (Johnson, 1996, p. 229)

Before trying to develop a model, most groups begin an initial "data run" with the goal of doing all possible crosses, to see all possible results. Students can usually do all phenotypic crosses; this is easy. But seeing all results requires doing all genotypic crosses; this is more difficult because genotypes are unobservable; Subsection S explains the need for logic and patience. An additional reason for patience and caution is that in each new round "the number of types of results" differs from that of old models. Because of this, in theory it is impossible to be certain that all possible results have been seen, but reasoning (using logic from genetics and statistics) can lead to a rationally justified confidence about this claim. With a brute force approach — doing the same phenotypic cross over and over, each time using a different set of parents — statistical logic indicates that as the number of crosses increases, so does the probability that all genotypic crosses have been seen. But statistical logic can be supplemented with genetics logic to form an eclectic strategy that is more effective for designing and interpreting experiments, and thus for testing models.

Genetics logic can be especially powerful when the genotype of one or more fruitflies is known. Knowledge about genotypes is useful because, as can be seen by carefully examining the Prediction Overviews in Sections B10-B14, genotype knowledge usually restricts the range of cross-result possibilities. This narrowing of range makes it easier to use hypothetico-deductive logic for making predictions, or to use retroductive logic for speculating about a model or system. In either case it increases the odds that logic will lead to a definite conclusion instead of a deferred judgment that requires patience and further experimentation.

W. Goal-Oriented Experimental Design

There are two basic strategies for the **goal-oriented design of experiments** { 2.61 }. First, there is the "do it all" strategy described above, which has the general goal of gathering a large

amount of data to use for the retroductive invention of models and system-theories.

A second type of goal-oriented strategy is to design experiments for a specific purpose, by first doing mental experiments to predict what might happen during a physical experiment; following this, the physical experiments that in some way look "interesting or important" can be done. But there is not much incentive to do thought-experiments with GCK, so this is rarely done, as discussed earlier. { **thought experiments**, 2.63_A }

In fact, according to researchers it is rare for students to do any type of experimental design that is guided by a specific goal, rather than the general goal of a "do it all" strategy. But experiments with a specific goal are possible, especially when a group's newly proposed model serves as a basis for determining which experiments might provide useful information. As might be expected, this type of design is more common (but is still rare) with the more successful groups. For example, a member of Morgan "does not see crosses as only a way to collect data, but as a task with a specific goal. (Finkel, 1993, p. 168)" And the teacher says,

> They [Morgan] used a variety of strategies to test each model, a process not common in the problem solving of the other groups. ... They looked at the Mendel model to see what portions needed to 'be adjusted,' applied their model to specific crosses with the use of Punnett squares, performed like crosses, repeated crosses and performed generational crosses. These tactics were not commonly used by the other groups. (Johnson, 1996, pp. 107, 113)
>
> In rounds two and especially three they [Pauling] used techniques not common to the other groups. First, they performed two types of crosses that were not typically used by the other groups. They bred pure strains to use in testing for the new pattern and they performed reciprocal crosses. Finally, the Pauling group would consider the cross results from multiple generations of organisms. ... This decision [to create pure strains] was pivotal in helping them to construct a model as it gave them a better phenotype/genotype map. ... They began making predictions early in round three, spent several pages of transcript setting up one particular predictive cross and then continued in successfully predicting the results of a number of subsequent crosses. (pp. 99-100, 71, 100)

Because it is useful to know a fruitfly's genotype — such as whether it is homozygous or heterozygous, and containing which allele(s) — strategies for generating this knowledge are also useful. With a little experience some of these observation-and-interpretation strategies become familiar algorithms that are easy to plan and do. But I don't know any examples of a novel 'crucial experiment' being intentionally designed by students. For example, although it is possible, after a theory of autosomal linkage has been proposed in Round 4, to design an 'abin x bbnn' that is a crucial experiment to distinguish between autosomal linkage and most alternative theories, I doubt

if students have ever done this. But it is common for students to *recognize* a highly discriminating experiment in retrospect, after its results have been observed and interpreted. For example, in Round 1 the results of CxE disagree with the predictions of an initial theory of dominance, but agree with a newly invented codominance theory. There is high predictive contrast (because the two theories make different predictions) and the results agree with one theory and not the other. This is a crucial experiment, but it is recognized as such only after it has been run and interpreted. { **crucial experiment**, $2.12_{B,}$ 2.61_{F} — these can be recognized after they are done, but they are rarely designed in advance by students }

Effective experimental design is based on a solid foundation of genetics knowledge. In all studies of model-revising problem solving in genetics — including those described in Section B20, and also others such as Stewart (1988) and Hafner & Stewart (1995) — adequate genetics knowledge is seen as an essential part of solving effect-to-cause problems. In fact, a good knowledge of genetics will improve all phases of the problem-solving process.

Genetics knowledge will help students invent ideas for new models, and make wise time-use decisions about which model(s) to pursue for testing and development. It will help, if an inadequate model is pursued, to quickly test this model and find the empirical limitations. It is also essential for conceptual evaluation, and for the retroductive revision of a newly proposed model (and its associated experimental systems) to develop an improved model. And it will help students thoroughly test a model so they can be confident of its scientific usefulness.

Of course, by itself genetics knowledge is not sufficient. An effective use of genetics concepts during problem solving requires the skillful blending of content knowledge (about genetics) and procedural knowledge (of problem solving), as described in Subsections S-W. When analyzing problem-solving actions, or when teaching students how to improve these actions, it is useful to think in terms of a *working knowledge* of genetics that includes skill in using genetics-based algorithms and heuristics, and in connecting genetics knowledge with procedural knowledge. { **content knowledge** plus **procedural knowledge**, 2.73 }

X. Trial-and-Error with Fluent Speed

To solve a problem the most efficient strategy is to move on a straight-line path toward a solution. But during model revising the nature of a solution is unknown until it is found, so knowing "which direction to go" is not obvious, and students usually follow a nonlinear route of meandering trial-and-error exploration. An important part of the process is the empirical testing of models, and although a recognition of anomaly usually lowers the status of a model this is not always sufficient reason for permanent rejection:

> Most of the time a tentative model was dropped at the first sign of an anomaly. In many of those cases the model was only temporarily dropped and reemerged several times during the model revision process. This reemergence seemed to be a part of a pattern in which groups would jump from one model to another, either because a group member still had confidence in a model and would resurrect it in an attempt to persuade the other group members or because the model had been previously dropped at the first sight of an anomaly, but later resurrected when more recent models also failed to explain their cross data. (Johnson, 1996, p. 199)

All four evaluation conclusions in ISM — **retain**, **revise**, **reject**, and **delay** { 2.42 } — occur during model revising. A delayed judgment can be viewed as a temporary rejection (that is tentative and reversible) or a temporary retaining (without pursuit, by ignoring a model so other actions can be pursued) until there is a reason to look at the model again. Or an initial rejection can be retrospectively seen as a delay if a rejected model is eventually reconsidered.

In Subsection E, anomaly resolution is described as a "messy, meandering process." This description is compatible with a view of problem solving that interprets it as repeated cycles of trial-and-error inventing and testing. When using this strategy, a useful skill is speed; if a group goes down a wrong path for awhile, this doesn't matter as much if they quickly discover that the path is not likely to be productive. Then they can return to looking for another option to pursue, repeating the cycle of exploring one potential solution after another until a satisfactory solution is found. Of course, skill in quickly completing invention-and-evaluation cycles is improved by a solid foundation that fluently combines content knowledge (of genetics) with procedural knowledge (of problem solving).

Subsection W describes another view, in which model revising is seen as a process that involves careful planning of each step. Both views have validity, and what students actually do

7

seems to be a combination that mixes trial-and-error wandering (sometimes) with careful planning (sometimes). Perhaps 'goal-oriented wandering' is a more accurate description, with the amount of guidance varying from one group to another, and from one situation to the next.

Y. A Story of Goal-Oriented Wandering

The following story, of the Morgan group near the end of Round 2, shows them wandering toward a solution, guided by the logic of genetics knowledge and problem-solving strategies.

First it is told by the teacher (Johnson, 1996, pp. 114-116):

> They did not give up the model [based on a "rock, paper, scissors" analogy] easily and continued testing it on all crosses. ... The group finally became frustrated with the model when it failed to explain the cross results. At this point M1 proposed a three allele model that would ultimately survive their rigorous testing. ... This model seemed better able to explain the crosses...but at this point they seemed tired and applying the model to many of the twenty-one crosses was cumbersome. They began to drift with just a few minutes left in the period. ... They started the third day playing around with the capabilities of the computer simulation. It was unusual for the Morgan group to be off task for very long, but today it took some encouragement to get them back on task again. They seemed to have dropped the three allele model from the previous day even though in their discussion it had been successful in explaining crosses. F even stated, 'We have three models and they all suck.' They then reviewed several of their models...
>
> This [three alleles] was the model that would survive their retesting but again they seemed to side step it and instead went on to describe an imprinting model in which their were two types of heterozygotes, depending on which allele came from which parent. They felt that none of their models explained all their observations and in frustration they went off task again. F still was interested in a three allele model and later, when M2 said they needed to get back to work, she decided they should proceed with a third allele. It was a tentative model that she felt was promising, but was dropped by M1 and M2 as they felt it didn't explain all the crosses. They spent the rest of the period on task, testing that model with cross after cross. It was impressive to see how they stayed with it for a number of pages of transcript until they seemed tentatively confident and began making a number of predictions until they felt confident they had it.

Finkel (1993) provides another perspective on the students' experience:

> Class ends [on Day 2] with F's remark: 'We are thoroughly frustrated today.' The final day of model revising...begins with the group deciding that they don't have any more crosses to make, they just need to 'figure out the model.' The students then spend several minutes, alone and with the researcher, trying to figure out some of the other functions available on GCK. The researcher leaves and then returns. Finding the students still involved in 'playing with the computer' she asks them how things are going. They tell her that they have 'three models and they all suck' and when pressed, explain to her why they don't work. ... Shortly after the researcher leaves the teacher stops by and asks many of the same questions. The students explain their models and their problems again. ... Considering all of the models they have developed, M1 suggests that maybe they need a 'little combination theory here' an idea that does not get elaborated. This is followed by a long period of time in which the students talk about the computer and do not do any genetics. After a while the teacher stops by again to give

> them some encouragement. She suggests that they not give up so easily and proposes that maybe it would help if they tried to make a 'Mendel bible' for this model, as they did for the last one. This discussion stirs the group into action, although slowly. ... [For awhile the details of their 'three alleles' model are developed, with some success.] This feeling of success is short-lived. When a subsequent prediction does not hold up, F suggests that maybe they just have the genotypes "mixed up"; she is not willing to discard the model. ... They stick with the three allele idea even when they encounter setbacks such as predictions which do not hold up. Instead the group is tinkering with how to apply the three allele model to the data they have collected, how to assign genotypes to phenotypes in such a way that they can accurately predict the results of crosses. ... Using the crosses they have already recorded in their lab notebook and making new crosses as needed, the group tinkers with the three allele model until they have it working. (pp. 187-191)

In this experience there is a nice balance between perseverance and flexibility. Each model is pursued for awhile, until this pursuit no longer seems productive, and eventually the group finds a promising model that, despite its initial flaws, they do not want to reject. Instead, despite the difficulties encountered, the students persevere and develop the details of their model, revising it "until they have it working."

Z. Competition and Cooperation

One potential obstacle to problem solving is based not on genetics concepts, but on human nature. There is a tendency to prefer one's own ideas over the ideas of others; this competitive "pride of personal ownership" can decrease the objectivity of evaluation and the cohesive synergistic teamwork of a group's problem-solving efforts.

On the other hand, sometimes there is an attitude of unselfish group-oriented cooperation. In the episode described above, F decides that they should develop the three-allele model even though it was not "her model." This model was originally proposed the previous day by M1, with M2 quickly contributing a useful suggestion for modification; but M1 and M2 abandon it until F becomes its advocate and (along with the other group members) rescues the model from permanent rejection. This is an interesting example of "group ownership" of an idea, where instead of a competition with "my model versus your model" there is a cooperative effort toward doing whatever is needed to solve the problem, no matter where the ideas come from. { **psychological motives**, $2.31_{A,}$ operate in both **competition and cooperation**, $2.72_{B\text{-}C}$; also **persuasion** (and a willingness to be persuaded for the good of the group) occurs, 2.71_{E} }

The previous paragraph praises *cooperation,* but *competition* can also be beneficial by

providing **motivation** { 2.73_A } for a group (competing with other groups) or for individuals (competing with other individuals). And an excessive desire for teamwork (for its own sake, rather than to improve the team's effectiveness in achieving a goal) can lead to uncritical ‘group think’ conformity, especially if mutual criticism of ideas is not encouraged. An ability to use the advantages of both competition and cooperation, to combine them in an effective blend, is important in science and in all other human activities.