CHAPTER 11

OPTIMIZATION OF RESEARCH PART 2

I. Generality Maximization

- A. Conceptual precursor, terms applicable to issue of generality
- B. Maximized reliability
- C. Maximized similarity
- D. Maximized paradigmatic validity (understanding)
- E. Conceptual precursor: the assumptions underlying the chosen basis for generality
- F. Applications issues
 - 1. subject variables
 - a. prediction via understanding: synthetic subjects
 - b. prediction via similarity: representative subjects
 - i. random sampling
 - ii. rational sampling
 - 2. apparatus or setting variables
 - 3. procedural variables

II. Detectability Maximization

- A. Maximization of statistical detectability
 - 1. decrease criterion stringency
 - 2. increase signal to noise ratio
 - a. increase the magnitude of the signal
 - i. methodologically
 - (1) subject
 - (2) apparatus / setting
 - (3) procedure
 - ii. mathematically
 - iii. graphical illustration of increasing magnitude of signal
 - b. decrease the magnitude of the noise
 - i. methodologically
 - (1) subject
 - (2) apparatus / setting
 - (3) procedure
 - ii. mathematically
 - iii. graphical illustration of decreasing background noise

- B. Maximization of paradigmatic detectability
 - 1. conceptual precursor
 - 2. decrease criterion stringency
 - 3. increase signal to noise ratio
 - a. increase compellingness of the new paradigm
 - b. demonstrate failures in the existing paradigm

III. Meaningfulness Maximization

- A. Review precursor covariance
- B. Maximization of statistical meaningfulness
 - 1. increase percentage of accountable variance
 - a. increase magnitude of effect (signal)
 - i. methodologically
 - ii. mathematically
 - b. increase magnitude of noise in baseline
 - i. methodologically
 - ii. mathematically
- C. Maximization of paradigmatic meaningfulness
 - 1. increase ability of paradigm to predict nature
 - a. your ability to use the paradigm
 - i. understand its assumptions
 - ii. understand how it works
 - b. the paradigm's ability to predict nature
 - i. very broadest most fundamental aspects
 - ii. middle level traditional aspects
 - iii. shallow or least established aspects

CHAPTER 11

OPTIMIZATION OF RESEARCH PART 2

I. Generality Maximization

Information which is undependable is not a fact. To say something is a fact is to say that it can be made to reoccur or be replicated, i.e., it must be reliable to be a fact. We know that if we follow a key peck with food then the response rate will increase. This effect occurs each time we do it. To say something is generalizable is to say the information applies to a new situation of interest or that it is applicable to a new situation. This is referred to as external validity. Not only will following a key peck with food increase its rate, but following singing with cheers will increase its rate. Keep in mind that we are not really interested in the "generality" of individual events but rather in the generality of functional relationships, in this case reinforcement. In addition to reliability then, it is necessary that functional relationships or knowledge have generality. Information which is not repeatable or does not generalize to other situations is worthless. The test of reliability is the ability to be directly replicated or repeated under exactly the same conditions. The test of generality is the ability of a relationship to be systematically replicated or repeated in somewhat different situations. As everything in science, generality, in the last analysis, is empirically determined. If the functional relationship predicts in a new situation, then it is said to generalize.

The ability to generalize information from one situation to another is a function of several factors. The reliability of the original information; the paradigm's validity; your understanding of the paradigm and the true determinants of the behavior and the relevant details of the situations in question; and the similarity between the original source of the data and the situation to which it is to be applied.

There are both similarities and differences between the terms "stimulus generalization" and "generality of a functional relationship." Stimulus generalization is the description of the fact that an organism behaves in a similar way to similar stimuli and that the more different the stimuli the more different the behavior. The generality of a finding refers to the degree to which a functional relationship obtained in one situation is able to predict the obtained relationship in a new situation.

A. Conceptual Precursor, Terms Applicable to Issue of Generality

All individuals together are called the <u>population</u> whereas a subset of the population is considered a <u>sample</u>. If information is applied to a new situation we speak of that situation as a <u>target</u>, and the situation which was used to generate information applicable to that target as a <u>model</u>. For example, typically a sample is measured and the information such as "response allocation matches reinforcement allocation" is taken as an indicator that that relationship also exists in the population.

B. Maximized Reliability

All the issues discussed in the section on reliability maximization are obviously appropriate in assuring generality. If two observers cannot agree on the nature of a functional relationship in the original situation, it is unlikely that agreement will occur from the original situation to a target situation. Therefore we can increase our expectation of generality by maximizing reliability.

C. Maximized Similarity

If a functional relationship is obtained and is then applied to itself, it, of course, applies; it is a simple description. If the original situation changes very slightly between the time the relationship is identified and then applied to itself the information will probably still correctly describe or predict. However, it can be seen that the less similar the original and target situation the less likely that the finding will apply to the dissimilar target. We may be able to increase the likelihood that information gained in one situation is applicable to another situation by making the two situations as similar as possible, hoping that the elements which are different are not important, and hoping that the factors which are the same are important. This is better than nothing but not entirely satisfactory. First, it may not work. For example, humans are similar to humans and fish are different from humans, however voting patterns of humans earning over \$50,000 per year; whereas the effects on a fish of being dropped onto the surface of the sun predicts very well what will happen to humans dropped onto the sun.

Secondly, a dependence on similarity to maximize generality represents a rather passive, timid and superstitious approach to science. If we deliberately minimize the difference between the model and target, then we minimize the opportunity to learn why things work the way they do by eliminating the knowledge gain provided by the generality of findings. In the extreme, we could maximize similarity by using each patient as both the model and the target for their own therapy. Obviously, that would be nothing less than the abandonment of science. We would have to start from scratch with the next patient.

It is clear that the issue of generality is related to the definition of the word "similar." If the model and target are similar with respect to the controlling variables of interest then the functional relationship will generalize. If they are similar with respect to irrelevant dimensions and dissimilar with respect to relevant dimensions, then the findings will not generalize. If we consider two groups similar because the effect of interest generalized, then we cannot then turn around and claim that the function generalized because the groups were similar. This is a tautology. Basing expectation of the generality of a functional relationship on naively perceived similarity represents either an unsubstantiated hope or a tautology, neither of which provide much in the way of advancement of understanding. Simply because the research was done on primates we are not assured that the functional relationship will generalize to humans. Simply because the functional relationships were obtained from worms we are not assured that those findings will not generalize to humans.

D. Maximized Paradigmatic Validity (understanding)

Paradigmatic validity refers to whether or not different situations are actually related to each other the way the paradigm suggests that they are. Whether or not a functional relationship obtained in one situation will generalize to some target situation depends on whether or not the actual causes of the effects in the original situation are going to reoccur in the necessary way in the target. Your ability to predict whether or not the effect will generalize depends on your understanding of the actual causes of the behavior in the original situation and your correct understanding of whether or not those same factors are operative in the target situation. You ability to generalize the finding is therefore affected by the validity of your paradigm. It is important to realize that two elements are needed to predict whether or not an effect will generalize to a new situation: 1) exactly why it occurred in the original situation and, 2) how the new situation relates to the original situation with respect to the relevant details. In an entirely alien reality you would have great difficulty predicting what would happen in new situations even if you knew why things happened in the original situation. This is because you would have a very poor idea of the circumstances in which to expect those same controlling factors again. Rather like Alice in Wonderland. You cannot make accurate predictions unless you know the rules by which the situations are connected. A well-known example of true rules of controlling variables being other than what they appear on the surface would be the Abbott and Costello comedy routine "Who's on First." The joke pivots on the realization that who (etc.) is a person's name and their assigned position is to play first base.

E. Conceptual Precursor: The Assumptions Underlying the Chosen Basis for Generality

In sum, knowledge can be generalized on the basis of simple naive notions of similarity, or on the basis of understanding what's going on. This has led to two

research strategies: 1) research emphasizing similarity to maximize generality, and 2) research emphasizing the understanding of the mechanisms of action and the control of confounding factors to maximize generality. "Similarity" research is often done in the "real" world with "real" people or at least primates while "understanding" research is often done in the laboratory with synthetic situations and the most apt non-human subjects. "Real world" subjects or settings are maximally similar along some dimension assumed to be important in an effort to establish generality of the obtained functional relationships. Unfortunately they typically introduce many confounds which seriously undercut the actually obtained generality, as well as limit the information they produce. The raison d'etre of synthetic subjects or settings is to maximize our ability to understand the true causal factors and thereby maximize the generality of the obtained functional relationships. Unfortunately synthetic subjects and settings can fail to produce generalizable information due to misunderstood or unaccounted for differences between the model and target.

If the task is simply to develop knowledge applicable to a narrow range of situations such as for a simple technical knowledge, then naive notions of similarity will often adequately serve as a basis for predicting generality. This is the rationale underlying vocational training programs, such as Nursing and Psy. D. degrees. It is important to note however that if the person is expected to predict in a variety of situations, errors of prediction will occur, and those errors could have important consequences. That is, of course, what separates professionals from technicians. It is the well accepted difference between nurses and physicians. Vocational training or practical experience need not prepare people to generalize into different situations while professional education must.

F. Applications Issues

Rarely will you be content to describe functional relationships applicable to only the exact research that you carried out. Typically you will want to make statements of wide generality. The generality of research is affected by choices made with respect to the subjects, apparatus, and procedure.

The choice of the determinant or basis of generality affects the decisions a researcher makes in each of these areas. In each case, the emphasis can be to maximize the similarity between the research and the target or to maximize control and aptness.

1. Subject Variables

Psychology is the study of behavior. Behavior comes only from living animals, therefore at least one organism is necessarily involved in any psychological research. The subjects used must in some way give us knowledge we can use in other situations.

a. Prediction Via Understanding: Synthetic Subjects

Synthetic subjects are often chosen to allow the maximum degree of control over confounds and to be maximally revealing of the phenomenon of interest. Generality in this case is based on understanding what is going on. Additional important considerations include the convenience of using a particular synthetic subject. A rat is simply easier to deal with than a whale or a person. The optimal synthetic subject is also one which is optimized to reveal information on the dimension of interest, e.g., giant axons of squid. This is especially true in pure research settings where the goal is understanding general mechanisms of action rather than developing an analog of a specific applied problem.

b. Prediction Via similarity: Representative Subjects

Alternatively, representative samples can be used. Generality in this case is based on making the subjects as similar to the target as possible. If you measure each of a small population from which you determine a functional relationship, you have therefore measured the entire target population and your finding, will obviously apply back to its source. Unfortunately, if you are interested in predicting the effects of some variable on any member of the entire population of humans on earth, you would have to measure the effects on every member of the entire population. The only alternative is to measure a sample and then infer information about the effects on any given member of the entire population. To be meaningful, the sample must provide information which generalizes to the entire target population, (i.e., all humans). The sampling techniques which you use are very important in that they can dramatically affect the generalizability of your finding. In order to be useful your sample must be similar to the target population or "representative."

i. Random Sampling

You may take advantage of the properties of chance to assure yourself of representativeness and therefore generality by randomly sampling from the population. Random sampling guarantees (other things being equal), that the obtained functional relationships are accurate because a random sample will produce as many deviation above as below the true value on all dimension simultaneously. If you randomly sample enough individuals you will get an accurate picture of the population and your findings will generalize. If your population or sample is small special precautions are necessary to maintain representativeness. For example, the probability of obtaining an individual can change as the population is depleted if subjects are not replaced.

ii. Rational Sampling

An alternate sampling procedure is to sample the population such that you

deliberately match the sample to the population according to the relevant factors you choose, such as proportionate numbers from each strata. This procedure is used when you cannot obtain a large enough random sample to make it representative. Rational sampling requires that you know in advance the relevant properties of the population and that your sampling procedure guarantees representativeness.

2. Apparatus or Setting Variables

Assuring that the subjects that you use are a representative sample of the population is not sufficient to assure generality of your obtained functional relationships. The impact on generality of the apparatus, setting, and procedure are equally important but often overlooked.

Apparatus or setting variables can strongly affect the generality or the applicability of the functional relationships derived in one situation and applied to another. The selection trade-offs are the same as those involved in selecting subjects. Maximum similarity versus maximum control.

"Real world" situations maximize generality through similarity, but also typically maximize error through lack of control. Additionally they are often not optimized to provide the data needed to understand the phenomenon. Synthetic situations on the other hand minimize error through the deliberate choice or design of the apparatus or setting. However, they are often specifically chosen for their ability to reveal fundamental relationships between classes of inputs and classes of outputs rather than to identify important controlling variables in specific applied settings, and therefore many not provide specific solutions to specific problems.

3. Procedural Variables

If a process inferred by our theoretical structure (e.g., fear) is in fact a correct description then the procedures we use must accurately reflect that inferred theoretical process. If electric shock is presumed to cause fear then it must do just that and all procedures which are assumed to produce fear must produce the same functional relationships. Otherwise, a finding demonstrated with one procedure may not generalize to a situation where the inferred variable is produced or measured in some other way. For example, if verbal behavior is used to index some actual behavior such as compassion, then the obtained data must generalize to actual situations. If someone says that they will help in a crisis and we use that fact in our theory, then they must help in actual crisis situations. Many confounds may interfere with the generality of findings obtained with a specific experimental procedure. For example, if the experimenter establishes demand characteristics or other confounds which do not exist in the target procedure, what the person says may not accurately reflect what the person does. The selection dilemma for procedural variations is the same as with subjects and the apparatus. Generality based on similarity versus generality based on understanding.

II. Detectability Maximization

A. Maximization of Statistical Detectability

Detectability refers to an ability to correctly identify events that are nonchance regardless of how dramatic or subtle. An emphasis on the detectability or power of a test (the avoidance of misses) generates the opposite pressure to that of reliability (the avoidance of false alarms). The more extreme the criteria for judging that an event is really unlikely by chance (e.g., only if you hear roaring and trees splintering in the outdoor camping example), the less likely you will mistake chance for real effect (i.e., make a false alarm) and the more likely it is to occur again given the same causes. Or in other words the more reliable the event will be. Unfortunately, having such a high criterion also makes it more likely that you will overlook subtle but none the less real effects (e.g., not run away when a bear is simply walking up to the camp quietly) (i.e., make a miss). If you decide to be extremely conservative before accepting that an event is an important discovery, it is likely that you will miss some important phenomena that are really there, but that are only subtle. Your decision mechanism will have poor detectability.

By simple reasoning, it can be seen that several things can be done to maximize your ability to realize when you have a research or therapeutic discovery when the effect is actually small or to increase the detectability of your procedures. You may wish to review the logic underlying the declaration of a true effect in Chapter X IV. B. 1. b before continuing with this section so that you can consider the two edges of the sword of decision making at the same time.

1. Decrease Criterion Stringency

The stringency of the criterion for claiming an effect can be decreased (from D to C in the illustration below). In this case, you would then be more likely to claim an effect when there actually was one (fewer misses). Unfortunately, however, you are also more likely to claim an effect when there is no real effect (more false alarms).



This is not an entirely desirable approach, but it does temper the tendency to make extremely stringent reliability requirements. In essence you can find more cures (fewer misses) by claiming that you found a cure for schizophrenia with less substantial results. However, a better strategy would be to decrease both false alarms and misses as discussed below.

2. Increase Signal to Noise Ratio

This refers to making the treatment effect stand out more. Anything that increases the signal to noise ratio will increase the likelihood that an effect will be detected at a given criterion of hits to false alarms (level of significance).

a. Increase the Magnitude of the Signal

i. Methodologically

(1) Subject

Choose subjects that react strongly to the treatment.

(2) Apparatus / Setting

Choose an apparatus which is especially sensitive to the treatment effect.

(3) Procedure

The more you are "in tune" with the natural laws governing the behavior then the greater the percentage of the whole "true" effect you will be able to control. Getting in tune depends on how well you understand the paradigm and how accurately your chosen paradigm describes nature.

Experimenter finesse is the term for your ability to devise an experiment which maximally illustrates the effect. This would mean that you are able to take advantage of special characteristics of a particularly apt subject; an apparatus which maximally facilitated the effect; a procedure which maximally reveals the effect; using independent variables which produced strong effects; or dependent variables which were very sensitive to the behavior. To exercise greater experimenter finesse in the camping example, string tin cans around the camp. Anything entering will make louder sounds. The signal (bear) will make much more noise than the wind in the trees. If you wish to hear a visitor at the door during a loud party, make the door knocker very large.

ii. Mathematically

Index the dependent variable over a wider time period. (e.g., whereas it may be unlikely that a single noise of a particular amplitude is indicative of danger, it is quite a different case if noises of that amplitude occur in a systematic pattern, like footsteps. In that case you should run. You would be able to detect a very quiet bear by sampling information over a wider time period.)



iii. Graphical Illustration of Increasing Magnitude of Signal

Increase Signal to Noise Ratio

The preceding figure illustrates the effect of increasing the signal (bear) to noise (wind) ratio by increasing the magnitude of the signal (bear) by stringing cans around your camp, for example. If a bear is around, the quietist noise it can make is the sound of tin cans. If a bear makes a noise at all it tends to be very loud (amplify or multiple all scores in bear noise distribution by ten). Where you were only able to detect signals of size "thirty" or greater initially, you can detect signals of size three with the same number of false alarms, because every bear sound is ten times bigger now. What this does geometrically is to spread out the bear noises: each point on the "bear curve" is moved to the right by a factor of ten. A noise that had been above the 3 on the x-axis is moved to the right to 30. This gives the appearance of shifting the bear curve up. To be most precise and most general, the interest is also to shift events on the other side of the mean down. The simplistic bear example is strained a bit with this addition, but it would be a dangerous animal signaled by a lower background noise level to make even deeper silence. For example, if some demon was walking in the woods, not only were the crickets quiet but all sounds including the wind stopped.

b. Decrease the Magnitude of the Noise

i. Methodologically

(1) Subject

Choose a subject which is especially insensitive to the most likely confounds.

(2) Apparatus / Setting

Use an apparatus which shields against the major confounds.

(3) Procedure

Exert tighter technological control to reduce noise levels. In the camping story, eliminate all the safe things that make noises throughout the night. You will then be able to easily hear a bear roar. Run pigeons in sealed unchanging chambers with few distractions. The variations in the behavior caused by chance or irrelevant factors would then drop to a very low level and the effects of very small treatment effects (signals) would be made obvious. If you wish to hear a quiet door bell at a party tell the people who are making noise to be quieter.

ii. Mathematically

Take a larger sample so that the likelihood that an event that large occurring by chance is lessened, only if the noise level is up for a whole 30 second period rather than only for 1 second. Simple random error will more closely approximate zero when sample sizes are large.



iii. Graphical Illustration of Decreasing Background Noise

The preceding figure illustrates the effect of increasing the signal (bear) to noise (wind) ratio by decreasing the magnitude of the noise (wind). Very few samples of ten consecutive chance noises get very loud (divide all the scores in noise distribution by some value). What this does geometrically is to squeeze in the noise distribution.

B. Maximization of Paradigmatic Detectability 1. Conceptual Precursor

As you have seen statistical decisions are based on an explicit acceptance of the trade-off between better detectability (decreasing misses) and better reliability (decreasing false alarms). In statistics we had decided to minimize misses consistent with no more than 5 false alarm errors per 100. We accepted some errors of one type in order to attenuate errors of the other type.

A similar trade-off between two types of errors also faces us when we make decisions about what obtained functional relationships are plausible, what interpretations are acceptable and even about what paradigm we accept. This choice of paradigms is the most fundamental question we will face, because it governs the types of questions we will ask and what we consider acceptable answers. There is a trade-off between, on the one hand, the absence of serious research on phenomena (and/or their interpretation) which may really be correct but which are inconsistent with the existing paradigm, and on the other hand the necessity of science to address itself to meaningful questions and to provide a credible theoretical framework for the phenomena on which it is focused. Consistency with the paradigm and paradigm revolution are the opposite poles of paradigmatic decisions just as reliability and detectability are opposite poles of statistical decisions.

Unfortunately, paradigmatic controversies are rarely presented in terms of how many false alarms we are willing to risk to decrease our misses by some amount, even though the statistical analog of this problem is well thought through. For example, at what point is it worth serious effort to investigate extraterrestrial causation of ancient artifacts. It is certainly a possibility, and it certainly provides a simple answer for some phenomena.

To elaborate the example, the paradigmatic issue of whether or not space travelers built the pyramids (i.e., an interpretation of data) can be plugged into a decision matrix.

_ _

		REALITY	
		TRUTH was spacemen	TRUTH was human
C L A I M	SAY was spacemen	Correct Detection of new paradigm	False Alarm (say spacemen when not)
	SAY was human	Miss (say human when was spacemen)	Correct Rejection of erroneous paradigm

Our interpretation could be wrong in two ways: 1) claiming that spacemen built the pyramids when humans built them, and 2) claiming that humans built the pyramids when spacemen really built them. Similarly, our paradigm could be right in two ways: 1) claiming that spacemen built the pyramids when they did, and 2) claiming they did not when they didn't. Both claiming spacemen causation and claiming human causation has its advantages and its problems and each has some evidence supporting it and some evidence refuting it. Additionally, there is a cost benefit ratio to each decision. (Note that there is no disagreement concerning the existence of the pyramids just where they came from.)

We must weigh all aspects of all factors. We must consider the weight of the evidence for each alternative as well as the cost and benefit of each possible outcome. What we must simultaneously consider is the gain if we accept extraterrestrial causation and it is true, with respect to the cost if it is false and we claim that it is true; versus the gain if the extraterrestrial causation is false and we claim it is false with respect to the loss if it is true and we claim that it is false with respect to the loss if it is true and we claim that it is false.

The issues involved can best be illustrated with several figures. They represent the factors underlying decisions and therefore are topologically similar to those used in statistical decisions. However, the actual functions could be quite different. Gaussian distributions are used to illustrate paradigmatic relationships but we have no evidence that the distributions relating frequency of occurrence to deviation from the mainstream paradigm are normally distributed. Consider the functions as simply falling off with increasing distance from the central tendency with the specific shape arbitrary.

The first figure below is intended to illustrate that: 1) only findings and their associated interpretation within the range of the paradigm's tolerance are accepted as plausible findings and acceptable interpretations; 2) Some findings or interpretations may be inappropriately integrated within the existing paradigm (missed opportunity to advance to a better paradigm); 3) some facts could falsely be attributed to error or to an alternate paradigm when they are better integrated within the existing paradigm (false alarm); and 4) some findings may be so aberrant that we must change our paradigm in order to accommodate them and we are correct in doing so.



Note that we presume that the existing paradigm is correct (temporally switching to the wind and bear example even though it is a strained example, we have total confidence that there are no bears). There are three aspects of this position: 1) all events are considered the result of the accepted paradigmatic laws (everything is wind); 2) there are no events which are the result of a different paradigm (there are no bears); and 3) the most deviant event possible is either the operation of the accepted paradigm in an unusual situation (strong wind) or is an error. This confidence in the paradigm is best illustrated by considering a magic trick. To a child, a new paradigm (magic causation) is required by the very abnormal event (coin disappearing). To the physicist, the event was <u>by definition</u>, an illusion and represented the operation of the accepted paradigm, just as well as any other observation. There <u>can be no such thing</u> as a

disappearing coin without our current physics paradigm.

However in principle, we must accept that both misses and false alarms could happen. Regardless of our confidence in the current paradigm, it is still possible for it to be wrong. Note the criterion lines in the above figure and that there is "current is correct" beyond that deviation and "current is wrong" before that criterional deviation. Misses are those findings which appeared to be well explained and integrated by the existing paradigm, but which are, in fact, instances of the operation of a different paradigm. False alarms are those aberrant findings and unacceptably different interpretations which are in fact simple examples of the operation of the current paradigm. Correct detections are, of course, instances where the existing paradigm correctly explains phenomena and correct rejections are instances where a new paradigm was correctly invoked to explain an event.

This second figure illustrates the gain/risk principle underlying choices of paradigms. Note again that the actual shape of the deviation that the prototypical paradigmatic statement and potential for gain functions are unknown. Take the shapes as arbitrary.



The preceding figure illustrates that as findings and their interpretation deviate from the normally accepted paradigm, the potential for gain if the findings and interpretations are correct increases. However, their probability of being correct dramatically decreases. A choice of doing paradigmatic research versus doing nonparadigmatic research can therefore be seen as somewhat like the choice between taking a single step in a direction very likely to be toward your destination or taking a step in a direction which is most likely wrong but has a remote possibility of moving you 10 yards toward your goal. The paradigm specifies a range of views which optimize the gain/risk decision. Many small but correct steps gain more in the long run than a few big steps likely to be wrong. This is one of the reasons that paradigm changes are reserved for extreme situations.

The next four figures illustrate the factors underlying an actual choice of a paradigm. In order to illustrate the factors, three dimensional figures would be needed. Unfortunately, you must, therefore, combine the next figure which gives the front view (right to left) with the second figure (which gives you the front to back) and then go back and similarly combine the next figure with the third figure (which gives a more correct version of the front to back information).

Absolute final wisdom could be considered as being at the minimum point of a function indicating complexity of conceptualization (the low point in the curved function provided below). Final wisdom could be seen as the point which minimizes complexity in the universe.



Our current paradigm (#) is that perspective which minimizes complexity to the best of our knowledge. Paradigm shifts in the past have occurred whenever net complexity was reduced by so changing (e.g., Ptolemy to Copernicus). All attempted shifts to some other paradigm could be seen, therefore, as either decreasing the eventual best possible net complexity, or increasing the eventual best possible net complexity. Those paradigm changes which reduce best net complexity are seen as good, those changes increasing net complexity are seen as foolish.

These next two figures illustrate the source of the difficulty in evaluating the appropriateness of various competing paradigms. If different paradigms were better or worse at handing all findings but within each paradigm every conceivable finding was handled equally well then net complexity and local complexity would be the same thing and the following figure would provide the front to back perspective of the preceding figure. (Imagine looking at the above figure from the right side, with the left side away from you.)



But, the above figure does not represent how various paradigms actually handle various local findings. Each paradigm does not handle every single finding equally well; as a result, net complexity is determined by how the paradigm handles many specific local phenomena in an overall sense. For example, burning down your house is a very simple and fast way to get rid of mice. That approach is extremely good at solving the mouse problem. However, you create a very large problem concerning where you are going to sleep at night. A paradigm does not handle every phenomena equally well. A good paradigm must be good in its net result. The following figure more correctly illustrates variations in the ability of various paradigms to handle both local complexity and net complexity.



As can be seen in the above figure, the net or average complexity of A is greater than the existing paradigm and the net complexity of paradigm B is less than the existing paradigm. This is the case even though for phenomenon I paradigm A is less complex than the existing paradigm. For phenomenon II paradigm B is worse than even paradigm A. Paradigm A may offer the mirage of simplicity when it claims that space men built the pyramids. The good part is that it solves in a very simple manner, the origin of the pyramids. The problem is that the other increases in complexity that that notion brings. Is there any evidence at all that space men did it other than our inability to figure out how they did it? How did they get here? Why did they use wood and stone tools? Where did the space men come from? Are the artifacts on their home planet the result of other ancient astronauts, perhaps ancient Egyptians? How do we explain the artifacts on the first planet in the universe? Where does it end?

When it comes to deciding whether or not to shift to a new paradigm because it handles some problem better than the existing paradigm, the task is how to gain by reducing our net false alarms without unacceptable increases in our net miss rate or vice versa. Historically, those people who have been most concerned with eliminating false alarms at the risk of accepting a few misses (stick with existing paradigm even when it seems like it should be abandoned) have been dramatically more productive at both arriving at the most consistently predictive models and in helping mankind. Wild flashy theories like wild goose chases have just not paid off. Still we can think through those factors which govern the tradeoff so we can better decide when to call for a paradigm change. The following figure is an expansion of Figure X with normal science and paradigm shift illustrated in more detail.



Advancement in understanding nature within a particular paradigm can be conceptualized as movement toward less net complexity of formulation by moving down the columns of numbers which represent degree of "error." Normal science can be seen as moving the implementation of a paradigm to its best possible implementation (movement from 1.00 to 0.00 down a vertical line which represents the reduction of net complexity within the paradigm). At any point a researcher can conduct normal science and better account for phenomena within the paradigm (i.e., in the existing paradigm movement from 0.90 toward 0.30) or the researcher can move to another paradigm such as paradigm shift to Paradigm A or to Paradigm B. For example, from "Existing Paradigm Position 0.60" a researcher could develop the following views: Paradigm A Position .70 (an increase in complexity); A 0.60 (no change in complexity); A 0.50 (a decrease in complexity); Existing Paradigm Position 0.50 (a decrease in complexity); Paradigm B position 0.70 (an increase in complexity); or to B 0.60 (no change in complexity); or finally to B 0.50 (a decrease in complexity).

Historically, researchers who advance the Existing Paradigm (i.e., E 0.60 to E 0.50) virtually always advance knowledge. Researchers who shift paradigms virtually never advance knowledge. It is not a 50/50 chance that the alternate paradigm will be better as the figure implies, rather it is thousands to one against.

Shifts to alternate paradigms are usually made because the researcher sees a local reduction in complexity (Point I in figure x). However, the chances are high that there will be an increase in net complexity. An additional problem with paradigm shifts is that science will advance by shifting to Paradigm A 0.50 from E 0.60 but A at its best (0.40) will have more complexity than the Existing Paradigm at its best (0.30).

The problem of choosing to conduct normal science (E 0.60 to E 0.50) and when to be the leader of revolutionary science (E 0.60 to A or B 0.50 or 0.60 or even 0.70) is more apparent than real. The articulation of a new paradigm (A or B) is extremely rare and universally rejected until the existing paradigm is at its maximum capacity (E 0.30) so that movement to a different paradigm is virtually assured to be in the right direction (E 0.30 to B 0.20). Going from E 0.60 to A 0.50 appears to be a good move but it is a dead end. Movement to paradigm A will not occur when the existing paradigm is at its limit. E 0.30 to A 0.40 will be an obvious increase in complexity.

The poor likelihood of success of a paradigm shift is well illustrated by the fact that the Ptolemaic paradigm was maintained for quite a while in spite of glaring anomalies. Because it had exhausted gains possible through normal science, the Copernican view was tried even though at that time it had more complexity and accounted for less than the Ptolemaic solar system. Given all these problems associated with paradigm change, we can consider those factors which may increase our ability to know when to abandon the existing paradigm and accept a new paradigm (maximize our detectability). That is, we can try to identify those totally anomalous but real phenomena or totally alien but correct interpretations which require a paradigm shift because they deviate more than is acceptable from the existing paradigm. If we can correctly spot them early in the game we are in a better position to leapfrog toward "total wisdom." Again the following figure shows Gaussian functions but the actual functions are unknown and could be anything.



One could hope that by decreasing the demand for paradigmatic consistency, we may increase the likelihood of arriving at a better paradigm, but this is both true and false. It will increase the probability of finding a better paradigm (correctly detecting new paradigms) but it will also increase the likelihood of arriving at a wrong one (false alarm). In point of fact, it is more likely to arrive at error. Because the number of ways to be wrong vastly outnumber the number of ways to be right, poorly thought through theories or sloppy research cannot be justified simply by claiming that they will decrease misses, even though to decrease misses we may need to be less stringent in our emphasis of factors which maximize theoretical consistency because the theory is immature. The problem is that increasing consistency may decrease detectability but decreasing consistency does not necessarily increase detectability. Sloppy thinking or sloppy research is actually less likely to discover a better paradigm than research with an extremely high criterion for consistency. This issue can be illustrated with the following figures. As can be seen, it is essentially the same as the one used to illustrate decision theory trade-offs.



Our competing interests are illustrated below.



Variations in the criterion for acceptable hypotheses varies the frequency of accepting false paradigms and missing correct paradigms. More stringent criteria are the result of demanding more consistency. Less stringent criteria attempt to increase detectability by decreasing the demand for consistency. It is a dilemma.

The key, of course, is what we can do to get the best of both worlds. What factors affect the proper execution of normal science and what factors affect the proper point of paradigm revolution and how do they interrelate. How can we have our cake and eat it, too.

The following are ways to increase paradigmatic detectability.

2. Decrease Criterion Stringency

Increasing the demand that data and interpretations be reliable or consistent will exclude with the paradigm increasingly more phenomena and from consideration. Typically consistency interpretations demands are deliberately set at some criterion with the intention to exclude questionable phenomena or interpretations that are beyond that deliberately established criterion. Note however that the deliberate attempt at other times is to find phenomena rather than to assure that the finding is consistent with a large theoretical structure.



What we could do is to decrease the stringency of when to ump to a new paradigm. We could be less diligent in finding a paradigmatic explanation for phenomena resistant to normal explanation. This would allow us to accept some other paradigm which has less support and to abandon our existing paradigm with less provocation (because we cannot make the paradigm explain our findings.) If we cannot figure out how the Egyptians drew straight lines, we can propose that the devices to draw straight lines were brought by space ships.

The cost of this decrease in our criterion for abandoning the existing paradigm is an increase in theoretical false alarms. We will more often propose erroneous paradigms which waste a lot of researcher's time. But we will also be more likely to find a better paradigm. The down side of this simplistic approach is the fact that the likelihood of jumping to a productive paradigm is nearly zero.





The cogent presentation of well-supported facts which are clearly outside the ability of the current paradigm's explanatory power. This is increasing the magnitude of the signal. Along with an explanatory framework which:

- 1. reduces complexity/increases scope
- 2. promises even further simplification with further normal science
- 3. is supported by experimentum crucis

The value of this type of research is that it decreases the likelihood that the proposed alternate paradigm is wrong. If we present enough solid evidence integrated into a coherent paradigm, we can make geologists of the 1940s believe plate tectonics.

In terms of the above figure, this type of research makes a particular finding (x-axis position) more compelling. The finding is moved out into a region clearly not covered by existing paradigm. The obtained functional relationship "stands alone" well outside the explanatory power of the existing paradigm.



b. Demonstrate Failures in the Existing Paradigm

The quantitative formalization of a general qualitative theory narrow the range of apparent applicability of that theory. After the fact, a vague theory would appear to account for everything. With clarification, many apparent explanations and consistencies are empty. With a quantitative model, errors of prediction are more apparent and a better model is easier to identify.

Alternatively, the "noise" distribution can be narrowed with the cogent presentation of well-supported problems with the explanatory power of the existing paradigm which:

- 1. dramatically limit the range of solutions which the paradigm provides
- 2. show complexity in existing paradigm
- 3. show that complexity is likely to increase

What this does is to "decrease" the number of phenomena which the existing paradigm handles well by pointing out inconsistencies. This is reducing the magnitude18

of the noise. Unfortunately, this is equivalent to increasing the heat under the "frying pan" in which someone sits. Without a better alternative to which to jump, the result of the jump is as likely to be into the fire, as anything else.

III. Meaningfulness Maximization

It is not enough simply to maximize the likelihood that someone else will get your results if they repeat your procedures; or that they will get your results if the methodology varies in some way; or that you made every effort to find small effects. Your findings must also be meaningful. If you are an emergency room medic it is more meaningful to discover arterial bleeding than ingrown toe nails. If you are charged with finding the source of a problem in your patient, it is more important to discover a willingness to commit suicide than a smoking problem.

This basic concept takes two major forms in psychological research: 1) it is important that variations in your independent or predictor variable account for a large proportion of the variation in your dependent or predicted variable and, 2) You must maximize the ability of your paradigm to make sense of all the "booming," "crashing," apparent randomness in nature by maximizing your understanding of the existing paradigm and/or making a better paradigm.

Maximal reliability, generality, and detectability are necessary for maximal meaningfulness, other things being equal.

A. Review Precursor - Covariance

Covariance is a very important concept in science. It has been discussed in Chapter 2. To review, it is the amount of uncertainty that has been reduced by the knowledge of something else. For example, if you had to guess a number from 1 to 10, you would have 10 possible guesses. The range across which you must guess is depicted by the size of the left circle. If you knew that the answer was an even number, then you would have reduced your uncertainty by half. You would only have 5 possible guesses. You would have accounted for half the variability. The overlap of the right circle and the left circle depicts the range you no loner have to worry about. The odd numbers are removed from the range across which you must guess. A Venn diagram illustration would show half the variability in Y accounted for by the variability in X.



Alternatively, we could use a scatter plot illustration. Suppose you had a distribution of grades from 0 to 100 that was normally distributed. It is shown plotted on the *y*-axis. Suppose further that you had a distribution of studying that varied from 0 to 40 hours per week that was plotted on the *x*-axis. Finally, suppose that the relationship between those two measures is as illustrated in the scatterplot. People who study less tend to get lower grades; people who study more get better grades.



In this case, accountable variance is the amount of variability on the *y*-axis that has been removed by the variability on the *x*-axis. Your original guess for a person's grade point average had to be across the entire range of the *y*-axis, while with a knowledge of x, your guess only had to be across the range of the dots at that point on the *x*-axis which is illustrated by the small distribution in the lower left (i.e., you can come considerably closer to a correct guess).

Both the Venn diagram and the scatter plot illustrate the same effect. It is important to note that the overall variability on y remains the same. What changes is the range across which you must guess. (In point of fact, it is not the range but rather the likelihood of being correct and the product moment of your error if you guess the mean.)

When research is reported, an important question to ask is "what is the percent of accountable variance?" You are asking "will this information contribute very much to my understanding of behavior." For example, if one person told you that spending 20 minutes sleeping just before class increased your grade and it was a reliable effect (p<0.05), and someone else said that reading the text for 20 minutes just before class increased your grade and that too was a reliable effect (p<0.05), you should ask "by what percent will my grade increase in each?" The first person would answer 1-5%; the second person would answer 60-90%. It is obvious what you should choose to do before class. The Venn diagrams below illustrate the difference in accountable variance in the too suggestions.



B. Maximization of Statistical Meaningfulness 1. Increase Percentage of Accountable Variance

In this case, it is the percent of the variance in the obtained measure accounted for by some other obtained measure.

a. Increase Magnitude of Effect (signal)

This approach includes all the previously discussed methods for increasing the magnitude of an effect either through methodological or mathematical procedures.

i. Methodologically

This is a choice of subjects, apparatus, and procedures that react strongly to the treatment. For example, procedure could specify a large dose of drugs or other independent variables in order to produce a strong effect.

ii. Mathematically

Large sampling windows capture more information, just as larger telescopes or larger satellite dishes capture more signal. While hearing a single thump may not mean much, a series of thumps in the pattern of foot falls getting louder are a strong signal that something walking this way comes!

b. Increase Magnitude of Noise in Baseline

Increasing the magnitude of noise in the baseline in order to increase the percent of variance accounted for requires that the effect is robust and will be equally obvious in spite of the increase in noise. Showing that the effect is robust in the face of noise increases our belief in the generality of the effect.

i. Methodologically

Start with only a few animals. Use a variety of subjects, apparatus and procedures. If the effect is reliable with a low power test, then it is a relatively

strong effect capable of swamping the noise.

ii. Mathematically

Take a smaller sample or use a statistical test of lower power. If the effects is clear with only a short measuring period or with only a few subjects, then it is a very strong effect.

C. Maximization of Paradigmatic Meaningfulness1. Increase Ability of Paradigm to Predict Nature

In this case, it is the percent of the variance in nature that is accounted for by the paradigm. the error in the paradigm's ability to predict can be partitioned into that attributable to the error you make in applying the paradigm to your obtained data, and the error attributable to the inability of the paradigm to account for the finding (this latter discrepancy is typically assumed to be zero in the practical conduct of science).

a. Your Ability to Use the Paradigm

The issue here is the degree to which you are able to use all of the explanatory power of the paradigm. The better you understand the existing paradigm, the better you will be able to use that knowledge base to solve problems and to understand your obtained functional relationships. It is pretty obvious that the more case precedents a lawyer knows the more likely they will win, and the better the vocabulary the better the poet. Equally so with psychology and every other science. Using the steam shovel at the bottom of the lake metaphor, the more you know what other people have already discovered about the sunken object and the more you know about lake bottoms and probing tools, and the more you know about what kinds of things are possible to be down there, the sooner you will be able to discover what is actually at the bottom.

It is for this reason that reading pays off in science at a greater rate than any other activity. It is impossible to advance the science or even keep up if you must reinvent what has gone before.

i. Understand Its Assumptions

There are foundation assumptions underlying every paradigm. You cannot understand or use the paradigm without them. These issues are the purpose of this manuscript and this course. You cannot be successful at explaining phenomena if you don't understand what the scientific community needs to know and believes is an acceptable explanation. A paradigm is like a partially completed jigsaw puzzle. Each new functional relationship fits within the existing framework and contributes to completing the picture. An individual piece in isolation is meaningless.

ii. Understand How it Works

Paradigms provide the machinery with which events in nature can be understood by its practitioners. For example, if you were trying to discover how to enable someone to become a "better" person, then you will have the largest impact on that person if you see the task from the correct level of molarity (probably behavioral adaptation) and from the correct time scale (probably shortterm or learning). Additionally, you will be most productive if you focus on objective measures (empiricism, multiple converging support, etc.).

The behavior analysis paradigm further argues that following a behavior with a particular consequence can increase the rate of that behavior, and that the deliberate manipulation of that contingency can change the behavior of the organism. Additionally, it argues that there can be no meaning to "the person" other than the person's behavior. It follows therefore that to create a better person, the task is to reinforce "better" behaviors and to extinguish "worse" behaviors.

b. The Paradigm's Ability to Predict Nature

The issue here is the degree to which the paradigm is able to perfectly capture all the variance in nature. As we have seen in several chapters not only are obtained data seen from a paradigmatic perspective and in that sense filtered or approved by the paradigm, but the paradigm is maintained only so long as it provides the most productive framework within which to view data. Ultimately, the paradigm must answer to the reality of the productive interaction with nature.

i. Very Broadest Most Fundamental Aspects

The most basic demands of the scientific paradigm such as requiring consensual validation have changed little in the last several centuries and are unlikely to change much in the future. These aspects have pretty much gotten to be the final word.

ii. Middle Level Traditional Aspects

These aspects of a paradigm are relatively recent developments in the bigger scheme of things but have come to be pretty basic to a paradigm. Given the breadth and depth of their support, they are unlikely to change much in the future. Examples of these types of elements would be the role of micro organisms in infection in the medical paradigm and the role of reinforcement in the field of psychology.

iii. Shallow or Least Established Aspects

These aspects of a paradigm are important elements of the current paradigm and provide the frame of reference for the practice of psychology but are not as stable as more fundamental elements. For example, the central role of matching in behavioral research and theorizing is likely to change.