The Scientific Process

Dominic W. Massaro

The fundamental question is ... how science is possible
because it includes the aberrant (features of human behavior)
as well as the rational.
Ian I. Mitroff (1974)

The endeavor of psychological inquiry is unique among all sciences in one important respect. The scientist hopes to understand phenomena which themselves are critically involved in the process of understanding. Ultimately, the subject matter of the psychologist is the psychologist. M. C. Escher's *Drawing Hands* (Figure 1) sums up our dilemma more exactly than might any verbal statement. In this chapter, we study how people investigate the world around them. By doing so, we gain insights not only into the processes involved in psychological inquiry but also into the appropriate framework for the scientific study of mind and behavior.

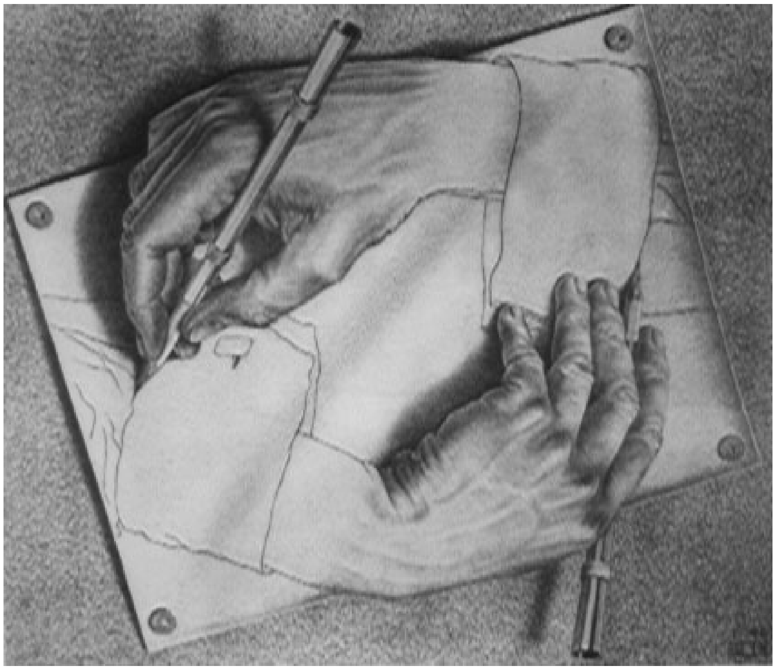

Figure 1. Drawing Hands by M. C. Escher

SCIENTIFIC INVESTIGATION

A rough characterization of scientific investigation includes a number of distinguishing attributes or features. The domain of the investigation centers around some form of systematic or controlled observations. The phenomenon under study is measured or quantified in a relatively precise manner. These observations are used to accumulate evidence, measure relationships, or to test one or more hypotheses. The nature of the scientific process is most easily illustrated in the context of hypothesis testing, that is, determining whether or not an explanation is correct. Any hypothesis must be testable, which means that the observations are capable of disconfirming it. Usually, certain assumptions are necessary in order to apply the hypothesis to the observations under study. Finally, a law of parsimony (simplicity) pervades the domain of scientific investigation. Given two equally good hypotheses, the most parsimonious one is preferred. A more complicated hypothesis includes more assumptions and the greater the number of assumptions, the easier it is to predict or describe some phenomenon.

Hypothesis Generation

Where do hypotheses come from? Given some preliminary observations, a person hits upon some explanation of the phenomenon of interest. Applied to a new situation, this explanation is referred to as a hypothesis. Figure 2 illustrates the ideal sequence of events in scientific endeavor. Needless to say, actual science does not always proceed in such a systematic fashion. A scientist has available a set of observations which consists of the research literature, current theory, and direct experience and hunches of his or her own. These phenomena are organized and interpreted in the framework of an explanatory system. This explanation should imply the outcomes of additional observations. These implications lead to specific hypotheses and tests of the hypotheses. The outcome of the test may falsify a hypothesis, requiring some revision of the explanatory system. If the outcome of the test is positive, the scientist continues to work within the same explanatory system, generating additional hypotheses and tests.

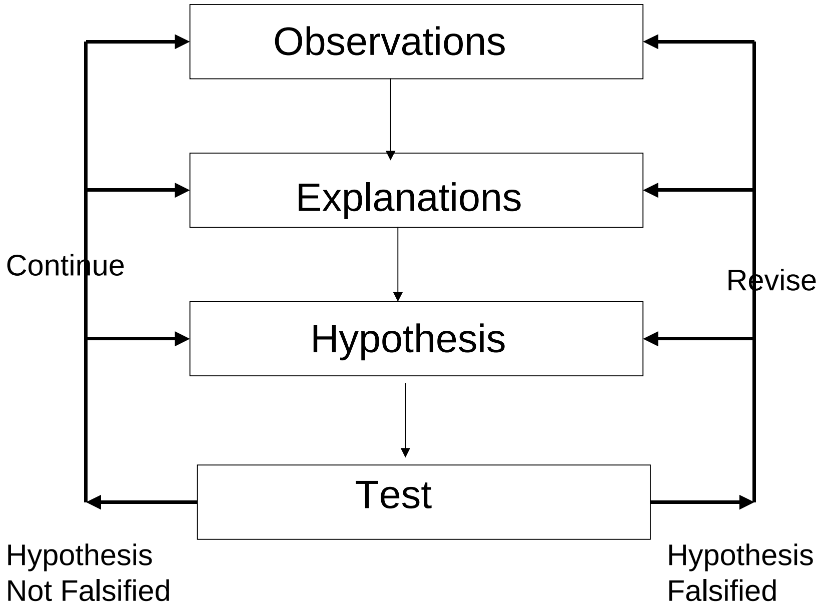

Figure 2. Ideal model of the scientific process.

Scientific Frameworks

Experiments are usually developed and conducted within general scientific frameworks, whether these frameworks are explicitly defined or only implicitly assumed by the experimenter. One framework for scientific endeavor has been expressed most succinctly by Popper (1959). The central assumption is that hypothesis testing must follow deductive rather that inductive methods.

Deductive reasoning is a process in which a conclusion follows logically from given assumptions, hypotheses, or premises. The conclusion contains no more information than that which is contained in the given assumptions. All humans are animals; I am a human; therefore, I am an animal. As can be seen from this example, the conclusion that "I am an animal" was directly deduced from the given information that all humans are animals, and I am a human. No new information is presupposed, for instance, by concluding that humans are a unique type of animal. Everything follows from what is given. On the other hand, inductive reasoning is a logical process in which the conclusion contains more information than that which is contained in the observations and experience on which it is based. Additional information is assumed in the conclusion. Every cat I have ever seen has hated being given a bath; therefore, all cats hate being given baths. From my experience it seems as though all cats do hate being given baths, but since it is impossible to give baths to every cat on planet earth, we are assuming that they all hate it. There is no certainty that tomorrow a dirty cat will not come to my door begging for a bath.

Following Hume, Popper claims that we are not justified in inferring universal or general statements from specific ones. Any conclusion drawn inductively might always turn out to be false. Although we may have many instances of white geese, this does not justify the conclusion that all geese are white. Therefore, the scientist should not try to verify a particular hypothesis, generalizing that it is universally true, by demonstrating that it works in specific instances. Since new instances can always falsify a given statement, no experimental observation can verify a hypothesis.

Hypothesis Testing

Popper proposes that hypotheses once constructed must be subjected to the following analyses. The investigator begins by comparing the conclusions derived from the hypothesis in order to determine whether they are internally consistent (that is, do not contradict one another). An analysis of the conclusions will also indicate whether or not the hypothesis is testable. By contrasting this hypothesis with other hypotheses, the investigator then determines whether the theory is unique and whether it would constitute a scientific advance should it survive experimental tests. Finally, if the conclusions drawn from the hypothesis meet these requirements, then it is worthwhile to subject its conclusions to experimental tests.

Experimental tests will decide how well a hypothesis or theory survives. If a theory survives the experimental tests, we should not discard it. On the other hand, if the experimental tests falsify conclusions drawn from the theory, then the theory should be rejected or modified accordingly. A critical feature of Popper's scientific framework is that verifiability and falsifiability do not have a symmetrical relationship. Although theories can be falsified, they cannot be truly verified. Positive results do not necessarily mean that the theory is true; they simply mean that the theory was not falsified. Popper proposes that it is best to conclude that positive results only corroborate a particular theory; they do not verify it.

Table 1 illustrates the falsification strategy of hypothesis testing. A hypothesis H predicts some observation O. Two outcomes of the experiment are possible. If the predicted observation O is not obtained (-O), the hypothesis is rejected. If the observation is obtained, the hypothesis is not rejected but neither is it verified. The bottom half of Table 1 acknowledges the fact that any experimental test of a hypothesis usually requires auxiliary assumptions relevant to the specific experimental situation. It could be the case that certain observations do not disprove a hypothesis if the assumptions involved in testing the hypothesis are not appropriate.

Prediction Observation	H ⇒ O -O	H ⇒ O O
Conclusion	-H	none
Prediction Observation	H + A ⇒ O -O	H + A ⇒ O -O
Conclusion	-(H + A)	none

Table 1. Illustration of the strategy of falsification in hypothesis testing.

In a more recent contribution, Popper (1976) acknowledged that models could be modified indefinitely to incorporate inconsistent results. This prolongation of falsification is called immunization. Successive modification of a model keeps it alive and holds off its eventual death. In this case, a better contribution is an alternative model rather than another inconsistent experimental result. As observed by Conant (1947, p. 36): "A theory is only overthrown by a better theory, never merely by contradictory facts." Surviving a particular experimental test only temporarily supports a theory since another investigator may soon provide a test that overthrows it.

In a slightly different approach to scientific endeavor, Platt (1964) encourages scientists to employ a strong inference strategy of testing hypotheses. In contrast to generating a single hypothesis, Platt would have the scientist generate multiple hypotheses relevant to a particular phenomenon of interest. The experimental test would be designed to eliminate (or in Popper's words, falsify) as many of these hypotheses as possible. The results of the experimentation would allow the generation of new hypotheses which could be subjected to further tests. Table 2 illustrates the testing of two hypotheses. One hypothesis predicts one observation (O1) and the other predicts (O2) If O1 and -O2 results from the experiment, H2 is rejected. Analogously, if -O1 and O2 results from the experiment, H1 is rejected.

Prediction Observation	H1 ⇒ O1; H2 ⇒ O2 O1 and –O2	H1 ⇒ O1; H2 ⇒ O2 -O1 and O2
Conclusion	-H2	-H1
Prediction Observation	H1 + A ⇒ O1; H2 + A ⇒ O2 O1 and –O2	H1 + A ⇒ O1; H2 + A ⇒ O2 -O1 and O2
Conclusion	-(H2 + A)	-(H1 + A)

Table 2. Illustration of the strategy of strong inference in hypothesis testing.

Both Platt and Popper adhere to Hume's axiom prohibiting inductive arguments. The message is that the scientist should not attempt to confirm a single pet hypothesis. However, Platt's solution seems more productive in that at least one of the multiple hypotheses under test should fail and can, therefore, be rejected. Strong inference has the potential of providing more information than falsification. If an experiment can be designed to falsify one hypothesis with one outcome and another hypothesis with another outcome, then there is a greater likelihood of rejecting a hypothesis. By making H_1 and H_2 mutually exclusive, then the experiment should be able to falsify one of the hypotheses. However, other outcomes might be possible; for example, neither or both of the outcomes may be obtained.

Hypotheses, Models, and Theories

A hypothesis is a conjecture to account for some phenomenon or set of phenomena not yet understood. In this regard, hypotheses, models, and theories are all defined similarly. Models and theories usually consist of a set of hypotheses or allow specific hypotheses to be derived from them. For our purposes, the terms will be used interchangeably, although they usually can be considered to be on a continuum of specific to general accounts of phenomena. A specific instance of a model and its role in scientific study is a computer program of memory search and comparison.

A good theory is 1) verifiable or testable, 2) parsimonious or relatively simple compared to the phenomena being explained, 3) comprehensive or reasonably complete and general, and 4) heuristic or a useful framework for generating predictions, analyzing results, and explaining them. A good theory provides an integrative framework for a set of hypotheses. It allows one to organize knowledge about the world. An important advance occurs when a theory is developed to integrate hypotheses and empirical findings that were previously believed to be unrelated. This integration makes available a larger domain for application of the theory. The theory becomes useful both in directing future research and in facilitating communication among scientists. To the extent that the theory applies to a large domain of phenomena, we can be increasingly confident that the theory will not be easily falsified.

Analogy in Science

Popper's and Platt's framework may be most appropriate for a relatively mature scientific discipline. In an early stage of model development, an empirical scientist might have to be content with a relatively facile acceptance of models or systems from other domains. As an example, Descartes used the machine as an analog of overt behavior of people; the behavior of a person was viewed as following the same principles as those for machines. Models derived from analogs can be tested to some extent in that conclusions drawn from the analog can be tested in the system of interest. Using the pump as an analog for a model of the circulatory system, it is possible to derive certain relationships between pressure and output flow in pumps and test whether this relationship holds in the circulatory system.

A familiar example in cognitive psychology is that of the digital computing device as a model for certain aspects of the human's mental functioning. The computing device can be described in terms of a number of stages beginning with input to the computer, operations on this input with respect to information stored in the memory of the computer, and finally output of the outcomes of the operations. In order to test whether the computer provides a good analog to mental functioning, a more specific analog is necessary to generate testable hypotheses. As an example, one could develop computer programs that carry out specific types of memory search and test if any describe how humans carry out memory search.

In some situations, an analog can be shown to produce the same outcome as a human, but with an entirely different set of processes. If this is the case, the analog is a poor model even though it makes correct predictions. Work in artificial intelligence has encouraged psychologists to adopt working systems in this area as models of psychological processing. For example, computing machines have been programmed to play games such as Chess, Master Mind, and Othello. One problem with this artificial system as a model for psychological functioning is that the computing capacity and speed can compensate for the deficiency of certain intelligent processes by sheer processing power and time. Although an artificial system may be shown to perform some task such as playing chess, the critical question is whether the artificial system performs the task in the same way as humans. This question requires the investigator to compare the nature and time course of the internal processes in humans to those of the artificial system. Only if the internal processes are similar can we say that the artificial system is a good model of the human system.

As students of psychological processes, we are interested in not only whether a model holds up to its experimental tests, but also whether or not practical applications can be derived from the model. Although the model helps us understand human conduct, we also want it to help us modify it. An obvious example is whether a model of the act of reading written language proves useful in assessment of reading skills and development of reading instruction. A model's adequacy in providing assessment and guiding instruction provides additional tests of the model. If a model assumes that a reader must have some knowledge of spelling constraints (orthographic structure) in order to recognize words for rapid reading, then readers without this knowledge should reveal reading problems. Acquiring this knowledge should lead to rapid reading if no other deficits exist.

J. Robert Oppenheimer, one of the best-known physical scientists of this century, rallied for greater scientific literacy. Scientific inquiry bears a strong resemblance to art, music, and sports and he believed that science could engage both the amateur and professional. Sheer joy would be the primary motivator. Our goal is captured by the cartoon in Figure 3; we inquire about inquiry but stop at this first level of regress.

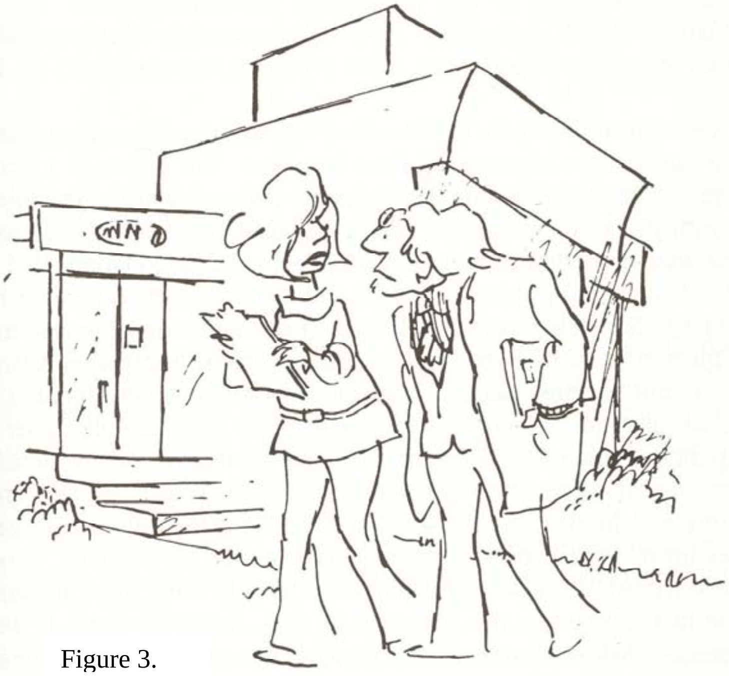

Figure 3.

"GRANTED, WE HAVE TO DO THE RESEARCH. AND WE CAN DO SOME RESEARCH ON THE RESEARCH. BUT I DON'T THINK WE SHOULD GET INVOLVED IN RESEARCH ON RESEARCH ON RESEARCH."

Psychology of Science

Scientific endeavor can be viewed as a natural but formal extension of human perception, memory, thinking, and problem solving. Within this light, we might expect to find many of the properties of these basic psychological processes to be reflected in scientific inquiry. This framework offers a productive exchange between the science of psychology and the psychology of science. We begin by reviewing some important attributes of psychological processes that play an important role in scientific investigation.

Campbell (1977) reviews evidence for the "idols" or "false images" that are characteristic of human thought. These less than optimal characteristics were provided by Francis Bacon, a famous philosopher of the early seventeenth century. Bacon believed that humans tend to suppose "a greater degree of order" than is really present. That is, our interpretation of an environmental situation is too orderly, too simple, ignores details, and essentially contains too few variables (parameters) of importance. There are several examples of this kind of behavior that have been observed in the psychological laboratory. If a hungry pigeon is placed in a box in which a small amount of grain is given randomly every 20 seconds or so, an orderly sequence of responses will develop. This "superstitious" behavior reflects the pigeon's "belief" that the food is contingent on his behavior and is not simply random.

A second proposition of Bacon's is that people tend to interpret new experiences as fresh evidence for strongly-held beliefs. In the more recent literature, this is called a confirmation bias in the evaluation of new evidence. A related proposition is that we tend to give more weight to positive than to negative evidence. These propositions capture the idea that the interpretation of evidence is heavily influenced by the eye and mind of the beholder. Consider an experiment in which subjects were shown a slide at various degrees of focus and were asked to guess the content. Subjects who were shown an out-of-focus slide and made a premature guess had more difficulty in seeing the slide correctly on successive trials as it was brought into focus. The subject required a less fuzzy slide to identify the content correctly when the slides were shown in a sequence of fuzzy to clear than when the slides were shown in isolation and, therefore, without premature guesses (Bruner & Potter, 1964).

One aspect of a confirmation bias is obvious in scientific practice. Scientists (or their research assistants) tend to check their method, procedure, and data analysis when the unexpected outcome obtains but not necessarily when the expected outcome materializes. However, if errors are found when unexpected outcomes are obtained, we might expect as many errors when expected outcomes are obtained.

Given these prescriptive frameworks of scientific endeavor, it is important to evaluate our natural disposition in the scientific process. Can scientists be expected to conform to some philosophers' ideal code of scientific inquiry? As psychologists, we can also ask whether this mode of scientific inquiry is compatible with our nature in addition to being optimal for the generation of knowledge.

PEOPLE AS SCIENTISTS

How do you evaluate your hypotheses of the phenomena surrounding you? Are you completely objective in your analysis of some situation? How do you react to negative evidence which does not support your hypothesis? Experiments have provided some answers to these questions.

Persons attempt to verify rather than falsify their ideas about the nature of their environment. Assume that there is a rule that generates a sequence of three numbers. For example, given the sequence 2 4 6, your task is to guess the rule governing the sequence (Wason, 1960). The rule can be tested by generating a new sequence of numbers, and asking whether the sequence follows the rule. Subjects are asked to test their hypothesized rule until they are fairly confident that it is correct. Given the sequence 2 4 6, most subjects test the rule with sequences that increase by 2, such as 6 8 10, 10 12 14, or 9 11 13. Subjects report that the sequences are being generated to test the rule that the three numbers must increase by two. This simple example illustrates the powerful verification (confirmation) bias found in most people. The subjects show a confirmation bias because they test the hypothesized rule with sequences that are consistent with the rule rather than with sequences that are inconsistent with the rule. A subject who believes that the rule is increasing by two will test it with sequences that increase by two rather than with sequences that do *not* increase by two. We will see this bias in even the most weathered scientists. Subjects generated many of these positive instances for test rather than negative instances such as 10 11 13. Finding that other sequences that increase by two are also consistent with the rule leads most subjects to believe that the correct rule is, in fact, a sequence that increases by two. In this example, however, the rule was any increasing sequence of three numbers. Accordingly, the conclusion that increases by 2 is the correct rule would be premature because other rules are equally consistent with the observations.

Another experimental task, somewhat more involved, further illuminates the hypothesis-testing processes (Johnson-Laird, 1983). Consider a situation involving 4 cards, each with a number on one side and a letter on the other. The cards are showing 4, K, E, and 7. How would you go about testing the rule, "if a card has a vowel on one side, then it has an even number on the other side"? Which cards are absolutely necessary to turn over to test the rule? It costs fifty dollars to turn over each card so it is desirable to turn over only the cards absolutely necessary to test the rule. The cost of research should not be ignored. Jot down your answers and let's proceed with another rule to test. Four new cards have a city on one side and a means of transport on the other. Given the cards showing (Leeds, train, Manchester, car), test the rule, "if I go to Manchester, then I travel by train"? Again, it costs fifty dollars to turn over each card. What cards are necessary to turn over to test the rule?

If you performed as most of the subjects did, you turned E and 4 over to test the first rule and Manchester and car to test the second rule (Wason & Johnson-Laird, 1972; Wason & Shapiro, 1971). However, 4 does not test the first rule since the rule is not bidirectional. An even number can occur without a vowel but not the reverse. Even if a vowel is found on the other side of 4, this result does not prove the rule since negative instances may still be found. The card with 7 can disconfirm the rule since it might have a vowel on the other side. Therefore, E and 7 are the appropriate tests for the first rule as are Manchester and car for the second rule.

These experiments and many others (Johnson-Laird, 1983, Rips & Marcus, 1977) reveal certain weaknesses in human deduction. First, people tend to interpret conditional statements of the form, If P, then Q, as bidirectional. Accordingly, they believe that the statement, If Q, then P, follows from the conditional statement, If P, then Q. In fact, the bidirectional is given by the statement, If and only if P, then Q. About half of the people believe that 4 tests the rule. Second, people fail to use the rule of inference called *modus tollens* in logic. This rule states the following with respect to the proposition, If P, then Q; if Q is false, then P is also false. In our example, *modus tollens* means that if Q does not occur, P cannot occur. Therefore, if the digit is not even, then the letter cannot be a vowel. According to *modus tollens*, 7 can disconfirm the rule and yet only about one person in ten chose 7 to test the rule.

People do better in testing conditional statements in the concrete world of travel than in the abstract world of letters and digits. This result illustrates the important interaction of problem solving and natural experience. The game with vowels and even numbers is extremely abstract whereas traveling is part of our every day experience. Errors of assuming that a statement is bidirectional are decreased when that statement is expressed in concrete, real life experiences rather than in abstract notions. In the former we have past knowledge and experience to help us picture the conditions in our mind. We can picture a situation in which the only means of transportation to a city called Manchester is by train. We know by experience or prior learning that this city of Manchester would not be the only place that a train traveled to. On the other hand, the letters and numbers are simply symbols that are related in an arbitrary way. We have no prior experience or knowledge to help us picture the test situation. In psychological science, an experienced investigator may see alternative tests which have gone unnoticed by an equally logical but less experienced colleague. One is the "better" scientist because he or she has more experience available for taking appropriate action.

The reasoning task studied by Wason and Johnson-Laird sheds some light on confirmation bias in hypothesis testing. A rule or hypothesis mentions some events explicitly and ignores others. As an example, one hypothesis might be that poor readers read poorly because they are easily distracted. In terms of the material implication statement, this hypothesis is: If a person is a poor reader, he will be easily distracted. The person testing a rule concentrates on the events mentioned in the hypothesis. Therefore, the obvious test of the example hypothesis is to evaluate whether poor readers are easily distracted. Another necessary test, however, is to ask whether good readers are easily distracted. Therefore, a general rule of thumb is to go beyond the events in the hypothesis to test its validity.

Confirmation Bias

Mynatt, Doherty, and Tweney (1977) initiated a systematic series of studies of the psychology of scientific inquiry. These researchers created an artificial research environment and allowed college subjects to make preliminary alterations in the environment. After observing the environment, the participants were instructed to formulate a hypothesis to account for the behavior of objects in the environment. The subjects were then given the opportunity to test their hypotheses. One group of subjects was instructed that the job of the scientist was to confirm hypotheses. Another group was told that the job was to disconfirm hypotheses. A third group was given neutral instructions in terms of simply testing hypotheses. Subjects were shown hypothetical environments that would allow tests of their hypotheses. Given a pair of environments, a subject was asked to choose one member of the pair to test the hypothesis. By evaluating these choices against the subject's original hypothesis, the authors determined whether the subjects chose situations to confirm or to disconfirm their hypotheses.

Subjects chose a situation to confirm their hypothesis about 7 times out of 10 regardless of how they were instructed. These results and other evidence adduced by the authors revealed a strong confirmation bias in this simulated research inquiry. People chose hypothetical situations to confirm their hypothesis and avoided consideration and testing of alternative hypotheses. A second result revealed that subjects, when faced with disconfirming evidence, changed to a new hypothesis. Negative results were usually effective in changing the participant's opinion of the truth of the hypothesis. Although subjects may not seek disconfirmatory evidence, they use it correctly when it is available.

In a second study by the same authors (Mynatt, Doherty, & Tweney, 1978), subjects were faced with a much more complex research environment. Subjects attempted to determine how 27 fixed objects influenced the direction of a moving particle in a two-dimensional display. The objects differed in size, shape, and brightness and their influence on the particle varied as a function of these dimensions. A confirmation bias was again observed and could not be modified with highly explicit training in strong inference. In contrast to the previous study, however, subjects often kept or returned to their disconfirmed hypotheses. The more complex environment seemed to limit the generation of new hypotheses and, therefore, subjects kept alive those few ideas that they had. This is reminiscent of the compulsive gambler's dilemma, "I know the game is crooked, but it is the only one in town." A final result indicated that complete abandonment of a disconfirmed hypothesis also proved unproductive. For example, a general hypothesis of how the objects influence particle direction was abandoned when it failed experimental test. The principle itself was correct but only for some of the objects. Accordingly, it can be worthwhile to modify disconfirmed hypotheses based on the previous results rather than rejecting all aspects of the disconfirmed hypothesis.

Continuing their research, Doherty, Mynatt, Tweney, and Schiavo (1979) had subjects decide from which of two islands an archaeological find had come. Subjects usually asked for information relevant only to their preferred island rather than for information relevant to both islands. The requested information was useless since knowing that a characteristic of the found object is representative of one island does not provide information about whether the characteristic is also representative of the other island. This confirmation bias led to an inappropriate confidence in the subject's hypothesis. If the subject believes that the pot with a curved handle came from one island and then finds out that 80% of the pots from that island have curved handles, he or she becomes even more convinced that the pot is from that island. What the subject fails to realize is that pots from the other island may be as likely or even more likely to have curved handles.

The behavior of college students in a simulated research environment might not be typical of professional scientists. However, observational and historical accounts of scientific endeavor argue just the opposite. Mitroff (1974) carried out an intensive study of 40 scientists participating in the Apollo lunar missions. These scientists had devoted much of their life to studying the nature of the universe and the missions offered a unique observation of that nature. The observation centered around the moon rocks that the astronauts carried back to earth. The majority of these scientists revealed a strong commitment to their scientific beliefs; in addition, these beliefs could be considered to be outright biases since they did not have sufficient supporting evidence. These scientists were aware of their biases and commitments and saw them as perfectly natural in the pursuit of knowledge. These commitments and biases were not completely disfunctional; scientists learn to cope with disconfirming data and usually modify their ideas accordingly. What is obvious is that the storybook picture of the completely objective, unbiased, and disinterested scientist requires modification.

Confirmation bias also reveals itself in terms of a reluctance to accept findings that do not fit into the current mold of thinking. Spallanzani's discovery of bat navigation met with this fate. Being a good scientist and possibly hoping that publications count as much as prayers for tenure in heaven, he rushed off his findings to the scientific community and described how bats navigate with their ears. He didn't know how but bats seem to get around not in terms of seeing but by way of hearing. As you might expect, he was ridiculed by the scientific community. At that time there was just no way that one could understand how bats could navigate with their ears. If bats see with their ears, does that mean they hear with their eyes? Not on your life: eyes are for seeing and ears are for hearing. The research done by Spallanzani was correct but it took at least 150 years for the scientific community to accept it. Given his difficulties, scientists, faced with a rejected paper, can take heart.

Competition in Science

Most of the theoretically driven research aims at supporting rather than falsifying theories. A scientist tends to view the world in terms of his or her theory; he or she may be the least equipped person to invent a critical test of the theory. Empirical tests usually turn out to provide results consistent with the theory without really providing a critical test of the theory. Given a natural predisposition for the scientist to verify rather than falsify his or her theory, it is important that scientific endeavor remain competitive. The competitive dimension is probably as fundamental to the game of scientific endeavor as it is in most games (McCain & Segal, 1973). Scientists are much more willing to falsify other theories while seeking support for their own. If one accepts the ideal of falsifying models, the scientist is usually making most progress when he or she is testing someone else's theory rather than his or her own. A model is a good model to the extent that its conclusions are easily tested by other scientists, just as a good scientific test is one that can be repeated by other scientists. Experimental tests should allow not only falsification of the model but should provide clues to building new models if falsification occurs. Of course, the new models must also be easily testable.

SCIENTISTS ARE PEOPLE

The defining characteristics of scientists are not any more obvious than those of other fuzzy concepts such as ethics, games, and happiness. At some level, scientists enjoy various aspects of the process of psychological inquiry and this reason alone may sustain a scientist's career. When challenged, scientists may also attempt to justify their life's work with more socially acceptable goals; however, our cognitive explanations of our behavior usually fall short of the actual motivational precursors (Nisbett & Wilson, 1977). As might be expected, scientists differ in their talents and styles of scientific inquiry (Mitroff & Kilmann, 1978; Wachtel, 1980). One person might be particularly adept at creative theorizing, but has difficulty mapping these ideas into an experimental framework. Another might be a concise and thorough experimentalist, but cannot advance beyond the current theoretical framework. Some scientists do both with similar competence. On the enjoyment side, some scientists delight in the stage of formulating and testing ideas, but dread the process of *writing up* their research. Others love writing, but find the mechanics of equipment design, testing subjects, and data analysis relatively boring.

Most young scientists are required to carry out all stages of psychological inquiry. As an example, few graduate programs in psychology permit a purely theoretical study to qualify as a dissertation and young faculty qualifying for tenure are expected to do both empirical and theoretical work. Wachtel (1980) has illuminated some of the limitations of this model of psychological research and suggests that the field accept and support a broader range of research styles. An alternative solution would involve teams of research scientists whose interests and skills are complementary rather than requiring the individual scientist to perform all stages of research enterprise. Some collaborations of this sort have been reasonably profitable and current scientific endeavor seems to be moving more in this direction. Programs in cognitive science have brought together scientists from a wide range of disciplines to collaborate on issues in cognitive science. Linguists, anthropologists, computer scientists, philosophers, and psychologists have uniquely different skills and styles of inquiry and their collaboration on an important problem such as reading offers a promising research strategy (Gardner, 1985).

Scientists learn and adopt not only intellectual frameworks for research, but also psychological and sociological frameworks. All scientists might be expected to be competitive, ambitious, narrow-minded, idealistic, practical, and emotional, but in fact they differ widely on these psychological characteristics. On the sociological dimension, scientists can usually be classified into one of two types. Some scientists are committed to their favorite hypotheses and take strong stands on most issues. Other scientists avoid this kind of polarization and see, at least, two sides to every issue.

It is apparent that the actual practice of the scientist does not follow an ideal set of procedures dictated by philosophers and logicians. Scientific inquiry is not performed by objective and unbiased humans but involves the participation of involved and committed observers. It is important to acknowledge that there are no unbiased facts; we are simply not equipped to interpret data passively. As Churchman (1971) observed, theories are not tested by observations arrived at independently of theory. Facts and interpretations are inseparable and both are human products. Although science is subjective, it is not completely arbitrary. The scientific community must assess the emotion and bias in the same way that it assesses the facts and theory.

One of the goals of this chapter is to describe the appropriate methods for psychological inquiry. The irony is that some of this kind of knowledge might be tacit knowledge (Weimer, 1979). Polanyi (1966) refers to tacit knowledge as information not directly available to conscious awareness and not capable of being directly communicated. In this regard, scientific methodology can be considered to be analogous to other skills such as riding a bicycle or playing good tennis. Instructors insist that a comprehensive description of these skills cannot be given explicitly. The coach teaches some things by example and the student must learn by observing and doing. In this book, we also ask the participant to learn from examples of scientific practice. Although there is no concrete set of rules for doing science, there are examples of good, mediocre, and poor research enterprises. The perceptive student will learn the rules of the science game by observing, participating, and practicing.

Intuitive Scientists

We might be led to believe that psychologists are a unique group among humans whose aim is to understand and predict psychological behavior. According to one view, however, all people are intuitive scientists and the scientific process can serve as a model for human action (Kelley, 1967; Weimer, 1979). People attempt to make sense out of the world around them and test their ideas with every new encounter. As an example, a person believes that a new acquaintance is insincere and future encounters with this person are used to confirm or disconfirm this belief. If this view approximates some of our behavior, we can see certain interpersonal conflicts materializing if intuitive scientists also have a confirmation bias. According to Kelley, proper resolution of a person's predictions takes on greater psychological significance than the material rewards usually claimed to motivate behavior.

If all people are intuitive scientists, psychological researchers might begin to explore the possibility of a more collaborative relationship with their subjects. Even if a collaboration is not possible, an experimenter might ask subjects to predict and explain the outcome of their study. The experimenter should be able to assess how the subject may have influenced the results in seeking to confirm certain beliefs about the experimental situation. As noted elsewhere in this book, people as the subject matter of scientific endeavor offer unique challenges to the scientist. Progress might depend on the degree to which intuitive scientists can be outwitted to reveal their mental functioning.

Potential Pitfalls in Research

The human element in the scientific process can distort the procedures, methods, results, and interpretations of experimentation. Scientists have not been unaware of this fact, and a good number of investigators have studied the problem intensively (Barber, 1976; Rosenthal, 1966, 1976). Rosenthal and Fode (1963) evaluated the influence of experimenter expectancy on the outcome of experimental tests. Here it is worthwhile to distinguish between the investigator and the experimenter. The investigator initiates the research, and the experimenter does the actual testing of the subjects. In many cases, the investigator serves as experimenter and sometimes even as subject. The investigators, Rosenthal and Fode, had 30 student experimenters test hundreds of subjects in a photo rating task. Subjects were asked to rate the people on a scale from -10 (extreme failure) through 0 (neutral) to +10 (extreme success). The photos were human faces cut from newsmagazines. The faces were selected to be relatively neutral in terms of whether or not they would be rated as being successful. There were two groups of experimenters. The investigators told one group of experimenters that they would obtain positive ratings of about +5, whereas the other group was told that they would obtain negative ratings of around -5.

The instructions to the subjects were identical for both groups of experimenters, and the same faces were shown. The results indicated higher ratings for subjects tested by experimenters who were told to expect higher ratings. There were many potential reasons why this result was obtained, and many later experiments were directed at assessing what accounted for the findings. Cheating by the experimenters and feedback from the experimenters did not seem to be responsible. Rosenthal (1976) concluded that nonverbal communication is responsible, although he has not delineated the nature of this communication. Careful observation of the testing procedure was unproductive in discovering the communicator cues that are used if, in fact, nonverbal communication is the culprit.

Rosenthal's research has been criticized for a wide variety of reasons (Barber, 1976). Much of the criticism centers around the use of inappropriate statistical methods for testing the significance of the results. We all know that it is easy to be misled with statistics, and Rosenthal seems to be guilty of this behavior. In many cases, the standard comparison was not significant, and either eliminating subjects from the analysis or performing a large number of additional statistical tests was needed to achieve statistical significance. In our framework, a between group difference is not psychologically significant if it doesn't reveal anything about the psychological processes responsible for the difference. There are probably many subtle ways the experimenter might cue the subject to adopt a slightly more positive criterion for rating the faces; however, the questions addressed in most research require a much more complex pattern of results to be of interest. The experimenter in these situations would have a difficult time communicating the exact pattern that is expected or desired. Also, the face-rating task is an ambiguous one since there is no well-established method for rating success or failure of people on the basis of pictures of their faces. It is not surprising that subjects in this situation will look to the experimenter for whatever hint they can get for performing the task. When considering these factors, one wonders why it isn't even easier to show experimenter-expectancy effects. For our purposes, it does not hurt to insure that the results of a particular study cannot be influenced by expectancies of the experimenter.

The use of computers can allow complete control over experimenter-expectancy influences. In this case, the experimenter can be blind as to the experimental condition given to the subject. The computer can be programmed to choose the specific condition for each subject and then present the experimental test without any human intervention. Here we might say that the subject was untouched by human hands. The human experimenter would welcome the subject, introduce her to the laboratory and computer, and then put her in the hands of the program. The program would choose the conditions randomly and test the subject. The human experimenter would return at the completion of the experiment and retrieve the results from the computer. Both the investigator and the experimenter will feel more comfortable with this procedure. In addition, the scientific community reading the results would not have to worry about experimenter-expectancy effects.

Dimensions of Science

There are many psychological and sociological dimensions to the scientific enterprise (Platt, 1970). Humans seek more than the simple fulfillment of needs as a naive behavioristic or sociobiological view might indicate. Harry Harlow (Harlow , 1953), an early critic of behaviorism, demonstrated that monkeys would learn and work to see novel situations, such as the picture of a colleague. Many of the early learning studies of rats in mazes were plagued with the rats' curiosity seeking behavior. Rather than turning to the arm of the maze that had a food reward earlier, the rat would venture down another arm. Later testing revealed that the rat "knew" were the food was. One of the biggest problems in human infant research is keeping them interested in the experimental task; they are easily bored and must be continuously challenged with new problems. Science is novelty seeking at its most challenging; it is exciting and provides a range of rewards for a wide number of participants. The rewards range from the most basic, such as earning a living, to the most abstract, such as a fulfilling and competent interaction with some aspect of nature or obtaining admiration from respected colleagues.

Style in Science

The style of the successful scientist is as varied as human nature itself. However, a critical ingredient is the willingness to participate in intellectual work. Platt suggests that struggling scientists structure their formal thinking by putting their reasoning into bound notebooks. The notebook serves as a diary and provides the needed continuity for any kind of intellectual inquiry. Our problem-solving skills are critically dependent on the information available. Most memories are not capable of providing the information when needed. The notebook holds the scientist's representation of the problem and frees the processing capacity to question and modify the representation. Knowledge is a natural extension of the mind in the same way that a tool is an extension of the hand.

Moral Responsibility

As in all human endeavors, scientific practice entails moral responsibility. Science and technology have placed our survival and destiny in our own hands. Scientific knowledge is not good or bad, but represents the closest approximation to objective truth. To the extent that scientific knowledge reduces uncertainty about many of the important issues facing society today, much of the debate will stop and we can get on with the business of making decisions and implementing those decisions. Scientists have little control over the nature of phenomena but all humans have the power to decide and act. Science only provides knowledge of the operations and consequences of these actions.

Ethical principles for research with human participants have been published (*American Psychologist*, 1981). The investigator assumes the responsibility for the ethical acceptability of the research. Before the study, the investigator describes to the participants the important aspects of the research that might influence their willingness to respond. Each participant's are respected and subjects are free to withdraw from the study at any time. After participating, subjects are informed about the nature of the study and any misconceptions that they might have had are addressed.