

***T. rex*, the Crater of Doom, and Scientific Method**

ANTON E. LAWSON
Department of Biology
Arizona State University
Tempe, AZ 85287 -1501
anton.lawson@asu.edu

March 2000

This material is based upon research partially supported by the National Science Foundation under grant No. DUE 9453610. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

***T. rex*, the Crater of Doom, and Scientific Method**

Abstract: Working from the 1970s through early 1990s, Walter Alvarez and his research team discovered the cause of the mass extinction that claimed the dinosaurs 65 million years ago, a massive meteor that slammed into the Yucatan Peninsula. The present paper discusses that research in terms of eight major arguments that were crucial to the ultimate discovery. The Alvarez case history reveals that science is at its core an enterprise in which causal questions are raised and answered through the creative use of analogical reasoning followed by an equally creative process of hypothesis testing in which predicted and observed results are compared. According to this hypothetico-predictive theory of scientific method, causal hypotheses drive investigations and enumerative induction plays no role. The central educational implication is that science classrooms should more frequently engage students in open inquiries that raise causal questions and encourage them to brainstorm alternative hypotheses, which are then rigorously and explicitly tested in a hypothetico-predictive fashion.

An unfinished task facing science educators interested in helping students become scientifically literate is the development of a valid and accepted theory of the nature of science (e.g., Alters, 1997). Usually the phrase "nature of science" is understood to include both scientific method, or methods, and the epistemological status of its products (e.g., its concepts, laws and theories). Both aspects are complex, so the present paper will restrict itself to a discussion of method. More specifically, the paper will introduce and discuss a case study of the research of Walter Alvarez from the 1970s through the early 1990s. Alvarez played the lead role in the discovery of a huge meteor that crashed into the Yucatan Peninsula and caused the mass extinction claiming the dinosaurs some 65 million years ago. Alvarez chronicles his research in book entitled *T. rex and the Crater of Doom* (Alvarez, 1997). Fortunately, Alvarez's account is very detailed as it chronologically lays out the key steps in his research and thinking. Thus the book provides an extraordinary opportunity to examine scientific method during an important scientific discovery. Let's first turn to Alvarez's account of that discovery. We will then analyze the account to derive a theory of scientific method. Lastly, we will compare the derived method to other proposed methods.

The Research of Walter Alvarez: Why Dinosaurs Became Extinct

Initial Explorations

Prior to embarking on his research career in the early 1970s, Alvarez was an oil company geologist working in Libya and Italy. During that time, he became interested in the geological history of the Mediterranean region, more specifically that of the Apennines. Based in part on the plate tectonics revolution of the 1960s, Alvarez and colleague Bill Lowrie speculated that a rotation of Italy's continental crust caused the observed deformation of the Apennines. Alvarez and Lowrie knew that past movements of the Earth's crust can be reconstructed using rocks containing magnetic mineral grains. The grains hold a record of the direction of the Earth's magnetic field when the rocks formed. So Alvarez and Lowrie set out to collect rocks hoping to use these tiny "fossil compasses" to piece together the region's geologic history. As Alvarez (1997) put it:

We were hoping to see a progressive rotation during the time of deposition of the Scaglia, with older fossil compasses twisted farther away from north than the younger ones. That would be the paleomagnetic signature of a rotation of the Italian plate. (ps. 35-36)

As it turned out, Alvarez and Lowrie found that the fossil compasses did in fact point to the west of north, indicating that the plate had indeed rotated. Disappointingly however, a detailed history of the rotation could not be worked out because local bed disruptions made it impossible to separate local movements from plate rotation.

But their disappointment was short lived. While investigating limestones collected near the Italian town of Gubbio, Alvarez and Lowrie quite unexpectedly discovered that, unlike most of the north pointing compasses, a few pointed in exactly the opposite direction! This was not the first time such magnetic "reversals" had been

discovered. In fact the reversals had played a key role in convincing geologists that the Earth's magnetic field had switched directions several times in the past.

Alvarez and Lowrie soon realized that the reversals held the key to dating past reversals because they occurred in sedimentary beds rich in fossil foraminifera, microscopic sea creatures known to provide the best means of dating marine sedimentary rock. Consequently, during the mid-1970s they made several trips to Italy working their way down through the Cretaceous rocks and up through the Tertiary charting and dating magnetic reversals with forams. Eventually, their investigation led to another surprising observation. Again in Alvarez's words:

With a hand lens you could spot the near extinction of the forams, which were abundant and as big as sand grains in the top beds of the Cretaceous, but with only the very smallest ones surviving into the first beds of the Tertiary. Bill and I learned to identify the KT boundary ourselves, and as we located this key break in outcrop after outcrop across the Apennines, we began to wonder about its significance. Why had the forams almost become extinct? What had happened to cause that extinction? And why was it so abrupt? (p. 40) [Note: Cretaceous is derived from the German word "Kreide" (chalk), hence the origin of the phrase KT boundary.]

Clues to Answering a "World-Class" Scientific Question

Alvarez and Lowrie made another puzzling observation. At each outcrop they noticed a fossil-less clay layer about a centimeter thick just above the highest limestone bed of Cretaceous forams and just below the lowest bed of Tertiary limestone. This observation suggested to them that perhaps the clay had something to do with the extinction. Soon, thanks in part to information provided in a talk by geologist Al Fischer, it became apparent that the extinction of the marine forams marking the KT boundary occurred at approximately the same time as the famous extinction of the dinosaurs. Once again in Alvarez's words:

I remember very clearly walking around the grounds at Lamont one day shortly after AL Fischer's talk there and realizing fully that this was a world-class scientific problem. ...The question of the KT extinction looked like one that could lead in totally new directions, and by the time I finished my walk, I had decided I would try to solve it. (p. 42)

The apparent rapid extinction of forams at the KT boundary contradicted the long standing doctrine of uniformitarianism, the view that geologic and biologic changes occur gradually. Nevertheless, Alvarez began seeking a catastrophic cause. Alvarez was well aware of meteor impact craters on Earth, the best example being Meteor Crater in Northern Arizona, as well as the mounting evidence gathered by both manned and unmanned probes of the solar system that the craters covering our moon and other planets occurred because of the impact of asteroids and comets. And more importantly, such impact craters are the rule, not the exception, in the solar system. Alvarez was also aware of two papers (Terry & Tucker, 1968; Russell & Tucker, 1971) in which the authors proposed that the dinosaur extinction had been caused by radiation triggered by the explosion of a nearby star, a supernova.

So by 1976, Alvarez began to focus his attention on the KT boundary, more specifically on the narrow layer of fossil-less clay. Alvarez he did not know how he could use the layer to test the catastrophic supernova explanation or the equally catastrophic impact explanation. Nevertheless, he suspected that the layer held the key to the dinosaur extinction and that it could be used to test the more global uniformitarian versus catastrophic views. As he put it: "Very rapid deposition of the clay would suggest a sudden cause for the extinction, but slow deposition would suggest a gradual mechanism." (p. 61)

But how could he find out how long it had taken to deposit the clay? What he needed was something that had been deposited in the limestone and in the clay at a constant rate. At this point Alvarez enlisted the expertise of his father Luis Alvarez, a physicist at Berkeley. The elder Alvarez knew that although meteors hit the Earth rarely and at random, meteorite dust, which contains iridium, falls from outer space at a constant rate across the entire Earth. Therefore, they came up with a way to indirectly measure the clay's deposition rate by measuring the amount of iridium in the layers.

Measuring the amount of iridium in sediments is no easy task. Nevertheless, thanks to the considerable analytical skills of Berkeley chemist Frank Asaro, by June of 1978 the initial measurements had been made, and they contained another surprise. Instead of the expected amount of about 0.1 parts-per-billion (ppb), assuming the

clay layer had been deposited slowly, a value of 9 ppb had been detected. Consider Alvarez's reaction:

Where had all the iridium come from? Possibilities quickly sprang to mind: Could it have come from the supernova that Dale Russell and Wallace Tucker had suggested to explain the dinosaur extinction? Did it come from an impacting asteroid or comet? Or could there be a noncatastrophic explanation? Maybe the iridium was deposited from seawater somehow. Or maybe the Earth had encountered a cloud of interstellar dust and gas. (p. 69)

Testing the Possibilities

Before investing time and energy in testing these possibilities, Alvarez needed to find out whether the iridium anomaly was restricted to the clay bed around Gubbio or whether it was a world wide phenomenon. So he went to the library in search of other known KT sites. At that time, the only other site was a seaside cliff called Stevns Klint in Denmark. As it turned out, the Stevns Klint deposits also contained a narrow clay layer with the unusually high concentration of iridium. So Alvarez concluded that it was time to think about a global explanation for the anomaly.

Of course the supernova explanation had already been advanced as a possibility. So Alvarez and his team sought a way to test it. Luis Alvarez knew that a supernova would deposit plutonium-244. So they could test the supernova explanation by analyzing the KT clay for plutonium-244. Analyzing geological deposits for plutonium-244 is extremely difficult. But, Frank Asaro and colleague Helen Michel, a skilled plutonium chemist, worked nonstop all day and night to perform the analysis. And when the results were in, Asaro and Michel had found plutonium-244. Alvarez's reacted as one would expect: "Dad and I were nearly jumping up and down with excitement - a supernova had killed the dinosaurs!" (p. 74) Unfortunately, their excitement was short lived as an attempt by Asaro and Michel to replicate the finding turned up empty. The initial finding of plutonium-244 had almost certainly been due to an impurity. Therefore, the supernova explanation was out. So Alvarez and his team began to take the impact explanation seriously.

But how could a meteor impact lead to mass extinction? It took several months of generating and rejecting possible mechanisms until Luis Alvarez recalled reading about the 1883 explosion of an Indonesian volcano named Krakatoa. The Krakatoa explosion had blown so much dust and ash into the air that it was detectable for months in places as far away as London. Could a large meteor impact blow enough ash and dust into the air to block the sun around the planet, kill plants, kill animals up the food chain, and lead to mass extinction? Based on his calculations of the amount ash and dust, and the amount of darkness, it seemed entirely possible. Again in the words of the younger Alvarez: "It was the first good hypothesis for why a large impact would cause a global mass extinction." (p. 77)

During a Copenhagen meeting in September 1979, Alvarez met Dutch geologist Jan Smit. It turned out that Smit and Jan Hertogen had independently confirmed a high iridium level in a KT boundary section in Spain. Soon several publications appeared documenting the iridium anomaly at several KT boundaries around the world. Given a plausible mechanism and this growing evidence for a world wide rapid rise in iridium, the scientific community took considerable notice. During the 1980s, over 2,000 publications appeared. For those favoring the impact explanation, the search for the impact site took center stage. Alternatively, those favoring the also popular volcanic eruption explanation searched for ancient volcanos. Alvarez summarized the situation this way:

We in the impact camp had evidence, from the anomalous iridium, spherules, and shocked quartz in the KT boundary layer, for the impact of a comet or an asteroid, but we could not locate the giant crater that would have resulted from the impact. The supporters of volcanism had no strong evidence in the boundary clay to support a giant eruption at KT time, but they could point to a huge volcanic outpouring of roughly the right age in India, called the Deccan Traps. (p. 98)

Searching for The Crater of Doom

Although many scientists still doubted the impact explanation, Alvarez and others continued the search for supporting evidence. Clearly if a huge impact had occurred in or near the ocean, it would have created a tidal wave, a tsunami. By the late 1980s, evidence of a tsunami had been sought in the now over 100 known KT sites with high iridium levels. Finally a KT boundary at the Brazos River in Texas was found that contained sandy

deposits providing evidence of a huge tsunami. But where had the tsunami come from? In 1988, Alan Hildebrand, a graduate student studying the Brazos tsunami deposits, reasoned that the impact site, the crater of doom, could not have been too far from Texas because the land mass around the Gulf of Mexico would have blocked a more distant tsunami. So he focused his search on the Gulf and the Caribbean. Eventually that search was rewarded by the discovery of a previously ignored, but enormous, circular pattern of gravity anomalies in the Yucatan Peninsula. As it turned out, the circular structure lay buried below the surface. Importantly, core samples previously extracted by Mexican petroleum geologists were found to contain impact-melt rocks consistent with the impact explanation. In fact, back in 1981 the geologists concluded that they had discovered a huge impact crater. But their discovery had not been published in the scientific literature.

Determination of the crater's age became of critical concern following publication of a 1991 paper entitled "Chicxulub crater: a possible Cretaceous/Tertiary boundary impact crater on the Yucatan Peninsula, Mexico." Unfortunately, the core samples from the crater, which could have been dated, had been misplaced. And because the crater lay deeply buried and out of reach, new samples could not be easily obtained. Consequently, Alvarez and his team decided to do the next best thing, which was to search for the closest place to the crater where KT sediments might lie on the surface. This turned out to be in northeastern Mexico. Alvarez summarized the situation like this: "If the KT boundary here, only a few hundred kilometers from Chicxulub, was quiet and undisturbed, Chicxulub could not be the KT impact site, and we would be back to zero." (p. 116)

So Alvarez and his team searched and searched in northeastern Mexico. For days they found nothing. But as luck would have it, on the final day of the expedition, a dry river bed called Arroyo el Mimbral was found with just what they were looking for, namely tsunami sediments nearly three meters in thickness precisely at the KT boundary! Other good news would soon follow. In late 1991, Mexican geologists turned up some stored crates of the original core samples from the Chicxulub crater. When the samples were analyzed, they not only contained melt-rock, but radiometric age determinations showed that the melt-rock was of KT age.

At this point, Alvarez and his colleagues thought that these pieces of evidence would be convincing and that their search was over. But they were wrong as a major challenge came from a Princeton geologist who disputed nearly every interpretation of their evidence from the Arroyo el Mimbral sediments. This challenge necessitated another expedition to Mexico to look for other KT boundary outcrops and still more evidence. That evidence was essentially in the form of rocks containing tiny gas bubbles that Alvarez and his team suspected were melt droplets that had been ejected from the Chicxulub crater. To help find additional KT outcrops, which hopefully would contain the gas-bubble-containing rocks, Alvarez enlisted the help of a team of Mexican geologists familiar with the region. The Mexican geologists had seen the bubbles before in many locations, but did not know how they had formed. But now with an idea of what to look for, and where to look, Alvarez and his team quickly located nine new KT outcrops, all containing evidence of a huge tsunami and all containing rocks with the tiny gas bubbles. Just as Alvarez had suspected, the gas bubbles had surely been created by a massive impact that created the crater, created a huge tsunami, and doomed the dinosaurs. As Alvarez put it:

...I had been convinced since 1980 that an impact had killed the dinosaurs. For more than ten years the evidence had looked better and better for a KT impact, but there were always serious questions and nagging doubts... Now there was a convincing crater at Chicxulub, and melt-rock samples from the crater, and glass from the KT boundary at Mimbral and Beloc and Leg77. (ps. 128-129)

What was the General Pattern of Alvarez's Research?

In general terms, what does the preceding account tell us about scientific method? Although multifaceted, Table 1 summarizes Alvarez's research as eight interconnected cycles of argumentation in which explanations are proposed and tested. Each test involves a plan, which together with the proposed explanation, leads to a predicted (i.e., expected) result. The plan is then carried out yielding an observed result, which is then compared to the predicted/expected result. When observed and predicted results are similar, the argument takes on an *If/and/then/And/Therefore* pattern. When observed and predicted results are dissimilar, the argument takes on an *If/and/then/But/Therefore* pattern. As you can see, each argument is preceded by an observation, a causal question, and a proposed explanation. Further, each argument embodies the proposed explanation, the planned test, predicted results, observed results, and a conclusion. Note that an argument may also include a theoretical rationale used to link the proposed explanation and planned test to the predicted result. In other words, the purpose of the

theoretical rationale is to convince others that the predicted result does in fact follow.

Insert Table 1 about here.

Did Alvarez Use Induction?

Induction is often cited as a key component of scientific method. For example, in writing about how scientists learn, Kuhn, Amsel & O'Loughlin (1988) stated: "Clearly some process of induction is involved. Associations, patterns, regularities are observed, and on this basis expectations or concepts regarding the way the world is organized are formed. Whether or not deductive logic plays a role in scientific thinking, inductive reasoning is clearly central to what scientists do." (p. 21) In spite of this endorsement, induction as central component of scientific method has been severely criticized by some scientists and philosophers of science (e.g., Darwin, as quoted in Schick & Vaughn, 1995; Hempel, 1966; Malherbe, 1996; Medawar, 1969; Musgrave, 1999; Popper, 1959; Popper, 1965). The primary criticism stems from the overwhelming number of possible observations one could make and induction's inability to guide the search for potentially useful observations. In other words, without knowing what to look for, one does not know where to look. This problem seems even more stark when one considers that most important scientific discoveries, including the present one, involve entities and processes that are not observable! Curd (1980) stated the problem like this: "Without a prior hypothesis to guide our investigation we have no idea which facts are relevant to our inquiry and in any case the sets of such facts could not be exhaustive." (p. 207) And Popper (1965) even denied the very existence of induction: "Induction, i.e., inference based on many observations, is a myth. It is neither a psychological fact, nor a fact of ordinary life, nor of scientific procedure." (p. 53) In Popper's view, science proceeds with conjectures, often after a single observation.

In spite of the criticism, one form of induction, called enumerative induction, retains some support among philosophers of science (e.g., Burks, 1980; Gustason, 1994; Jung, 1996). The classic example of enumerative induction involves the initial observation of one white swan. This initial observation is then followed by the observation of a second white swan, then a third white swan, and so on. Eventually after many such observations, at least more than one, the observer presumably uses enumerative induction to draw the *conclusion* that "All swans are white." But there is a problem here, a problem even recognized by the acknowledged father of inductive science himself, Sir Francis Bacon. The problem is that the conclusion is naive. After all, logical falsification merely requires finding a single non-white swan. Thus, a less naive use of enumerative induction might be its use as the source of the generalizing *hypothesis* that "All swans are white." This generalizing hypothesis would then be tested deductively (i.e., If...all swans are in fact white, and...I check out some more swans, then...they should also be white).

Does Alvarez's research provide some evidence for use of enumerative induction in this way? By the late 1980s over 100 KT sites had been found with high iridium levels. Consequently, did Alvarez, or someone else, use these data to generate the generalizing hypothesis that "All KT sites have high iridium levels?" I think not, for two reasons. First, as revealed by Argument Three, the observation of only a *single* KT site with high a iridium level provoked Alvarez to generate a hypothesis, not two or more observations as required by enumerative induction. And second, observation of a high iridium level in a single KT layer did not lead Alvarez to generate a *generalizing* hypothesis, as required by enumerative induction. Instead, the observation prompted him to first raise a causal question (i.e., Why is there a high iridium level in this KT layer?) and then immediately generate a *causal* hypothesis (i.e., The high level might have been caused by a world-wide catastrophic event). With this causal hypothesis in mind, Alvarez then sought a way to test it (i.e., If...a world-wide catastrophic event had taken place, and...the only other known KT layer is observed, then...it should also contain a high level of iridium). Thus, as summarized by Arguments Two and Three, the causal hypothesis was generated after finding a high iridium level in a single bed and was tested by looking for a high iridium level in the only other known KT bed at the time. Clearly in this case, Popper's view seems more accurate than the inductivist view. In other words, Alvarez did not use enumerative induction. Instead, his thinking was hypothetico-predictive in form. The use of causal hypotheses, planned tests and predictions as guides to inquiry seem particularly transparent in other aspects of Alvarez's research as well. Note in Argument One how Alvarez and Lowrie's causal crustal-rotation hypothesis led them to collect rocks in the Apennines and chart the magnetic direction of their mineral grains. Indeed, a look at all eight arguments reveals that hypotheses and planned tests direct the collection of data rather than use of induction.

But what provoked Alvarez to generate causal hypotheses? The answer clearly seems to be the causal questions, which in turn were provoked by *single* observations. Interestingly, these observations were not always the expected ones. In other words, sometimes while looking for one thing, something quite unexpected was found (e.g., many foram species apparently became extinct abruptly at the KT boundary; an unusually large amount of iridium was found in a KT boundary deposit). In retrospect, these unexpected results played a crucial role in guiding the researchers toward their ultimate discovery. Thus, in a sense, Alvarez was lucky to have made these unexpected observations. But as pointed out, these observations did not inductively lead to hypotheses.

Where Do Hypotheses Come From?

If hypotheses are not derived inductively, then where do they come from? Consider, for example, the meteor impact hypothesis of Argument Five, which was generated in response to the question: What sort of catastrophic event caused the apparent global distribution of iridium at the KT boundary? Recall that the Alvarez team knew about Meteor Crater in Northern Arizona. They also knew about the many craters on the surface of the moon and on other planets. And they believed that these craters were caused by impacts of asteroids and comets. Was this information used and if so how? The answer can be found in Alvarez's account. Here is what he said:

I believe that as Dad, Frank, Helen and I tried to make sense of the iridium anomaly, we sometimes talked about a giant impact, but we could not understand why an impact would cause worldwide extinction... Finally Dad started thinking about the dust that would be thrown into the air by an impact. He remembered reading that the 1883 explosion of the Indonesian volcano, Krakatoa, had blown so much dust and ash into the atmosphere that brightly colored sunsets were seen for months in London, on the other side of the world... Scale the Krakatoa event up to the size of a giant impact, thought Dad, and there would be so much dust in the air that it would get dark all around the world. With no sunlight, plants would stop growing, the whole food chain would collapse, and the result would be a mass extinction. It was the first good hypothesis for why a large impact would cause a global mass extinction. (ps. 76-77)

This quotation is clear. The impact hypothesis arose not through descriptive investigations and induction. Rather, in this case, the process of hypothesis generation seems to involve taking an idea from one or more previous contexts (e.g., Krakatoa exploded and blew dust into the air and blocked the sun) and borrowing it for use in a new context (e.g., so a really big meteor might have blasted the Earth and blown enough dust into the air to block the sun for long enough to cause mass extinction). Apparently, the idea was borrowed due to some perceived similarity of the contexts. This is not induction. Rather it is better characterized as one in which analogies are drawn and used. In other words, it is a process of reasoning by analogy, analogical reasoning, or analogical transfer, a process that has long been seen as central to scientific thought (cf. Biela, 1993; Bruner, 1962; Dreistadt, 1968; Finke, Ward & Smith, 1992; Gentner, 1989; Hestenes, 1992; Hoffman, 1980; Hofstadter, 1981; Hofstadter, 1995; Holland, Holyoak, Nisbett & Thagard, 1986; Johnson, 1987; Koestler, 1964; Oppenheimer, 1956; Wong, 1993).

Did Alvarez also use analogical reasoning to generate his other causal hypotheses? The present claim is that he did. For example, in response to the previous causal question of Argument Three, i.e., Where did the unusually high iridium level in the initial KT boundary layer at Gubbio come from?, Alvarez quickly generated these four alternative hypotheses: 1) a supernova, 2) an impacting asteroid or comet, 3) seawater, and 4) a cloud of interstellar dust and gas. Why would these four possibilities have come to mind? As previously mentioned, the supernova hypothesis came from the literature. As also mentioned, the origin of the initial undifferentiated version of the impacting asteroid or comet hypothesis most likely came from an analogy (e.g., a meteor blasted Northern Arizona and most likely killed nearby organisms; so a really big asteroid or comet might have blasted the Earth and killed lots of organisms). But what about the seawater hypothesis? Presumably Alvarez knew about previous contexts in which seawater had deposited iridium. Consequently, he reasoned by analogy that seawater could have deposited iridium in this new context as well. Presumably, his reasoning used to generate the interstellar dust and gas hypothesis was similar, i.e., interstellar dust and gas was known to have deposited iridium in previous contexts, so the dust and gas might also have done so in this context.

The preceding analysis of Alvarez's research makes a case for a theory of scientific method consisting of the following components:

- (1) Scientific discovery begins when one encounters a puzzling phenomenon that cannot be explained by one's currently accepted conceptual systems, theories, mental "models" (e.g., Johnson-Laird, 1983). Thus a causal question is raised (i.e., Why did such and such happen?)
- (2) One then generates one or more tentative explanations (i.e., hypotheses or theories) based on the creative use of analogical reasoning (i.e., borrowing one or more related mental models to see if it/they might serve as explanations in the present context). The tentative explanations are said to answer the causal question such that if a tentative explanation is "correct," the phenomenon would be explained. Of course the discovery process may be cut short if no plausible explanations can be generated. And, as nicely pointed out by Chamberlain (1965), if the initial list of alternatives is too limited, the process may degenerate into a search only for supporting evidence of one's pet explanation, which of course is not to say that scientists would prefer to have their ideas rejected rather than supported (cf., Woodward & Goodstein, 1996).
- (3) Given that more than one alternative explanation has been generated, one selects the most plausible (in terms of its internal consistency, its lack of conflict with previously accepted explanations, its lack of conflict with known evidence, and its parsimony) and attempts to design a test. The test is accomplished through the generation of implied consequences using a "quasi-logical" form of conditional reasoning that employs an *If... and... then...* linguistic form. This reasoning can be characterized as deductive, but this is not a strictly a matter of applying a formal logical rule as one may not be able to imagine a good test. Nevertheless, deriving the test requires that one assume that the tentative explanation is correct and then imagine some test condition(s), which if carried out, would lead to a specific observation, or set of observations. In other words, *If... p* is correct, *and... one does such and such, then... q* should be observed. Here *p* refers to the hypothesis or theoretical postulate being tested and *q* refers to the predicted consequence, the expected observation, the prediction.
- (4) One now conducts the actual test (as proposed above) and observes its results. The actual test requires the collection of one or more of three types of evidence (i.e., circumstantial, correlational, experimental).
- (5) One then compares observed and expected result(s). This may or may not require use of statistical procedures.
- (6) Next one draws a conclusion about the relative support or lack of support for the proposed explanation based on the degree of correspondence between predicted and observed results. Thus the pattern of scientific discovery in cases in which a good match is obtained is similar to that of the logical fallacy known as affirming the consequent (i.e., If *p* then *q*, *q*. Therefore *p* is supported, but not proved). And in cases in which a poor match between predicted and observed results is obtained, the pattern is similar to that of modus tollens (i.e., If *p* then *q*, not *q*, therefore *p* is not supported, but not disproved).
- (7) Next, one recycles steps 3-6 until all of the plausible hypotheses have been tested and either "accepted," or "rejected," or a new puzzling observation is encountered that raises a new causal question. Depending on the nature of the new causal question, it may lead to a "closer" inspection of the phenomenon in question, or it may divert the scientist's attention and start the entire process over again, but in a new direction.

Importantly, in this theory, puzzling observations raise causal questions. Causal questions in turn provoke the generation of alternative explanations, which in turn drive the search for additional observations used in their test. The process is cyclical such that new puzzling observations may arise during the process of hypothesis testing, which may lead to another cycle of explanation generation and so on. Indeed, observations are puzzling if they are unexpected in the sense that they contradict predictions derived from prior explanations. In agreement with Cohen & Nagel (1934), Popper (1959; 1965), Platt (1964), Chamberlain (1965), Hempel (1966), Medawar (1969), Lewis (1988), Moore, (1993), and Lawson (1995; 1999), the discovery process is essentially hypothetico-predictive in nature. And in agreement with Popper (1965) and Musgrave (1999), enumerative induction plays no role.

Jung's Interrogative Theory as an Alternative

The above hypothetico-predictive theory shares several characteristics with that of Jung's interrogative theory of scientific method (Jung, 1996). However, in Jung's theory, initial causal questions do not lead immediately to the generation of hypotheses. Instead, Jung argues for an intermediate "analytic" phase in which: "The next moves should be made so as to turn the principle *why*-question into a *whether*-question." (p.134) According to Jung, this requires that the scientist, "Analyze the phenomenon in question into its essential

ingredients or factors whose properties, interrelations or patterns *E* are known or knowable." (p. 134) Jung likens this phase to strategies used by detectives handling a crime scene in search of clues that might say something about the unknown criminal. In the case of Alvarez's research, after raising the key causal question (i.e., What caused the KT mass extinction?), Jung, in his analysis of Alvarez's research, claims that Alvarez entered the analytic phase in search of clues. Consequently, in Jung's view, Alvarez raised and answered a whole host of descriptive questions, all prior to the generation of hypotheses. According to Jung, these questions included the following:

"How long had it taken to deposit that layer of clay at the KT boundary?" (p. 158)

"Is there anything by which one can measure the duration of the KT boundary with precision to less than one year?" (p. 159)

"How much iridium does the KT boundary layer from Gubbio contain?" (p. 160)

"Is the anomalous amount of iridium in the KT boundary peculiar to and an indispensable part of the unknown cause of the mass extinction?" (p.161)

"Is this iridium anomaly a worldwide feature of the Cretaceous-Tertiary boundary?" (p. 161)

"Where had the KT boundary iridium come from?" (p. 162)

"Is there any way by which the iridium of the earth's core could have accumulated in the KT boundary?" (p. 162)

In Jung's view, after answering these descriptive questions, Alvarez was at last in a position to consider alternative hypotheses that might answer the initial causal question. Thus Jung's view of Alvarez's specific research procedure, and of scientific method in general, is at odds with the above hypothetico-predictive (HP) account. According to the HP account, causal questions first lead to alternative hypotheses, which then direct further observations that are used in their test. In contrast, in Jung's interrogative theory, the causal questions first lead to a series of descriptive questions, which when eventually answered lead to alternative hypotheses. So the specific question in the Alvarez case is this: Were the above descriptive questions raised and answered before or after Alvarez generated one or more hypotheses to answer the initial causal question? If Argument Two of Table 1 adequately represents what Alvarez was thinking, the answer is that the hypotheses came first. In that argument, the catastrophic event hypothesis directed Alvarez to examine the KT layer for iridium. In other words, without the hypothesis, he would have had no reason to look. Alvarez's own account (Alvarez, 1997) makes this point very clear:

It took me a while to realize that the thin bed of clay at the KT boundary at Gubbio not only raised the question of what caused the mass extinction, but that it also seemed to contradict the gradualistic mind-set of geologists. The near extinction of forams at Gubbio looked abrupt. Al Fischer stressed in his talk at Lamont that the dinosaur extinction had happened at the same time. Could *T. rex* have perished in a catastrophic event? (p. 58)

So here we have it. Alvarez has raised the causal question and immediately understood that the evidence seemed to contradict the gradualist view and seemed to support the alternative catastrophic hypothesis. So instead of attempting to answer a host of descriptive questions, as required by Jung's theory, Alvarez immediately sought a way to test the alternative hypotheses, i.e.,:

By 1976, I was focusing my attention on the KT extinction and I began to discuss it with Dad. What specific question could we ask about the clay bed at Gubbio? What useful measurements could we make? I pointed out that it would be valuable to know how long it had taken to deposit the clay layer. Very rapid deposition of the clay would suggest a sudden cause for extinction, but a slow deposition would suggest a gradual mechanism. (ps. 60-61)

So Alvarez's attempt to find out how long it had taken to deposit the clay, Jung's second descriptive question, was not the result of a search for clues. Rather it was motivated by a desire to test the alternative gradualist and catastrophic hypotheses (i.e., If...the mass extinction was caused by a gradual phenomena, and...we measure how long it had taken to deposit the clay layer, then...a long time period should be found. On the other hand, if...a catastrophic event was the cause, then...a very short time period should be found). Therefore, this analysis of the Alvarez case history supports the hypothetico-predictive theory of scientific method and does not support Jung's alternative interrogative theory. Likewise, the analysis does not support an inductivist theory. Importantly, after reading the present account Alvarez had this to say: "I am very much in agreement with with you. I can not imagine putting off the interesting why questions and generating causal hypotheses until later. I also found that the eight arguments in Table 1 not only captured the essence of our past research, but the pattern has proved helpful in thinking about our present research." (Alvarez, personal communication, March 2000)

A personal aside may be appropriate at this point. If science is indeed hypothetico-predictive as claimed, then why do so many non-scientists continue to view it as something else? I think the answer lies at least partly in the way scientists often portray their work. I have attended many scientific presentations outside my research field and come away thinking that the presenter had done nothing but raised and answered a seemingly disconnected series of descriptive questions. Only upon talking to the presenters, or to others in the presenters' fields, did I discover that the descriptive questions were not disconnected. They had been driven by causal hypotheses after all. But the presenters had omitted mention of the hypotheses and predictions because they assumed that the audience already knew what they were. Another source of confusion lies in fact that many scientific papers and texts are equally misleading about method (cf., Gibbs & Lawson, 1992; Medawar, 1990). Indeed, as another source of confusion, consider how statistics texts often confuse students about hypotheses and predictions when they routinely refer to null *predictions* as null *hypotheses*!

Conclusions and Implications

The Alvarez case history reveals that at the core of scientific discovery lies a hypothetico-predictive method in which observations first provoke the raising of causal questions. The causal questions are followed by the generation of alternative causal hypotheses via a process of analogical reasoning. Enumerative induction plays no role in hypothesis generation. The generated hypotheses are then tested through the derivation of their implied consequences and by a comparison of predicted and observed results. The process of deriving implied consequences, although often labeled as one of deduction, is not strictly a matter of applying a rule of logic. Rather, a element of creativity is implied as one may have a hypothesis but be unable to imagine any good way to test it. In the sense that neither hypothesis generation, nor hypothesis test, involves the strict application of logical rules, a logic of scientific discovery does not exist.

Although the present theory of scientific method shares many characteristics with previous hypothetico-deductive accounts, it rejects logical positivism in that strict hypothesis falsification is not possible. Further, science is seen primarily as a process that proceeds by first testing the most plausible hypotheses in terms of consistency with prior experience and with accepted theory before moving on to those that seem less likely. Hence the present view avoids the problem of having to test all possible hypotheses for "progress" to occur. The view that induction plays no role in scientific method draws into serious question, not only the analytic phase of Jung's interrogative theory of scientific method, but also attempts by some social science researchers to employ the inductivist-inspired grounded theory approach to inquiry (e.g., Strauss & Corbin, 1997; Creswell, 1998; Dey, 1999). The present view is that, although interesting causal questions emerge from observations, theory does not. Consequently, the grounded theorists' efforts would be better spent identifying interesting causal questions and using analogical reasoning to brainstorm alternative hypotheses, rather than gathering mountains of uninformative descriptive data.

As far as science classrooms are concerned, the implications are clear. Considerably more time should be spent having students conduct rich explorations in which they encounter puzzling observations in need of explanation. After causal questions are raised, students should brainstorm lists of alternative hypotheses. Next, instructional time should be spent having students, singly and in cooperative groups, design and conduct tests of their hypotheses. Importantly, students should be required to generate explicitly stated predictions prior to data collection. The expectation of such instruction is that students would benefit in terms of improved scientific reasoning skills (cf. Johnson & Lawson, 1998; Lawson, Clark, Cramer-Meldrum, Falconer, Sequist & Yong, 2000). They can also be expected to improve their understanding of the nature of scientific method, especially if aspects of the method were explicitly discussed (cf. Bell, Lederman & Abd-El-Khalick, 1998; Abd-El-Khalick, 1999).

Acknowledgement

The author wishes to thank John Alcock, Walter Alvarez, Russell Benford, and Lisa Wolpowitz for their helpful comments on an earlier version of this paper.

References

- Abd-El-Khalick, F. (1999). The influence of history of science courses on students' conceptions of the nature of science. Unpublished doctoral dissertation. Oregon State University, Corvallis, OR.
- Alters, B.J. (1997). Whose nature of science? *Journal of Research in Science Teaching*, 34(1), 39-55.
- Alvarez, W. (1997). *T. rex and the crater of doom*. Princeton, NJ: Princeton University Press.
- Bell, R.L., Lederman, N.G. & Abd-El-Khalick, F. (1998). Implicit versus explicit nature of science instruction: An explicit response to Palmquist and Finley. *Journal of Research in Science Teaching*, 35(9), 1057-1061.
- Biela, A. (1993). *Psychology of analogical inference*. Stuttgart, Germany: Hirzel Verlag.
- Bruner, J. (1962). *On knowing: Essays for the left hand*. Cambridge: Belknap Press., Harvard University Press.
- Burks, A.W. (1980). Enumerative induction versus eliminative induction. In L.J. Cohen & M. Hesse (Eds.), *Applications of inductive logic*. New York: Oxford University press.
- Cohen, M.R. & Nagel, E. (1934). *An introduction to logic and scientific method*. London: Routledge and Kegan Paul.
- Chamberlain, T.C. (1965). The method of multiple working hypotheses. *Science*, 148, 754-759. (Original work published 1897)
- Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. London: Sage.
- Curd, M. (1980). The logic of discovery: An analysis of three approaches. In T. Nickles (Ed.), *Scientific discovery, logic, and rationality*. Dordrecht: D. Reidel.
- Dey, I. (1999). *Grounding grounded theory: Guidelines for inquiry*. San Diego, CA: Academic Press.
- Dreistadt, R. (1968). An analysis of the use of analogies and metaphors in science. *Journal of Psychology*, 68, 97-116.
- Finke, R.A., Ward, T.B. & Smith, S.M. (1992). *Creative cognition: Theory, research and applications*. Cambridge, MA: The MIT Press.
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge: Cambridge University Press.
- Gibbs, A. & Lawson, A.E. (1992). The nature of scientific thinking as reflected by the work of biologists and biology textbooks. *The American Biology Teacher*, 54 (3), 137-152.
- Gustason, W. (1994). *Reasoning from evidence: Inductive logic*. New York: Macmillan College Publishing.
- Hestenes, D. (1992). Modeling games in a Newtonian world. *American Journal of Physics*, 55, 440-454.
- Hempel, C. (1966). *Philosophy of natural science*. Upper Saddle River, NJ: Prentice-Hall.
- Hoffman, R.R. (1980). Metaphor in science. In P.R. Honeck & R.R. Hoffman (eds.), *The psycholinguistics of figurative language*. Hillsdale: Erlbaum.
- Hofstadter, D.R. (1981). Metamagical themas: How might analogy, the core of human thinking, be understood by computers? *Scientific American*, 249, 18-29.
- Hofstadter, D.R. (1995). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*. New York: Basic Books.
- Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P.R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: The MIT Press. (see especially Chapter 11)
- Johnson, M.A. & Lawson, A.E. (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? *Journal of Research in Science Teaching*, 35(1), 89-103.
- Johnson, M. (1987). *The body in the mind*. Chicago: University of Chicago Press.
- Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Jung, S. (1996). *The logic of discovery: An interrogative approach to scientific inquiry*. New York: Peter

Lang.

Koestler, A. (1964). *The act of creation*. London: Hutchinson.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.

Lawson, A.E. (1995). *Science teaching and the development of thinking*. Belmont, CA: Wadsworth.

Lawson, A.E. (1999). What should students learn about the nature of science and how should we teach it? *Journal of College Science Teaching*, 28(6), 401-411.

Lawson, A.E., Clark, B., Cramer-Meldrum, E., Falconer, K.A., Kwon, Y.J., & Sequist, J.M. (2000). The development of reasoning skills in college biology: Do two levels of general hypothesis-testing skills exist? *Journal of Research in Science Teaching*, 37(1), 81-101.

Lewis, R.W. (1988). Biology: A hypothetico-deductive science. *The American Biology Teacher*, 50(6), 362-367.

Malherbe, M. (1996). Bacon's method of science. In Peltonen, M. (Ed.), *The Cambridge companion to Bacon*. Cambridge, England: Cambridge University Press.

Medawar, P.B. (1969). *Induction and intuition in scientific thought*. Philadelphia: American Philosophical Society.

Medawar, P.B. (1990). Is the scientific paper a fraud? In Medawar, P.B., *The threat and the glory: Reflections on science and scientists*. Pyke, D. (Ed.). Oxford, England: Oxford University Press. Moore, J.A. (1993). *Science as a way of knowing: The foundations of modern biology*. Cambridge, MA: Harvard University Press.

Musgrave, A. (1999). How to do without inductive logic. *Science & Education*, 8, 395-412.

Platt, J.R. (1964). Strong inference. *Science*, 146, 347-353.

Popper, K.R. (1959). *The logic of scientific discovery*. New York: Basic Books.

Popper, K.R. (1965). *Conjectures and refutations: The growth of scientific knowledge*. New York: Basic Books.

Russell, D. & Tucker, W. H. (1971). Supernovae and the extinction of the dinosaurs. *Nature*, 229, 553-554.

Schick, T.S. Jr. & Vaughn, L. (1995). *How to think about weird things: Critical thinking for a new age*. Mountain View, CA: Mayfield.

Strauss, A., & Corbin, J. (Eds.), (1997). *Grounded theory in practice*. London: Sage.

Terry, K.D. & Tucker, W.H. (1968). Biologic effects of supernovae. *Science*, 159, 421-423.

Wong, E.D. (1993). Self-generated analogies as a tool for constructing and evaluating explanations of scientific phenomena. *Journal of Research in Science Teaching*, 30(4), 367-380.

Woodward, J., & Goodstein, D. (1996). Conduct, misconduct and the structure of science. *American Scientist*, 84, 479-490.

Table 1

Eight Major Hypothetico-Predictive Arguments Advanced by Alvarez During His Research

	Argument One
	The Apennines mountains are deformed. (observation)
	What caused deformation of the Apennines? (causal question)
	Rotation of the Earth's crust caused the deformation. (proposed explanation)
	<i>If...</i> rotation of the Earth's crust caused deformation of the Apennines, <i>and...</i> rocks that contain magnetic mineral grains are collected from various locations within the Apennines (planned test), <i>then...</i> a progressive rotation of the magnetic grains should be found with the older "fossil compasses" twisted farther away from north than the younger ones (predicted result).
	<i>But...</i> although the fossil compasses did generally point west of north, local bed disruptions made it impossible to separate local movements from plate rotation (observed result).
	<i>Therefore...</i> one can not tell whether or not rotation of the Earth's crust did in fact cause deformation of the Apennines (conclusion).
	Argument Two
	While working their way down through the Cretaceous rocks and up through the Tertiary, Alvarez and Lowrie found fossil forams abundant and as big as sand grains in the top beds of the Cretaceous, but found only the very smallest forams in the first beds of the Tertiary. (observation)
	This observation implied that the forams had very nearly, and abruptly, become extinct. Thus the following questions were raised: Why had the forams almost become extinct? And why was it so abrupt? (causal questions)
	The extinction of many foram species, and possibly the dinosaurs, was caused by a catastrophic event. (proposed explanation)
	<i>If...</i> the extinction of many foram species, and possibly the dinosaurs, was caused by a catastrophic event, <i>and...</i> the amount of iridium contained in the clay layer at the KT boundary is measured (planned test), <i>then...</i> a relatively small amount of iridium should be present, i.e., about 0.1 ppb (predicted result). Iridium in meteorite dust falls to Earth at a constant rate, thus the less iridium in the layer, the less time it must have taken for its deposition (theoretical rationale).
	<i>But...</i> a huge value of 9 ppb was detected (observed result).
	<i>Therefore...</i> either the extinction of many foram species, possibly the dinosaurs, was not caused by a catastrophic event, or perhaps the catastrophic event itself deposited the unusually large amount of iridium (conclusion).
	Argument Three
	An unusually large amount of iridium was found in the KT clay layer (observation)
	What caused the unusually large amount of iridium? (causal question)
	Perhaps it was caused by a world-wide catastrophic event such as a supernova, an impacting asteroid or comet; perhaps it was somehow deposited from sea water; or perhaps the Earth had encountered a cloud of interstellar dust and gas. (proposed explanations)
	<i>If...</i> the unusually large amount of iridium was caused by some sort of world-wide catastrophic event, <i>and...</i> the amount of iridium is measured in the other known KT boundary layer at Stevens Klint (planned test)
	<i>then...</i> an unusually high level of iridium should be found in that layer (predicted result).
	<i>And...</i> the Stevens Klint KT layer did contain an unusually large amount of iridium (observed result).
	<i>Therefore...</i> the world-wide catastrophic event explanation was supported (conclusion).
	Argument Four
	Unusually high levels of iridium are found in two geographically distant KT boundary layers. (observation)
	What sort of catastrophic event caused the unusually high iridium levels in the two geographically distant KT boundary layers? (causal question)
	A supernova caused the unusually high iridium levels. (proposed explanation)
	<i>If...</i> a supernova caused the unusually high iridium levels,

*and...*the amount of plutonium-244, which is known to be deposited by supernovas, is measured in a KT boundary layer (planned test),

*then...*plutonium-244 should be detected in that boundary layer (predicted result).

*But...*after one false positive reading, plutonium-244 was not detected (observed result).

*Therefore...*the supernova explanation was not supported (conclusion).

Argument Five

Unusually high levels of iridium are found in two geographically distant KT boundary layers. (observation)

What sort of catastrophic event caused the unusually high iridium levels in the two geographically distant KT boundary layers? (causal question)

A meteor impact in or near an ocean (which ejected dust and ash into the atmosphere blocking the sun and leading to mass deaths throughout food chains), caused the unusually high iridium levels in the two geographically distant KT boundary layers. (proposed explanation)

*If...*the unusually high levels of iridium were caused by a meteor impact in or near an ocean,

*and...*several KT boundary layers throughout the world are examined (planned test),

*then...*one or more of those layers should contain tsunami deposits (predicted result). The tsunami deposits are expected based on the assumption that the oceanic impact created a tsunami (theoretical rationale).

And... the KT boundary layer at the Brazos River in Texas was found to contain a tsunami deposit (observed result).

*Therefore...*the oceanic meteor impact explanation was supported (conclusion).

Argument Six

Tsunami deposits are found at the Brazos River in Texas. (observation)

What caused the tsunami that left the deposits? (causal question)

The tsunami was caused by the meteor that created the Chicxulub crater. (proposed explanation)

*If...*the tsunami was caused by the Chicxulub meteor,

*and...*a KT layer near the Chicxulub crater in nearby northeastern Mexico is examined (planned test),

*then...*it should contain a tsunami deposit (predicted result).

*And...*the KT layer at Arroyo el Mimbral in northeastern Mexico was found to contain a tsunami deposit nearly three meters thick precisely at the KT boundary (observed result).

*Therefore...*the Chicxulub meteor explanation was supported (conclusion).

Argument Seven

Furthermore,

*If...*the Chicxulub meteor caused the tsunami,

*and...*core samples taken from the Chicxulub crater are examined for melt-rock (planned test),

*then...*those core samples should contain melt-rock of KT age (predicted result).

*And...*the core samples did contain melt-rock of KT age (observed result).

*Therefore...*the Chicxulub meteor explanation was further supported (conclusion).

Argument Eight

Alvarez and Mexican geologists found rocks in northeastern Mexico containing tiny gas bubbles. (observation)

What caused the tiny gas bubbles? (causal question)

When the Chicxulub meteor slammed in Earth creating the tsunami, it also ejected melt rocks. When the rocks cooled, tiny gas bubbles were trapped inside. (proposed explanation)

*If...*the tiny gas bubbles formed when the ejected melt rocks cooled,

*and...*other KT layers in northeastern Mexico are examined (planned test),

*then...*the layers should also contain tsunami deposits as well as rocks containing tiny gas bubbles (predicted result).

*And...*nine new KT layers were found, each containing tsunami deposits and rocks with tiny gas bubbles (observed result).

*Therefore...*support was obtained for the cooling rocks explanation (conclusion).