JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 35, NO. 2, PP. 189–212 (1998)

Criteria of Excellence for Geological Inquiry: The Necessity of Ambiguity

Charles R. Ault, Jr.

Teacher Education Program, Lewis & Clark College, Box 14, 0615 Palatine Hill Rd.,
Portland, Oregon 97219-7899

Abstract: According to Gowin, a curriculum properly derives its authority by representing the "criteria of excellence" for evaluating the claims produced within a field of inquiry. Gowin's epistemology applied to examples from geological inquiry yields criteria of excellence responsive to the demands characteristic of geological problems. Student efforts to learn these criteria hold the promise of making progress toward independence in accessing, using, and evaluating knowledge. This understanding contributes to the reformation of the concept of inquiry as a "step beyond science as process" called for in the *National Science Education Standards* and reinforces the need to consider the diversity as well as unity of styles of scientific reasoning. Geological inquiries differ from those of other sciences because they refer to objects with histories. These histories create a demand for concepts that necessarily contain an irreducible element of ambiguity, thus permitting comparison and contrast of geological objects. A case study of how geologists apply analogies, impose boundaries on categories of thought, and constrain the ambiguity of key concepts in reasoning about the accumulation of sediments at a continental margin is used to support this argument. Such examples of geological reasoning support a skeptical attitude toward interdisciplinary curricula that omit or oversimplify criteria of excellence. © 1998 John Wiley & Sons, Inc. J Res Sci Teach **35:** 189–212, 1998.

"In the vision presented by the [*National Science Education*] *Standards* [NSES], inquiry is a step beyond 'science as process'" [National Research Council (NRC), 1996, p. 105]. Typical of ideas for teaching about this "step beyond" is Gowin's knowledge vee or epistemological vee (Gowin, 1981; Novak, 1990; Novak & Gowin, 1984). Many science educators are familiar with Gowin's vee, and there is an emerging body of literature on its efficacy in science teaching (e.g., Esiobu & Soyibo, 1995; Lehman et al., 1985; Okebukola, 1992; Roth & Roychoudhury, 1993). The vee is a convention for diagraming relations among concepts and methods of inquiry, framing both in the context of a focus question and the event of interest. However, the vee greatly abridges Gowin's philosophical program in answer to his own focus question, "What is to be taught?" (Gowin, 1981, p. 120). No response is possible without addressing how a curriculum might properly derive its authority. To answer this challenge, one must begin by making explicit the criteria of excellence for evaluating the goodness of an inquiry within a particular context. Gowin first offered the vee as a heuristic useful for dissecting the claims and practices of original inquiries in the search for criteria of excellence applicable to specific domains. However, the aim of this essay is neither to summarize nor to extend psychological studies of the efficacy of the vee. Instead, the objective is to inquire into the history and philosophy of science applied to geological inquiry to obtain insights about geology's criteria of excellence.

CCC 0022-4308/98/020189-24

At the same time, the goal is to understand a fundamental and distinctive feature of some important geological concepts: their essential ambiguity. These ends—knowledge of criteria of excellence and understanding the necessity of ambiguity—serve a pragmatic purpose: determining *which* aspects of original inquiry and primary sources merit translation into educative materials for teaching geology widely and properly. The exercise models how analyses might proceed in search of criteria of excellence appropriate for other disciplines.

This analysis of the role of ambiguity in forming geological concepts and the search for criteria of excellence particular to this field prompt skepticism toward efforts to portray science in school as abstract ideas and general processes independent of well-defined problem domains. Reasoning, most simply characterized as compare and contrast, typifies geological inquiry. For example, unlike atoms of carbon-12, submarine fans are not utterly indistinguishable and completely interchangeable parts of a conceptual system. The depositionary history of a particular submarine fan makes it distinct from, while being similar to, other members of this class of phenomena. Effective geological reasoning ultimately depends upon constraining the ambiguity of key concepts. Procedures for doing so depend upon conscious use of metaphor and analogy as well as upon an explicit grasp of the purposes concepts serve in a well-defined context. In conjunction with judicious ways of constraining ambiguity, excellent geological reasoning requires that (a) independent lines of evidence converge upon the same conclusion, and (b) explanations on different scales in space and time cohere. Differentiated as several principles of historical methodology in science, these criteria suggest the importance to curriculum construction of showing how disciplined thinking responds to the demands characteristic of particular phenomena of interest. In addition, highlighting the irreducible and inevitable ambiguity embedded in geological concepts helps to illuminate the psychological obstacles to learning this subject meaningfully.

A Basis for Authority in Science Teaching: What Makes Good Science Good?

Criteria of Excellence

According to the educational epistemology developed by Gowin (1981), among the most important and well-warranted claims derived from a context of disciplined inquiry are the field's criteria of excellence. These criteria serve as the basis for making judgments about the merits of the inquiry process itself, the persuasiveness of explanatory findings, and the grounds for adjudicating disputes regarding interpretation and significance. In short, the articulation of criteria of excellence is necessary to answer the question, "What makes good science good?" In turn, answers to this question are needed in guiding reform efforts in science education. Ideas about what makes good science good may not inform teaching practices directly, but they do bear upon the authority of the curriculum and its status as the carrier of a trusted heritage from disciplined inquiry.

Criteria of excellence are hierarchical in nature and often common across disciplines at high levels, though higher level principles are neither reducible to nor derivable from those they subsume. Thus, one may speak of the criteria of excellence for science, biology, or population ecology. What makes particular fields of inquiry distinct resides in the uniqueness of their criteria of excellence or in their relative strengths of attachment to criteria of excellence they hold in common with other fields. In progressing through time, fields of inquiry produce their own methods and strategies for solving problems, and these methods and strategies are characteristic of the phenomena under scrutiny. Geology is not physics. Ecology is in very important ways not reducible to chemistry. Neither geology nor ecology is evolving to becoming better adapted to the criteria of excellence governing the search for greater unity in physics theory. Fields produce and follow their own criteria, appropriate to the class of problems defined in large measure by how the conceptual apparatus of the field imposes boundaries—inventions of mind—upon experience. As summarized by Kitcher, ". . .we discover more about the world while simultaneously learning how to investigate the world . . ." (1993, p. 202).

According to this view, becoming disciplined in the pursuit of inquiry means becoming adept in applying criteria of excellence appropriate to a particular phenomenon of interest. As a discipline of inquiry, a field must remain open to the possibility for change even in its criteria of excellence. The reasons justifying such change comprise the highest level of criteria for making judgments about the reasonableness of current knowledge, the productive potential of (or need for) other approaches to inquiry, or the direction of attention to newly appreciated phenomena. Fields of inquiry are not closed circles.

According to Gowin, the distillation of a field's criteria of excellence is a crucial first step toward construction of a valid curriculum. Once a tradition of disciplined inquiry lays claim to a productive heritage (e.g., how DNA works, how continental margins have formed), it may rightfully impose itself upon the attention of others outside the research community. The name given this imposition and the methods of intervention it justifies is "schooling."

A vital analogy exists between the open pursuit of inquiry and the design of school curricula. When in the realm of university education the goal of convincing citizens to hold a proper conception of what it means to be good supersedes commitment to open inquiry, perverse social conditions may thrive—witness Germany in the 1930s. For schooling to avoid promoting an analogous pathway to social injustice, its curriculum ought to reflect the diverse heritages of disciplined inquiry and present these heritages to students in valid forms of representation. In surveying a proper pathway for schooling to follow, the signs to post are the criteria of excellence that warrant the claims of various traditions of disciplined inquiry. To grasp such criteria is to liberate thinking and promote democratic justice by sharing academic competence widely in a society. To ignore them is to promote dependency and encourage the power of an elite by virtue of its access to the production of knowledge.

The heritage of disciplined inquiry and the promise its understanding holds for the good of a student's life are indispensable requirements for justifying the coercive nature of schooling. To teach knowledge that fails a test of excellence, to indoctrinate students with misrepresentations of the heritage of a field of inquiry, is grossly irresponsible. But it happens—each time compromising the intellectual liberty in a student's life. One step away from this unhappy situation is an understanding of what makes good science good. That may be the most practical knowledge brought by a teacher to the classroom. If it is missing, all else is adrift; the activities of teaching are without a rudder—or a basis of authority.

In Gowin's *Educating* (1981), the first reference to criteria of excellence pertains to all subjects. Later, he enumerates criteria of excellence for the natural sciences. In reference to educative materials and their significance as representations of primary sources, he wrote:

> First of all, they carry with them appropriate criteria of excellence. For some materials the central excellence is a version of knowledge and truth. For others, criteria of beauty, elegance, and artistic merit dominate. Still other materials exhibit criteria of human judgment, like justice in politics, fairness in ethics, meaningfulness in education. Since these criteria are to be found in [primary source] materials and retrieved for use in educating, they are criteria that are public, open, definite. These public, open qualities help enormously in reducing oppression in educative episodes. The teacher does not unilaterally control the access to [primary source] materials; the student can learn to be free of any teacher once a legitimate way into the [primary source] materials has been found and tested, and as it makes sense of the students' experience. (p. 54)

Several pages later, he continued:

> Among the most important claims of any field are the *criteria of excellence* the field uses. In science, two criteria often found together are reliability and validity. The reliability criterion is used to judge the excellence of knowledge claims produced by an inquiry by analyzing whether techniques of measurement pass tests of repeatability, reproducibility, universalizability. . . . A claim is judged valid if it refers to the piece of reality it purports to refer to, and if that portion of reality is thought to be important to the science. . . . [Some] theories are judged better than others because they are more coherent or more comprehensive or more fruitful in generating telling questions. . . . What makes good science good is a fundamental concern of scientists and those who use scientific knowledge. (pp. 85–86)

Curriculum

Educative materials can be seen as records of prior events which human beings can use to make new events happen. Gowin wrote of the curriculum as analogous to a musical score (the curriculum) read to make music happen (the educative event), or architectural drawings read in constructing a house. The most educative curricula are those most successful in making new events happen from the record of prior events *so long as that record is a valid representation of the original inquiry*. Validity implies embodiment of criteria of excellence.

Defining criteria of excellence begins with a search through relevant exemplars of original inquiry or primary resources. The purpose is to select studies acknowledged, through subsequent practice within a particular field, to "carry with them the criteria of excellence helpful in judging the works themselves" (p. 116). A telling example from biology is Darwin's *Origin of Species* (Gould, 1986). Not only did Darwin claim that descent with modification by means of natural selection explains the origin of species, he also explained the reasonableness of his method of argument.

Educative materials of high worth embody criteria of excellence. These are principles open to public scrutiny and they serve to make judgments about the merits of claims of knowledge in the context of a particular line of inquiry. When the consistency of a claim with its appropriate criteria of excellence is persuasively argued, the claim is well warranted. When this consistency fails, examples of pseudo-science, creationism, perpetual motion machines, and controversies over ethnocentric sciences abound.

Trust in knowledge depends upon appeal to criteria of excellence—the working capital of a community of scholars. These criteria do not guarantee truth by any means; they serve only to adjudicate disputes about the reasonableness of claims. When learned meaningfully (i.e., connected to the personal interpretation of experience), they enhance the liberty of an individual to make judgments within a field. Coupled to the recognition of the function of ambiguity and knowledge of appropriate ways to constrain this ambiguity in the context of an inquiry, such criteria yield intellectual power.

Authority, Rationality, and the NSES

Authority and Rationality

According to Gowin, good science teaching depends upon intervening in students' lives with educative materials that embody criteria of excellence. In deciding what and what *not* to teach, educators must choose carefully how to represent science. Their authority and the legitimacy of the curriculum demand consistency between the curriculum and the criteria of excellence characteristic of a field of inquiry. Moreover, the curriculum ought to reflect the importance of conceptual features distinctive to the field, as distilled through repeated application of the criteria of excellence to knowledge claimed by inquiry.

Thus, a search for criteria of excellence and an understanding of the distinctive features of concepts essential to a particular field of inquiry contribute to defining one of the key standards for science education: teaching science as inquiry. The newly released NSES presented a vision of reform in which "inquiry is a step beyond 'science as process.' . . . [The] new vision . . . requires that students combine processes and scientific knowledge as they use scientific reasoning and critical thinking to develop their understanding of science" (NRC, 1996, p. 105). What, precisely, does the phrase "step beyond" mean? In the quotation, it refers to scientific knowledge and scientific reasoning. The premise of this essay is that examining scientific knowledge and reasoning in the context of a particular field of inquiry contributes substantially to grasping what these terms mean, and in turn suggests what making a step beyond science as process might necessarily entail.

Earth Science Curriculum Content

The NSES for the content of earth and space science teaching have already incorporated, to a limited degree and in very generalized terms, some of the themes developed here. For example, in setting objectives for student understanding in the middle grades, Content Standard D expresses how scientific reasoning in the earth and space sciences depends upon an understanding of "the earth and solar system as a set of closely coupled systems" (p. 158). In addition to expressing the importance of thinking in terms of systems and system components to the rationality of understanding in the earth and space sciences, the standards document added acknowledgement of the challenge presented to students of earth science by "evidence from sources that range over immense time scales" (p. 188) and the additional constraint that "direct experimentation is usually not possible for many concepts associated with earth and space science" (pp. 188–189).

Earth systems evolve historically across many scales in time and space. Competing inferences about this evolving past unaided by direct experimentation demand other methods of adjudication. Often these methods focus on comparisons and contrasts among different examples of how a system works; the boundaries for defining not only the system but also its components require careful, yet inevitably incomplete, constraints upon the ambiguity of concepts (e.g., the case of the Astoria submarine fan detailed later).

At stake in the teaching of science, as the national standards aptly express, are the conceptions of rationality intimately associated with scientific inquiries. Too hastily and too often, in this author's judgment, the science education community has tried to enumerate the processes of science disembodied from particular domains of inquiry. In summary, making the step beyond science as process and examining reasoning within a well-defined context hold promise for clarifying scientific notions of rationality.

Toulmin's Pragmatic Notion of Rationality

Philosopher Stephen Toulmin argued for the acceptance of distinct methodologies and multiple models of rationality, each well suited to the characteristics of specific problems and phenomena (Toulmin, 1990). He urged greater stature for knowledge that is timely, local, and particular, while cautioning restraint in the pursuit of knowledge that is universal, general, and timeless. In many respects, the disciplines of physics attempt to warrant universal, general, and timeless explanations of phenomena. In contrast, attention to timely, local, and particular aspects of phenomena is critical to disciplined geological inquiries.

In the 20th century, physics has moved from a Euclidean geometry of nonintersecting parallel lines expressing the predictions of Newton's laws of motion to a non-Euclidean geometry of space–time curvature expressing the predictions of Einstein's relativity theory. Throughout the evolution of physical theory from the 16th century to the present, geometrical abstraction has proven crucial, as emphasized in this quotation from an interview with physicist Edward Witten of Princeton's Institute for Advanced Study: "What we have learned in the 20th century, is that the great ideas in physics have geometric foundations" (Witten, in Horgan, 1991, p. 46).

Science Methods Match the Demands Characteristic of Problems

What we are learning in the late-20th century, argued Toulmin (1990), is that what is true for physics (the necessity of and quest for axiomatic geometric foundations) is not necessarily true for other fields of inquiry. The rejection of a criterion defining excellence for physics theory as a standard of rationality for other domains does not, however, mean rejection of physics theory within those other domains. Toulmin forcefully argued these positions, and in response to recognizing the demise of the goal of finding a Euclidean, axiomatic model of theoretical rationality in all domains, concluded, "every science will need to employ those specific methods that have proved, in concrete experience, to match the characteristic demands of its own intellectual problems" (p. 193). Toulmin's position, however, does not call for increasing the compartmentalization of knowledge. Quite to the contrary, he argued:

> The intellectual tasks for a science in which all the branches are accepted as equally serious call for more subdisciplinary, transdisciplinary, and multidisciplinary reasoning. Like the informal procedures of the common law when it is functioning at its best, these interlocking modes of investigation and explanation check exaggerated claims on behalf of all universal theories, and reinstate respect for pragmatic methods appropriate in dealing with concrete human problems. (p. 193)

The primary goal of Toulmin's work was, as quoted above, "to reinstate respect for pragmatic methods appropriate in dealing with concrete human problems." Problems stemming from concrete human experience—as opposed to those defined by theoretical structures—are of necessity transdisciplinary, with methods shaped in accordance with the nature of the phenomena of interest. They are also local, particular, and timely, and share features in common with geological reasoning. This kind of problem solving calls for a variety of ways of thinking, each matching the "characteristic demands of its own intellectual problems."

Retrodiction (or Postdiction) Is a Characteristic Demand in Geology

According to Kitts' seminal work *The Structure of Geology* (1977), the most "characteristic demand" of geological problem solving is the concern for singular, not universal, statements. Such statements are about particular events and are by nature probabilistic, genetic, and "sketchy" (p. 25). By genetic, Kitts meant that the aim of such singular statements is to provide an account of the causal derivation of a particular feature of the earth in terms of plausible and empirically verified processes. In Kitts' words, "The goal of geology is the derivation and testing of singular descriptive statements about the past. . . . Retrodiction, not prediction, is the most characteristic inference . . . past events must be inferred" (p. 39). To retrodict is to infer the past from the present. Retrodictions are, in addition, predictions of what else to find in the record of the earth's past, given sets of trusted statements about its history.

Geologists often reason by comparing and contrasting structures of interest such as mountains, valleys, and river deltas. Through such reasoning, they acquire a "trained judgment of exceptions" (p. 35). They argue what is critical to resolve and what can be left unresolved. For this reason, phrases such as "tends to," "distinctive of," "resembles," "typically," "distinct from," "in contrast to," and so forth dominate explanatory language. The goal is confidence in the retrodiction; certainty remains elusive, although widespread agreement has a rational foundation.

"To judge geological generalizations by their ability to support predictions is absurd" claimed Kitts, for "the future is more uncertain than the past" (p. 46). One can at best imagine and assign probabilities to the plausible possibilities the future holds, constrained by the facts of the past, as reliably known.

Geology is a science whose realm is change over vast scales in time and space, and not a science in search of universal theoretical structures. These characteristics of geology, despite the advent of plate tectonic theory, have not changed as much as the demands on it. Geologists have been asked to guarantee—by legal standards of liability—the safety of nuclear waste repositories for 10,000 years, predict earthquake and volcanic eruptions in time to evacuate population centers, determine the earth's petroleum reserves, calculate the effect of changes in atmospheric composition on the polar ice caps, and warn us of the next impending collision with an intruder to the earth's orbit, to name a few.

Kitts judged the demand for geological prediction absurd because he acknowledged that extrapolation multiplies uncertainty. Still, the judgment is too harsh. Retrodiction provides a background against which humans might evaluate the scale of their own interventions in nature, from trends in the composition of the global atmosphere to the deterioration of local watersheds. Awareness of the scale of human perturbations of nature by comparison with geological ones means learning to appreciate inescapable risks while acknowledging an inability to predict the future precisely. The "pragmatic rationality" of geology does include prediction, once the sense of prediction is detached from the formal "geometric foundations" of physics and crafted in keeping with the "characteristics demands of its [geology's] own intellectual problems." Kitts' concept of retrodiction provides an essential insight into these characteristic demands. Furthermore, such retrodictions or historical inferences become the foundation for making comparisons and contrasts among different systems or structures of the globe, all in keeping with the explanatory ideal of making the dynamic systems of the earth intelligible.

Gould's Second Style of Science

Gould (1986) argued that the "most enduring impact" (p. 60) of the work of Darwin is the prominence accorded historical methodology as the basis "for an entire second style of science" (p. 60). Darwin, claimed Gould, taught how and why to study homology and the more general concept of contingency (or change dependent upon the accumulation of changes) as a means to render complex systems intelligible.

Following Gould, the second style of science is the result of the historical method of inquiry descended from Darwin and Lyell. For Gould, the historical method is the appropriate response to the demand for extrapolation from observable events across time and place on unobservable scales. Four principles dominate this method: (a) uniformitarianism, the classic dictim "the present is key to the past"; (b) stage theorizing, the substitution of place for time; (c) relic interpretation, the examination of oddities or complexly superimposed events; and (d) proper taxonomy, the knowledge of how descriptive categories themselves embody the causal principles of an explanatory framework.

Uniformitarianism Assumes No Divine Intervention

What Darwin adopted from Lyell in the most straightforward sense was the principle of uniformitarianism, elegantly case as the present is key to the past. To a first approximation, this principle directs a scientist to examine small, steady processes of change in the present, then extrapolate their effects over geological time. Darwin applied this principle to life (for example, by extrapolating the process of selective breeding to infer the effects of natural selection over time) just as Lyell had to rock strata. The principle of uniformitarianism validates prediction either forward or backward (retrodiction) in time.

Once again, something distinctive about geological reasoning appears to be apparent. Despite whatever promise a rule guiding thinking about the past may have, it ultimately fails as a scientific law in the sense of universal applicability (Rogers, 1989). Why? Because the phenomena of interest are inextricably connected to consequences of historical changes. As objects, they accumulate characteristics on account of their histories. Features of their past exist in the present, limiting the possibilities of change in the future. This complexity holds equally true whether the object of interest is a layer of volcanic ash, a fold texture in a metamorphic rock, a deep-sea sedimentary fan, a hanging glacial valley, or a floodplain levee.

The principle of uniformitarianism is the foremost guide to making reliable historical inferences, but it is not a law. In the often intractable context of immense spans of time, uniformitarianism suggests plausible extrapolations. Cast as "actualism" (Leveson, 1980), this principle constrains geological inferences to the range of possibilities not forbidden by the trusted theories of chemistry and physics. From the point of view of actualism, the events that constitute geological phenomena are among the universe of possibilities subsumed by physical laws extrapolated through time and space. Very importantly, this modern sense of uniformitarianism makes no theoretical distinction between the imagery of catastrophic and gradual rates of change as did geologists in the time of Lyell. Any rate of change is entirely possible on virtually any scale, if reconciled with the consensus judgments of present-day chemists and physicists.

Place Substitutes for Time

Consider that geological processes, some cyclical, some irreversible, occur across many time scales and locations, progressing at various rates and commencing at different times. Geologic objects in present time in effect sample moments from these processes of change. Because there are so many sequences commencing and progressing through time, the objects one observes now very likely capture salient features of change processes which cannot be directly observed on the human time scale. The challenge to retrodiction ("historical inference" to Gould) is to hypothesize an arrangement by stages for what is observed. In an arrangement by stages, place substitutes for time. This hypothesis includes a commitment to categories of description—in effect, stages. Classifying features of the earth as stages of a process conforms to how the features of the earth are conceptualized in keeping with some hypothesis of an accurate historical sequence. Descriptive category and explanatory mechanism ride in the same boat. As Schwab instructed, the conception of stages—a mental invention—tells us what facts to look for. In addition, stages are part of the conceptions telling us what meaning to assign to these facts. In Schwab's words:

> Today, almost all parts of subject-matter sciences proceed in this way. A fresh line of scientific research has its origin not in objective facts alone but in a conception, a deliberate construction of the mind. On this conception, all else depends. It tells us what facts to look for in research. It tells us what meaning to assign to these facts. (Schwab, 1962, p. 199)

What did Darwin have in mind as an illustration of the principle of arrangement in stages? Gould cited Darwin's *The Structure and Distribution of Coral Reefs* (1842) as the best example. In this treatise, Darwin described reefs as "fringing, barrier, and atoll." This classification conforms to his historical hypothesis that these three types of atolls observed in the present are the historical consequence of slowly sinking islands over different periods of time. One island is an example of another island's past. A different island is an example of its future.

Knowledge of coral reef growth is necessary to the generation of this hypothesis. Alternative hypotheses without any claim of a mechanism responsible for producing an orderly sequence of island stages are certainly plausible: Erosion or volcanic processes, for example, might compete with the hypothesis of sinking. Any island might be any island's past or future; no island is a stage, only a state.

Darwin's contemporaries, explained Gould, made counterarguments based upon colonization by corals of differently eroded platforms in the ocean. Fringing reefs, barriers, and atolls exhibited no process orderly in time. They reflected the happenstance of erosion and the opportunistic growth of coral.

Modern science has vindicated Darwin's case for the sinking island and enhanced our understanding of the volcanic processes shaping the origins and fates of ocean islands. Drilling technology in the 20th century has confirmed Darwin's hypothesis by revealing that reefs thicken progressively from fringing reef to atoll, consistent with reef growth while subsidence occurs once volcanism has ceased. The convergence of independent lines of evidence on the same explanation increases its credibility, and in this way historical inferences (retrodictions) become convincing arguments.

Arrangement in stages abounds in geology as a goal of problem solving. Stages describe volcanic arcs, river basins, plate margins, etc. As a principle governing reasoning, it first requires description of present patterns, ultimately in terms of categories based upon some historical hypothesis of orderly sequence. This sequence reflects the accumulation of changes and the historical contingency of succession of one stage upon the completion of another.

Most likely, more than one hypothesis will emerge when "deliberate constructions of mind" wrestle with complex histories. The problem then becomes one of searching for critical tests which may confirm or contradict the plausibility of a causal mechanism linking a sequence of stages. Alternatively, inquiry may turn to the search for an independent line of thought that converges upon the same arrangement of stages. When the search for evidence in support of a stage hypothesis returns empty-handed, Gould's second style of science makes an appeal to a fourth principle of reasoning in this context: the interpretation of relics.

Interpretation Unravels Relics

Stage hypotheses assume that all stages exist somewhere and that these stages contain clues about their derivation. However, when both uniformitarian and stage principles cannot suffice to draw historical inferences, a third principle comes into play: the "principle of imperfection" or "oddity" (Gould, 1986, p. 63). This principle builds upon the premise that sequences of historical change overprint one another, superimposing new patterns on the relics of past events. (Gould chose to use the words *imperfection* and *oddity* in reference to biological structures, implying that evolution must make do with what already exists. Reasonably analogous words in geology that label this principle for making historical inferences in the absence of observable stages are *superposition* and *relic*.)

In the history of an object, what happens next may alter but not obliterate the effects of what has happened before. A single object becomes a complex puzzle of clues about its own history and derivation. Although some of the most important clues may be missing entirely, the relics of processes no longer exerting effects may persist in various states of alteration. Of course, several series of changes may overprint each others' effects with little clue as to sequence. The way out of such puzzles, if there is any, is found by interpreting the record of overprinting within a framework of stage theory, even if examples of the presumed stages do not exist in the present. A view of a full moon on a clear night illustrates the application of this style of reasoning. In the lunar highlands there are relic craters from an ancient period of intense bombardment; this period itself was presumably superimposed after accretion and impact melting had arrived at the stage of a frozen surface. Overprinting the ancient bombardment are the multiringed astroblems—massive craters that define the face of the man in the moon to the naked eye. Other, smaller craters such as Copernicus, surrounded by ray structures formed from their ejecta, cut across all of the older features of the moon. A light bombardment from the time of the Copernican impact presumably continues into the present. Of course, no person has witnessed any other than the last of these stages, and the processes responsible for the moon's current features may no longer be operating in the solar system. Yet, this sketch of lunar history remains a convincing series of inferences logically consistent with what anyone can observe with a good pair of binoculars.

Rocks on many scales from lunar body to crystal lattice often have relics of their past embedded in their structure and composition, obliterated or preserved to almost any degree. To decipher these clues requires knowledge of the processes which yield them—an exercise in the "trained judgment" of the geologist (Kitts, 1977, p. 35). Inferring the order of events in time is essential and perhaps the most difficult aspect of interpreting relics, for a given rock sample may preserve the stresses and strains of very different epochs in its history. Interpreted according to the concepts of plate tectonic theory, its fabric and minerals may record the tortuous stresses accumulated while in the proximity of an active island arc system of volcanoes as well as the intense pressures impressed by a subsequent collision between the island arc and a continent. Sorting amalgamation effects from accretionary ones is very troublesome (especially if they both result in similar patterns), but goes to the crux of the problem of reconstructing the history of a continental margin from the rock record. The problem becomes which events were superimposed upon which others. The logic, however, analogous to the unraveling of the lunar past is nearly the same as a child might use in deciphering the history of a garage-sale toy with an obvious history of repair and use: How was the toy damaged in what order did the repairs occur?

Proper Taxonomies Link Causal Reasoning and Explanatory Categories, Subsuming Stages and Relics

Gould (1986) expressed the relationship between an explanatory goal and the categories of description to apply in pursuit of this goal as "the importance of proper taxonomies" (p. 63). Earlier, Schwab (1962) developed much the same notion in his paper, "The Concept of the Structure of a Discipline." Schwab claimed, in reference to the "deliberate constructions of mind," that "on this conception, all else depends" (1962, p. 199).

Consider as a telling example of the significance of the proper taxonomy principle how in the history of the theory of plate tectonics working hypotheses influenced even what counted as data (LeGrand, 1988). The example illustrates how closely descriptive categorization conformed to the causal presuppositions of competing explanations. According to LeGrand, in the study of the magnetism observed to vary among seafloor rocks from place to place, there existed disagreement over what was "pattern" and what was "noise" in the data, depending on whatever working hypothesis was favored. LeGrand's account of this process classically typified Schwab's position that theory serves as the principles for inquiry ("on this conception, all else depends"), guiding not only the interpretation of the data but the judgment of what constitutes data itself:

> Specialists in palaeomagnetism routinely "cleaned" their samples and their data, whether the focus was directionalist or reversal studies or magnetic anomaly mapping. It was neither the "raw" specimens nor the "raw" data which found their way into the literature, but rather specimens and data adjusted according to various theoretical beliefs and expectations. Consider the famous Eltanin-19 profile (the nineteenth of more than twenty profiles from the study). The data from which it was constructed were gathered using a complex network of theories and methodological presuppositions and equipment which itself was a concrete manifestation of theory. Marine scientists had no direct observations of magnetic stripes. The data were then processed and "cleaned" in accordance with theory and so on. But, that to which I wish to draw attention is that the other profiles were tossed aside in favor of Eltanin-19: it demonstrated clearly the bilateral symmetry expected on the basis of Vine's refined model. In a sense, Eltanin-19 "made" Vine-Matthews and Vine-Matthews gave Eltanin-19 a significance it would not otherwise have possessed. It is an apparently clear-cut case of selecting the "best" data where "best" is defined as fitting most closely theoretical expectations and thereby "confirming" them. The other profiles were judged to contain more "noise"; i.e., the symmetry which they displayed was less striking, less convincing. (LeGrand, 1988, p. 220)

The resolution of this debate about magnetic anomaly data required something more than greater precision. Rejection of one argument (susceptibility to magnetization owing to compression) and confirmation of the competing argument (seafloor spreading and the generation of new oceanic crust) depended, ultimately, upon reconciliation with other lines of inquiry (recognition that seismicity and volcanism outline seafloor plate boundaries, for example), independent of the taxonomic categories and their built-in tendency for self-confirmation used to account for the remnant magnetism of ocean crust.

Resolved as parallel magnetic stripes laid down symmetrically about an axis of magma generation, the seafloor, magnetic data reconcile midoceanic ridges with the fate of ocean crust in deep marine trenches. Most old seafloor sinks back into the mantle, and fresh upwellings replenish the oceanic crust. If these mechanisms are trusted, new interpretations of puzzling continental landforms become possible. There are mountains in California (and throughout the world) understood to be remnants of seafloor rock that escaped the fateful marine trenches as plates collided. Now plate tectonic theory enjoys a cornucopia of descriptive categories—a proper taxonomy—such as *ophiolitic suite, tectonic melange, spreading center, convergent boundary, arc volcanics,* etc. Each theory-laden term carries causal implications.

Descriptive Categories and Causal Frameworks in Geology

While Schwab's perspective may apply to all sciences, it seems to acquire particular importance in geology. Arguments over the descriptive categories (e.g., magnetic stripes) and the causal frameworks (e.g., sea-floor spreading) they imply is often close to the surface of debate, in part owing to the complexity of the systems studied, and in part to the impossibility of conducting controlled experiments (there is only one earth with only one history). Laws are unobtainable in the same sense as in physics and chemistry. Of course, crucial observations are often possible (magnetic anomaly), models desirable (ocean floor cycling through the mantle), and predictions testable (the chemistry of magmas based on mantle, ocean floor, or continental crust origin), but the science remains fundamentally interpretive rather than experimental, with the warrant for an argument dependent upon integration and reconciliation of independent lines of inquiry.

Indeed, in geology, ordering into categories is often derived from a theory of causal ordering independent of a criterion of experimental reproducibility, a situation placing special demands upon the proper taxonomy principle:

> In a profession more observational and comparative than experimental, the ordering of diverse objects into sensible categories becomes a sine qua non of causal interpretation. A taxonomy is not a mindless allocation of objective entities into self-evident pigeon-holes, but a theory of causal ordering. Proper taxonomies require two separate insights: the identification and segregation of the basic phenomenon itself, and the division of its diverse manifestations into subcategories that reflect process and cause. (Gould, 1986, p. 63)

The principle of proper taxonomies begins with the virtually self-evident assertion that a system of classification serves a purpose, and continues with acknowledgment that knowing this purpose is critical to drawing the boundaries among categories. Moreover, grasp of the causal interpretation embodied in a system of categories—realization of the worth of a mental invention in rendering phenomena intelligible—prevents the perception of a taxonomy as being arbitrary or arcane. As a corollary, an improper taxonomy is one that either fails to serve its explanatory purpose or fails to render phenomena intelligible.

Fundamental Entities in Geology Contrasted with Those of Chemistry

In very large measure, the fundamental entities of the proper taxonomies for chemistry and physics are dealt with theoretically as interchangeable parts utterly independent of histories. "If you have seen one carbon-12 atom, you have seen them all," one might say, for the history of events in which a particular atom of carbon-12 has participated is of no interest in carbon chemistry. The single atom may have been decomposed from an oak tree leaf, escaped from the soil into the atmosphere as gaseous carbon dioxide, absorbed by a planktonic creature to make a carbonate skeleton, buried in seafloor sediment adjacent to the Astoria fan, subducted beneath the margin of North American, and, 500,000 years later, erupted in an episode of Cascade Arc volcanics.

Of course, given any particular atom of carbon-12, there is no means of determining such a history of interactions. Therefore, if one attempts to predict what might happen next to this particular atom of carbon-12, none of this history (except perhaps a recent change in energy state) has (or perhaps can have) any bearing on an answer. Rather, it is the universal, timeless, mathematical abstraction of the properties of carbon-12 and the mathematical characterization of the conditions surrounding an atom of carbon-12 that make possible any such prediction. Predictions remain intractable for any individual atom of carbon-12 but apply successfully to large numbers of these fundamental entities because they have identical properties and are all equally good examples of their class.

Equally importantly, there is in principle no ambiguity in determining the identity of a substance such as carbon-12, for the task of observing carbon-12 is fully operationalized into instrumental procedures for measuring isotope mass and emission spectra. Even more telling, the borders between an example of carbon-12 and of nitrogen-14, for example, are sharp and well defined in principle, regardless of the sensitivity of any measuring device. This precise border is possible because the members of a class are utterly indistinguishable from each other. It is the properties defining class membership which count. These properties—defined by mathematically formalized relations as, for example, among force, mass, and inertia—count equally well for all members of the class. There is no atom of carbon-12 more typical than another, none somewhat distinctive from other examples of its class. In fact, to use carbon chemistry in explaining complex phenomena with histories—such as the carbon cycle phases of plant growth and volcanic eruptions—these features of the conceptual framework of chemistry are prerequisites. They are inherited from the more general scope of particle theory and the foundational premises of quantum mechanics.

It may seem that the seen-one-seen-all proposition about the constituent elements of matter is an empirically testable hypothesis. Hanson (1958) argued convincingly against this position. Even without recounting his argument fully, one can realize the importance of a conceptual framework in deciding this issue from the following summary:

> It is an indispensable condition of quantum theory that all electrons, all protons, all neutrons, must be identical . . . from no more data than the number of electrons in an atom of some element, many of its physical and chemical properties [can be calculated]. These predictions are supported in experiment . . . however, certain properties must be "worked into" the concept of the electron . . . the absolute identity of all electrons is a property they must have if they are to explain and pattern all the observations which gave rise to the electron-concept in the first place.
>
> The power and success of quantum theory consists in the pattern of interlocked, systematic accounts it gives of the behaviour of complex bodies. Since it does this only by postulating the absolute identity of all elementary particles of the same type, what better reason could there be for saying that all elementary particles of the same type are identical? . . . Quantum theory is not mathematics: a conception like this would not have been formed *no matter what*. It is justified in every microphysical environment, now and during many years past; indeed, without this conception experiments would not even make sense. All the data, the facts, the observations, bear the stamp of this unifying conception.
>
> Why does a body free of impressed forces move in a straight line? Why does Mars describe an elliptical orbit? Why does the force of gravity vary inversely with the square of the distance? Why does nothing more faster than light? Why are all electrons identical? Because the world as we now know it becomes intelligible by supposing these things to be the case. What better reason for saying they are the case? (pp. 133–134)

A criterion of intelligibility provides a means of judging the worth, not the ultimate truth, of a concept. This criterion applies to all conceptual frameworks in science and provides both start and end points in the examination of the rationality of atom and sediment. Both concepts meet the criterion of making the world intelligible. They serve an explanatory purpose. However, they may still differ in important conceptual respects as well. Accounts of sediments depend upon the tenets of the molecular-kinetic theory of matter. Yet, satisfying explanations of sedimentary structures require articulation in a language of categories irreducible to particle theory. Beds, ripples, transport mode, and strain are not the sediments themselves but categories for thinking about muds, sands, and cobbles. Force, mass, energy, and chemical change in turn are classes of phenomena for thinking about ripple formation, transport velocity, and sediment deformation. As Hanson (1958) asserted, "In general, though each member of a class of events may be explained by other members, the *totality* of the class cannot be explained by any member of the class. The totality of movement cannot be explained by anything which moves. The totality of red things cannot be explained by anything which is red" (p. 120). The totality of geological processes cannot be explained solely in terms of geological processes, as any geologist will concur. However, it does not follow that the conceptualization of geological phenomena will conform to the same patterns of rationality built into the superordinate class of phenomena.

Each stage through which the imaginary atom of carbon-12 passed in the earlier example represents a fundamental geological category: soil, atmosphere, sediment, subduction, or volcanism. At every stage in this carbon cycle, each geological object or process is at some unique stage in its own historical development and shares at some level of description features of other soils, sediments, submarine fans, and volcanic arcs. At another and equally crucial level of description, each object or process diverges in a unique way from all other examples of its category. And at each stage of the cycle, an understanding of the chemistry of carbon matters—its reactivity, its volatility in gaseous form, etc. What matters about the carbon matters equally, no matter where and when in space and time. What matters about each geological object does not extend so universally.

How Geological Reasoning Constrains Ambiguity: The Case of the Astoria Submarine Fan

Consider one phase in the carbon-cycle example as exemplified by the submarine fan off the coast of Astoria, Oregon, by beginning with a text excerpt defining the geological category "submarine fan":

> Great quantities of sediment accumulate at the base of the continental slope in the form of the continental rise. The rise is a system of numerous coalescing *submarine fans* or cones, each of which typically contains *fan valleys,* distributary channels, and *suprafans*. . . . The fan valley typically displays levees and is a basal continuation of a submarine canyon. As the turbidity currents and other gravity-flow phenomena exit the canyon across the fan apex, there is a loss of confinement and distributaries are formed in much the same fashion as on alluvial fans or riverine deltas. (Davis, 1983, pp. 537–538)

The phrase "submarine fan" labels an object in a way responsive to the purpose of interpreting the development of continental margins through geological history. The term functions to "signify regularity" (Gowin, 1981) by describing several defining criteria. Its relationships with other concepts and regularities impose critical constraints on its meaning. Neither the defining criteria nor the constraining relationships define "submarine fan" in an absolute sense. Unlike concepts in physics such as mass or inertia, this geological concept apparently functions quite satisfactorily without being defined in terms of formal mathematical abstraction of its relations to other concepts. Nor is there a fully specified operational definition of how to identify this feature of the continental margin and ocean floor completely free of ambiguity.

In a general sense, grasping the meaning of "submarine fan" appears to depend upon (a) examining how (or under what conditions) the phrase is used, and (b) ascertaining the purpose of its use. Examination of the text passage reveals how the meaning of "submarine fan" is constrained by use and purpose when thinking geologically. The constraints needed to communicate the meaning of "submarine fan" with a minimum of ambiguity stem from several sources, including (a) metaphor, (b) physical processes, (c) analogy, (d) boundary conditions (counter-case examples), (e) explanatory framework, (f) explanatory goals, and (g) clear case example. Each of these seven constraints is crucial to establishing the meaning of the category "submarine fan."

Yet, even when all of these constraints are taken together, ambiguity remains in the meaning of "submarine fan." This ambiguity is a necessary hallmark of concepts used in geological reasoning. The generation of well-warranted singular statements regarding the histories of geological structures (the goal of geological reasoning) depends upon comparing and contrasting them. Because of their separate histories, no two members of a class of phenomena are identical. Hence, the ultimate boundaries of the class remain inevitably obscure. A fully formalized definition would limit a category to a single member and preclude thinking in terms of comparisons and contrasts among members of the same category.

Certainly there are reasonable and clearly expressible ways of setting limits to class membership. These limits, if not understood within the context of an explanatory purpose, however, appear to be arbitrary and arcane to the novice learner (Finley 1981, 1982). Misunderstanding of how purpose imposes reasonable boundaries (what Gould called "reasonable taxonomies") among concepts may compound the frustration caused by a geological concept's necessary ambiguity.

The selection of the example submarine fan in arguing this case about the nature of conceptualization in geology is illustrative of these points. Whatever the case selected as a category for reasoning in geology, it is certain to illustrate how important it is to ascertain the use and purpose of a concept and how its ambiguous meaning at some level is a necessary response to the characteristics of geological problems and historical methods. Interestingly, the playfulness found in asking questions such as "Is ice an igneous rock? Glacial ice a metamorphic one? Snow a sedimentary one?" may offer the best means of appreciating the necessarily ambiguous status of concepts and how human-centered purpose sets reasonable boundaries for their meanings (cf., Custer, 1991; Romey, 1983).

Returning to the case in point, how is one to determine, given the text excerpt above (Davis, 1983, pp. 537–539, 546), what "submarine fan" means? By the use of (a) metaphor, (b) physical processes, (c) analogy, (d) boundary conditions (counter-case examples), (e) explanatory framework, (f) explanatory goals, and (g) clear case example.

Seafloors Have Metaphorical Fans. First, there is the metaphor "fan." Notice that Davis' description also includes the phrase "or cone." Sediments accumulate in the form that is to be understood as like a fan-shaped section of a cone. What has this initial use of metaphorical imagery accomplished, and to what advantage or disadvantage? It rests upon the reasonable assumption that those curious in making sense out of the margin of continents share images of cones and fans. However, the initial definition is not operational—it does not specify the activity or measurements an observer must engage in to determine the existence of a submarine fan. The advantage of the metaphor is its efficient evocation of prior knowledge. The disadvantage is the amount of ambiguity in the initial definition. The metaphor alerts thinking to certain salient features of the object of interest interpreted within an already familiar context.

Fans Share Processes of Fan Making. The connection of submarine fans to a series of processes moving and depositing sediment dramatically reduces the ambiguity of the metaphor. The processes themselves are cast in general terms—interchangeable in time and place. The category is now construed as a product, the end effect of a series of events (e.g., turbidity currents, gravity flow phenomena, formation of distributaries). However, this series comprises a generalized, not specific, order in time. Moreover, each process may occur to some greater or lesser extent than another. The identification of processes creating submarine fans works in many respects like an operational definition, but because of the variation of these processes in time and magnitude, ambiguity remains. The products of these processes may exhibit (in fact, necessarily exhibit) variation. Hence, process fails to specify fully what the category "submarine fan" means. Some might say these processes constitute a model of submarine fan formation. At the same time, they would acknowledge that each example of a fan might mean adjusting the model.

Fans Resemble Other Fans and Deltas, But Not Canyons. The next move is to invoke analogies to other systems and products of deposition: alluvial fans and riverine delta, canyons and levees. As with the fan-cone metaphor, this analogy connects prior knowledge to the task of new conceptualization. The analogy presumes grasp of a body of domain-specific knowledge and helps to clarify the relation of process to product. Grasp of the meaning of "submarine fan" proceeds according to an understanding of its resemblance to other systems. At this stage, comparing and contrasting dominate thought as a search for resemblances.

Boundaries Have the Most Interesting Conditions. Of course, establishing resemblance to another structure cannot go too far or the analogy will have erased anything distinctive about submarine fans. There needs to be a description of differences between deltas and fans, for example. The category requires the delineation of "boundary conditions" (Kitts, 1977). Thus, the differentiation of submarine fans from other depositionary structures continues in the text (not quoted here), with detail about the nature of turbidite deposits, sheetflow, and the role of the levees in channelization, in relation to "upper, mid, supra, and lower fan" structures when depicted in cross-sectional profile. There is no abrupt break with the submarine fan feature and adjacent structures: "The distal margin of the fan lacks definition because it merges with similar sediments of the abyssal plain" (Davis, 1983, p. 538). The definition began with the acknowledgment that individual fans "coalesce." Thus, even with the attempt to describe boundary conditions, ambiguity remains. The learner may likely have an unclear understanding of how to distinguish borderline examples of deltas extending past a narrow continental shelf from true submarine fans. There cannot be a boundary to the concept of submarine fan without comparison to other related categories, such as "delta."

Frameworks Are for Explaining the System as a Whole. The fifth move is to embed the concept of "submarine fan" in a wider framework of thought. First, the text authors refer to the fan as a "fan complex," then remark, "The discussion above considered the general characteristics and dominant processes of the canyon and fan environments separately. To understand the sedimentology and stratigraphy of the accumulated materials in these environments, it is necessary to consider the system as a whole" (Davis, 1983, p. 538). For several pages, the text explains the concept of a "submarine canyon-fan system," using the example of the Astoria Canyon-Fan adjacent to the coast of Oregon. The explanatory framework is one of systems thinking, where an account of order is constructed in terms of inputs, energy dissipation, interactions, and products. This general framework contains the principles and laws of chemistry and physics as well as general models of geological processes. This "system as a whole" quotation contains the sixth and seventh means of constraining the meaning of "submarine fan," as well as any other concept within this framework.

The Grasp of Meaning Depends on Knowing the Purpose. Most importantly, from the "system as a whole" quotation one learns the goals for problem solving in this domain. These goals are to understand (a) the sedimentology and (b) the stratigraphy of "*these* [emphasis added] environments." Geologists impose boundaries, describe processes, select analogies, and invoke metaphor in order that they might invent concepts useful for this purpose. The "worth of the concept" (in the sense developed by Toulmin (1972), for whom "worth" implies "fruitful of inquiry") depends upon its facilitating the goals of understanding sedimentology and stratigraphy in the environments of the continental *shelf*. Italicizing *shelf* emphasizes that this entire seven-step process of constructing meaning may recurse, beginning with this new metaphor evocative of shape. "Shelf" makes reference to another object of interest on a larger scale.

A Good Explanatory Example Tells the Worth of a Concept. Does the submarine fan-canyon concept have worth? Does it work to solve a geological problem? How does an example bring closure to the meaning-making process for "submarine fan"? The last question is perhaps the easiest to answer. The concept submarine fan has a restricted context in which it is presumed useful, further constraining its meaning but still without reducing ambiguity to zero. The text selection uses the example of the Astoria Canyon-Fan system off the Oregon coast as "a rather typical example of this deposition system" (Davis, 1983, p. 539). Note the qualifier: "rather typical." It is neither the perfect example nor the example with exactly the same features found in all other examples—just "rather typical." This qualifier in effect acknowledges that there is one and only one Astoria Canyon-Fan system on earth (and no other anywhere else, for that matter). *The understanding of its sedimentology will eventually incorporate what is distinctive as well as typical about its sediments*—for they are derived from a complex of sources from Oregon to Idaho and beyond, including flood basalts, accreted (and metamorphosed) terranes, granitic batholiths, multiple cycles of arc volcanics, and recycled sedimentary rocks. The geology of each of these sources in turn is understood by the same emphasis on its resemblance to other examples, yet with crucial differences existing on account of separate histories for each example. The sought-after statements are indeed singular.

The understanding of the stratigraphy of the Astoria Canyon-Fan must also incorporate what is distinctive as well as typical about the sedimentary environments and their response to tectonic changes in the alignment of basins, the location of trenches, and the disruption of coastal settings by compressional folding, extensional unroofing of ancient rocks, and transform (or strike-slip) faulting altering coastline and drainages. Each of these tectonic features is similarly understood within a framework of their resemblance to other regimes, yet again with historical reasons for uniqueness among examples. Coordination of when each happened with respect to another is absolutely essential to the reasonableness of explanations about the construction of the Astoria Canyon-Fan. Of course, in both sedimentology and stratigraphy the laws of chemistry and physics are necessarily presumed not only to decipher the settings of geological clocks, but also to determine resemblances and differences among examples of geological phenomena—sediments and their structures—in one time and place from another.

How does the Astoria Canyon-Fan example help bring closure to the meaning-making process? "Its dimensions and general characteristics fall near the middle of the spectrum for such systems" (Davis, 1983, p. 539). The phrase "near the middle" is especially revealing. Real-world examples can be averaged along the defining dimensions of the category: number of distributary channels, cross-sectional profile, etc. There exists no operationally defined procedure for deciding unequivocally in all cases if a given feature of the earth is or is not a submarine fan; yet, defining ways of measuring departure from the category mean for a particular feature approaches an operational definition. If resemblances to other "rather typical" examples are strong, then the case for thinking in terms of canyon-fan systems is called for.

For several pages the text selection describes in detail the sequence of sediments from core locations throughout the Astoria Canyon-Fan system. From such sequences, geologists construct models integrating patterns in the sequence and distribution of sediments and processes of transporting, depositing, and reworking deposits. Diagrams based on these models depict the system from top-down and cross-sectional perspectives, the former representing processes over a region at a particular moment in time and the latter the pattern of accumulation of sediments through time. If no regularity exists from either perspective, no model is possible. Gradients may change, walls of canyons may slump, and currents may change direction. Nevertheless, "The simplified model does serve as an excellent tool to aid in interpreting the rock record" (Davis, 1983, p. 546).

Conclusions

The Worth of Constraining Yet Accepting Ambiguity

The concept of canyon-fan systems has worth: It aids in interpreting the rock record. Much hinges on the single word *interpret*. Professors of geology frequently acknowledge that the exciting (and sometimes ego-bruising) aspects of work in the field turn on arguments over interpretation. Geology majors may even take a course in geological interpretation. Geological interpretation is not simply the everyday notion of individual and subjective differences in opinion. Rather, a fundamental aspect of geological argumentation is at stake. Interpretation is a carefully considered goal of geological explanation—a goal with the same status as a criterion of excellence as the geometrical foundation has for physics.

In the simplest sense, interpretation is a retrodiction—a warranted series of inferences yielding a historical sequence of events accompanied by a causal model (when plausible) for their occurrence. Think of an interpretation as a chronicle with diagrams. Once, for example, an investigation of the Astoria Canyon-Fan has achieved this goal, the geologist may go public and commence to debate the interpretation. The community of inquiry presented this interpretation should agree on the data directly tied to measurements of physical properties, or be able to fault the measuring procedures. They may disagree on which data to collect or even which measurements count as noise, because they differ on the descriptive categories to apply (the proper taxonomy) as a consequence of their different trainings and theoretical preferences. They may disagree on the importance of data no one has been able to gather, and so on, with the irreducible core of agreement anchored in the unproblematic aspects of applying the laws of chemical and physical theory (Leveson's actualism).

Obviously, for many reasons, seldom is retrodiction unequivocal. High among these reasons is the ambiguity of the categories for thinking geologically and the failure of virtually any geological principle to generalize to all cases. Although they may contain quite a few carbon atoms, submarine fans are very different objects of scientific interest. However, if concepts such as submarine fans were defined in such a way as to eliminate ambiguity completely, they would cease to serve a useful purpose in geological explanation. The ambiguity is built into the role of using resemblances, of calling upon analogy, of judging similarities and differences with other cases. It appears that categories for thinking geologically necessarily entail ambiguity to draw conclusions in the form of "distinct from" or "typical of." ("The Astoria Canyon-Fan is typical of sedimentary systems of the continental rise but with features distinctive of its relation to sediment sources and regional tectonics," for example.) In both of these senses, the inference is historical in nature, made with confidence based upon principles of historical argument and the explanatory ideal of interpretation, not with certainty based upon deduction from universally trusted premises.

Since explanatory purpose is essential in constraining ambiguity and establishing meaning, formalized predictions in geology face tremendous skepticism. Formalized predictive theory (as contrasted with informal, retrodictive theory) responds to criteria governing scientific reasoning that can prove misleading in making historical sense of complex systems. In writing about his attempts to understand the evolution of the continents, for example, Oliver (1983) concluded, ". . .the earth may not operate according to the principle of Occam's razor. There is no natural requirement that the earth do things in the simplest possible manner or that it have the simplest possible structure" (pp. 293–294).

Keeping in mind this anti-Occam's razor principle leads to the realization that making sense of the continental shelf off the coast of Oregon depends upon thinking about it in relation to other such environments present and past. In doing so, concepts helpful in ordering resemblances and productive in organizing critically distinctive features are required. "Submarine fan" works because it establishes an appropriate balance between imagery and operational definition, remaining inevitably and sufficiently ambiguous.

Criteria of Excellence for Geology

The criteria of excellence for geology are those for evaluating its responses to the demands characteristic of interpreting the earth. As constructed in this article, there are several candidates for these criteria: (a) constraining ambiguity when reasoning by comparison, (b) evaluating independent lines of inquiry for convergence, (c) constructing proper taxonomies, (d) accommodating methods of historical inference to the problems of extrapolating systems, and (e) integrating geological arguments across scales. A suitable label and summary for each of these criteria follows. The concluding Implications section conforms to this same sequence and set of labels.

Constraints on Ambiguity. Because no two examples of any category in geological thought (such as submarine fan) are identical, ambiguity remains in any definition of the concept. Procedures for constraining the ambiguity of concepts are essential for comparing and contrasting objects with histories, but these same histories require definitions of categories with some degree of ambiguity. Ambiguity both enables and limits inquiry.

Independent, Converging Lines of Inquiry. Geological investigations benefit from commitment to generating multiple and independent lines of inquiry to evaluate the degree to which they converge upon a common solution. Demonstrations of independence and convergence increase confidence in solutions. (This criterion is a more general form of integration across scales.)

Proper Taxonomies. Classifying serves a purpose. Deliberate expression of the explanatory purpose makes categories of description meaningful. Categories are wrong to the degree they obstruct the explanatory purpose, and right to the degree they promote its accomplishment.

Extrapolating Systems through Time. In response to the demands of extrapolation across immense spans of time and the scale of missing evidence, geologists differentiate the historical method of science into several strategies that promote extrapolation. These strategies assume no divine intervention while adhering to a modern principle of uniformitarianism, substituting place for time to arrange objects in stages, and interpreting relics as overprinted processes.

Integration across Scales. Perhaps no criterion is more important to the evaluation of geological claims than the determination of whether they cohere into arguments well integrated across different temporal and spatial scales. McPhee's conversations with the geologist Eldridge Moores (McPhee, 1993) repeatedly revealed how the test of integration across scales warrants geological claims. From studies at the scale of mineralization and fault propagation to hypotheses about plate subduction and mountain building, good geological inquiries interlock in mutually supporting and noncontradictory ways. When they do not (perhaps always the norm at some level of analysis), the mismatch prompts further inquiry. Consider McPhee's interpretation of the basis for trusting geological arguments:

> Where pictures are clearest, the data cross-check with confirming frequency . . . the ophiolitic narrative [formation of seafloor rock] will conform with ancient latitudes preserved in the remanent [sic] magnetism of rock. The fossil record must not disagree. Where strike-slip faults have sliced a landscape and carried two sides apart, matchups can be traced in time and space. Sedimentary sequences, blue-schist belts, batholithic belts, thrust belts, and mélanges will orchestrally tell what happened. If they are not synchronous, it didn't happen. (McPhee, 1993, pp. 216–217)

The chronology of events must cohere synchronously in time across wide spaces. A failed chronology or a lack of synchrony casts doubt upon current claims, demanding new inquiries that satisfy the criterion of integration across scales.

Implications of the Criteria of Excellence in Geology for Science Teaching and Learning

Constraints on Ambiguity. Rationality is enhanced, not compromised, by an understanding of the role ambiguity plays in geological problem solving. Reasonable ways of constraining potential confusion stemming from this necessary ambiguity contribute to making geology distinct from other fields, and managing this ambiguity in an instructional context is of critical importance. The disparity between what lay people and scientists perceive as different purposes served by the same concept may pose obstacles in communicating the rationality of problem solutions (or simple inferences) and lead to false characterizations of nonscientists' thinking as illogical.

As illustrated by the example of sediments at the margin of a continent, this reasoning progresses through the generation of a category from a general metaphor (seafloor fan) to the identification of processes common to examples of this category (erosion, transport, and sedimentation). Next comes a comparison by analogy with other categories of the same geological type as determined by processes of formation (seafloor fan with deltas, canyons, and levees formed by erosion, transport and sedimentation). The imposition of boundary conditions distinguishes the category in question from others that share processes of formation (the distal margin of a fan and the abyssal plain). The problem solution is located within an explanatory framework (sedimentology and stratigraphy for the system as a whole), and the conditions to satisfy to accomplish the explanatory goal (account for the structure and environment of the continental shelf) are enumerated. Finally, the analysis of the case illustrates the explanatory worth of the system of categories and processes (Astoria submarine, fan-canyon complex). Each phase entails necessary ambiguity, yet together they work to constrain ambiguity.

"Regrettably, there is insufficient research into the nature of learning and problem solving . . ." in geological education (Ault, 1994, p. 271). To progress, such research must have available a body of knowledge about the nature of problem solving particular to geology itself. The strategies described as productive for thinking about the Astoria seafloor fan exemplify a starting point.

Independent, Converging Lines of Inquiry. Chamberlin (1890) articulated his method of multiple working hypotheses and its relation to good science at the end of the 19th century (reprinted 1965). The founder of the *Journal of Geology,* he developed this method with particular emphasis on geological examples. Writing with characteristic elegance, he stated:

> In the use of the multiple method, the re-action of one hypothesis upon another tends to amplify the recognized scope of each, and their mutual conflicts whet the discriminative edge of each. The analytic process, the development and demonstration of criteria, and the sharpening of discrimination, receive powerful impulse from the co-ordinate working of several hypotheses. . . . Each hypothesis suggests its own methods of proof, its own methods of developing the truth . . . the total outcome of means and of methods is full and rich. (p. 756)

Do students in typical school science classes entertain multiple working hypotheses in response to a problem? Do they encounter examples of science being conducted as competitions among hypotheses? (Recall the example of magnetic anomaly data.) Do they recognize how solutions earn high levels of trust because independent lines of inquiry converge on the same answer? (Refer back to the example of coral growth and volcanic islands.) When they do, they are experiencing a curriculum that embodies a very high-level criterion of excellence.

Proper Taxonomies. Examples of the principle of proper taxonomies abound, and its implications for instruction and meaningful learning, although far from prescriptive, appear to be profound. Many science students very likely hold a pigeon-hole conception of taxonomic categories, whether rock type, volcanic cone shape, geological period, cloud form, or plate boundary. They likely encounter text materials which recount the succession of discoveries by those clever and imaginative scientists who got it right and who with perseverance overcame their obstinate, short-sighted competitors. The categories in the text may look to students like a series of sharply defined right answers—a progression of, in retrospect, inevitable discoveries, not an invented conception fruitful of inquiry. Without a good grasp of the purpose a category serves in an explanatory context, the category may appear to be arcane and arbitrary—a strange discovery, not a productive idea for solving problems. Progress in understanding what makes good science good may stumble unless students comprehend the purpose served by a concept in the context of an inquiry. The proper taxonomy criterion is one expression of this principle.

Extrapolating Systems through Time. Extrapolation of geological processes to account for past and present patterns of the earth's landscapes and structures is the explanatory goal of geological inquiry. The principles that guide reasoning toward this end are adapted to difficulties of scale and omission of clues (cf. Schumm, 1991). The most encompassing name for principles of extrapolation in geology is “uniformitarianism,” and most earth science texts introduce this idea very deliberately. However, “uniformitarianism” needs vigilant updating and careful differentiation to organize school science in keeping with criteria of excellence. Most importantly, principles of reasoning under the general heading of uniformitarianism must address difficulties of scale in time and space. Students ought to encounter problems demanding such a style of reasoning repeatedly.

Integration across Scales. Scientists have conceived of the earth as a complex, dynamic system of vast duration (Earth Systems Science Committee and National Aeronautics and Space Administration, 1986). Systems thinking links volcanism and orbital parameters with changes in global atmospheric conditions, for example. In many respects, the history of the earth is a series of impossible or unfundable experiments carried out without design or controls. Orians (1986) drew a similar conclusion about the complexity of perturbations human society imposes on the present-day ecosystems of the earth (e.g., no one would design and fund an Exxon Valdez oil spill as an experiment).

Many students in school science classes learn about the scientific method. They ought to learn as well the value of skepticism by asking how good the evidence is (Sagan, 1995). Although a stereotypical version of the scientific method may best fit the design of controlled experiments, Sagan’s evidence question applies equally well to claims about the history of the earth and the nature of earth processes—but how? There is, for example, no prospect of replicating the deposition of the Astoria fan under carefully controlled, experimental conditions. Given the obstacles to performing tests about earth systems phenomena as a whole (while admitting the value of modeling, simulation, and reduction of subsystems processes to laboratory-scale investigations), the challenge to educators is one of communicating criteria of excellence for evaluating claims about the earth without oversimplification and with acknowledgment that this context of inquiry requires proper elaboration of scientific methods.

Students ought to learn to evaluate how well scientists (including students) have selected, defined, and controlled variables and in what ways they have applied careful logic to the test of a hypothesis. When they do so, they are applying criteria of excellence to evaluate the goodness of a scientific claim based upon experiment. Gould’s second style of science demands an equally rigorous and productive set of criteria of excellence. For inquiries about the earth as a complex system vast in time and space, the criterion of integration across scales responds to this demand very successfully. It embodies the educational merits of avoiding oversimplification and promoting communication, advancing the concern that “Our children must understand ideas from earth science. . . . Nothing less than the survival of our children and our planet is at stake today” (National Center for Earth Science Education and American Geological Institute, 1991, p. 1).

Closing Comment. This article focused on the constraint of ambiguity in geological concepts and ended with a list of criteria of excellence for geological inquiry. The analysis prompts caution about making only the processes presumed common to all sciences central to curriculum construction. Unless balanced by attention to how disciplines have evolved in response to particular phenomena of interest, the representation of science as process fails to portray the rationality of scientific inquiry properly. Good science teaching depends upon resources that represent the demands characteristic of solving problems in distinctive fields of inquiry. Very suspect from this perspective are extreme examples of curriculum reform that treat science not only as a methodology undifferentiated by discipline, but also as simply one of many perspectives subsumed by a seamless interdisciplinary curriculum.

Valid understanding of any science subject predicates appreciation of the rationality characteristic of solving its particular problems. Geology is similar to but distinct from other sciences. Its distinctiveness makes it interesting, and geological thinking matches the "characteristic demands of its own intellectual problems" (Toulmin, 1990, p. 193). The same holds true for other subjects. In summary, geology is not physics. Nor is physics—with its timeless, universal, and general criteria of excellence—a better science than the local, timely, and particular claims of geology. The answer to Gowin's focus question ("What is to be taught?") is "an adequately large conception of events such that the regularities in those events can be described and judged" (Gowin, 1981, p. 196). The role of the curriculum comes to an end once the learner achieves independence in accessing, using, and *evaluating* knowledge according to the relevant criteria of excellence.

References

Ault, C.R., Jr. (1994). Research on problem solving: Earth science. In D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 269–283). New York: Macmillan.

Chamberlin, T.C. (1890). The method of multiple working hypotheses. *Science, 15,* 92–96. (Reprinted in 1965, *Science, 148,* 754–759)

Custer, S.G. (1991). Snow as a field-teaching medium for earth science. *Journal of Geological Education, 39,* 34–43.

Davis, R.A., Jr. (1983). *Depositional systems*. Englewood Cliffs, NJ: Prentice-Hall.

Earth Systems Sciences Committee (ESSC) and National Aeronautics and Space Administration (NASA). (1986). *Earth system science overview: A program for global change*. Washington, DC: NASA.

Esiobu, G.O., & Soyibo, K. (1995). Effects of concept and vee mappings under three learning modes on students' cognitive achievement in ecology and genetics. *Journal of Research in Science Teaching, 32,* 971–995.

Finley, F.N. (1981). A philosophical approach to describing science content: An example from geologic classification. *Science Education, 65,* 513–519.

Finley, F.N. (1982). An empirical determination of concepts contributing to successful performance of a science process: A study in mineral classification. *Journal of Research in Science Teaching, 19,* 689–696.

Gould, S.J. (1986). Evolution and the triumph of homology, or why history matters. *American Scientist, 74,* 60–69.

Gowin, D.B. (1981). *Educating*. Ithaca, NY: Cornell University Press.

Hanson, N.R. (1958). *Patterns of discovery.* Cambridge: Cambridge University Press.

Horgan, J. (1991). Profile: Edward Witten, the Pied Piper of superstrings. *Scientific American, 265,* 42–47.

Kitcher, P. (1993). *The advancement of science*. New York: Oxford University Press.

Kitts, D.B. (1977). *The structure of geology.* Dallas, TX: Southern Methodist University Press.

LeGrand, H.E. (1988). *Drifting continents and shifting theories*. New York: Cambridge University Press.

Lehman, J.D., Carter, C.O., & Kahle, J.B. (1985). Concept mapping, vee mapping and achievement: Results of a field study with black high school students. *Journal of Research in Science Teaching, 22,* 663–673.

Leveson, D.J. (1980). Uniformitarianism and its alternatives. *Journal of Geological Education, 28,* 244–248.

McPhee, J. (1993). *Assembling California*. New York: Noonday Press.

National Center for Earth Science Education and American Geological Institute. (1991). *Earth science education for the 21st century: A planning guide*. Alexandria, VA: Author.

National Research Council. (1996). *National science education standards*. Washington, DC: National science education standards. Washington, DC: National Academy Press.

Novak, J.D. (1990). Concept maps and vee diagrams: Two metacognitive tools to facilitate meaningful learning. *Instructional Science, 19,* 29–52.

Novak, J.D., & Gowin, D.B. (1984). *Learning how to learn*. New York: Cambridge University Press.

Okebukola, P.A. (1992). Attitude of teachers toward concept mapping and vee diagramming as metalearning tools in science and mathematics. *Educational Research, 34,* 201–213.

Oliver, J. (1983). Exploring the geological structure of the continental crust. *Journal of Geological Education, 31,* 277–294.

Orians, G.H. (1986). The place of science in environmental problem solving. *Environment, 28,* 12–41.

Rogers, R.D. (1989). Use of observational patterns in geology. *Geology, 17,* 131–134.

Romey, W.D. (1983). Winter ice and snow as models of igneous rock formation. *Journal of Geological Education, 31,* 9–12.

Roth, W., & Roychoudhury, A. (1993). Using vee and concept maps in collaborative settings: Elementary education majors construct meaning in physical science courses. *School Science and Mathematics, 93,* 237–244.

Sagan, C. (1995). *The demon-haunted world: Science as a candle in the dark*. New York: Random House.

Schwab, J.J. (1962). The concept of the structure of a discipline. *Educational Record, 43,* 197–205.

Schumm, S.A. (1991). *To interpret the earth: Ten ways to be wrong*. New York: Cambridge University Press.

Toulmin, S. (1972). *Human understanding*. Princeton, NJ: Princeton University Press.

Toulmin, S. (1990). *Cosmopolis*. New York: Macmillan.