How Mathematicians Think

How Mathematicians Think

USING AMBIGUITY, CONTRADICTION, AND PARADOX TO CREATE MATHEMATICS

*

William Byers

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press, 6 Oxford Street,
Woodstock, Oxfordshire OX20 1TW

press.princeton.edu

Sixth printing, and first paperback printing, 2010
Paperback ISBN: 978-0-691-14599-0

The Library of Congress has cataloged the cloth edition of this book as follows

Byers, William, 1943–
How mathematicians think : using ambiguity, contradiction, and
paradox to create mathematics / William Byers.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-691-12738-5 (acid-free paper)
ISBN-10: 0-691-12738-7 (acid-free paper)
1. Mathematicians—Psychology. 2. Mathematics—Psychological aspects.
3. Mathematics—Philosophy. I. Title.
BF456.N7B94 2007
510.92—dc22
2006033160

British Library Cataloging-in-Publication Data is available

This book has been composed in Palatino

Printed on acid-free paper. ∞

Printed in the United States of America

7 9 10 8 6

The point of view taken in what follows is that the experience Wiles describes is the essence of mathematics. It is of the utmost importance for mathematics, for science, and beyond that for our understanding of human beings, to develop a way of talking about mathematics that contains the entire mathematical experience, not just some formalized version of the results of that experience. It is not possible to do justice to mathematics, or to explain its importance in human culture, by separating the content of mathematical theory from the process through which that theory is developed and understood.

Different Ways of Using the Mind

Mathematics has something to teach us, all of us, whether or not we like mathematics or use it very much. This lesson has to do with thinking, the way we use our minds to draw conclusions about the world around us. When most people think about mathematics they think about the logic of mathematics. They think that mathematics is characterized by a certain mode of using the mind, a mode I shall henceforth refer to as "algorithmic." By this I mean a step-by-step, rule-based procedure for going from old truths to new ones through a process of logical reasoning. But is this really the only way that we think in mathematics? Is this the way that new mathematical truths are brought into being? Most people are not aware that there are, in fact, other ways of using the mind that are at play in mathematics. After all, where do the new ideas come from? Do they come from logic or from algorithmic processes? In mathematical research, logic is used in a most complex way, as a constraint on what is possible, as a goad to creativity, or as a kind of verification device, a way of checking whether some conjecture is valid. Nevertheless, the creativity of mathematics—the turning on of the light switch—cannot be reduced to its logical structure.

Where *does* mathematical creativity come from? This book will point toward a certain kind of situation that produces creative insights. This situation, which I call "ambiguity," also provides a mechanism for acts of creativity. The "ambiguous" could be contrasted to the "deductive," yet the two are not mutually exclusive. Strictly speaking, the "logical" should be contrasted to the "intuitive." The ambiguous situation may contain elements of the logical and the intuitive, but it is not restricted to such elements. An ambiguous situation may even involve the contradictory, but it would be wrong to say that the ambiguous is necessarily illogical.

Of course, it is not my intention to produce some sort of recipe for creativity. On the contrary, my argument is precisely that such a recipe cannot exist. This book directs our attention toward the problematic and the ambiguous because these situations so often form the contexts that produce creative insights.

Normally, the development of mathematics is reconstructed as a rational flow from assumptions to conclusions. In this reconstruction, the problematic is avoided, deleted, or at best minimized. What is radical about the approach in this book is the assertion that creativity and understanding arise out of the problematic, out of situations I am calling "ambiguous." Logic abhors the ambiguous, the paradoxical, and especially the contradictory, but the creative mathematician welcomes such problematic situations because they raise the question, "What is going on here?" Thus the problematic signals a situation that is worth investigating. The problematic is a potential source of new mathematics. How a person responds to the problematic tells you a great deal about them. Does the problematic pose a challenge or is it a threat to be avoided? It is the answer to this question, not raw intelligence, that determines who will become the successful researcher or, for that matter, the successful student.

The Importance of Talking about Mathematics

In preparing to write this introduction, I went back to reread the introductory remarks in that wonderful and influential book, *The Mathematical Experience*. I was struck by the following paragraph:

> I started to talk to other mathematicians about proof, knowledge, and reality in mathematics and I found that my situation of confused uncertainty was typical. But I also found a remarkable thirst for conversation and discussion about our private experiences and inner beliefs.

I've had the same experience. People want to talk about mathematics but they don't. They don't know how. Perhaps they don't have the language, perhaps there are other reasons. Many mathematicians usually don't talk about mathematics because talking is not their thing—their thing is the "doing" of mathematics. Educators talk about teaching mathematics but rarely about mathematics itself. Some educators, like scientists, engineers, and many other professionals who use mathematics, don't talk about mathematics because they feel that they don't possess the expertise that would be required to speak intelligently about mathematics. Thus, there is very little discussion about mathematics going on. Yet, as I shall argue below, there is a great need to think about the nature of mathematics.

What is the audience for a book that unifies the content with the "doing" of mathematics? Is it restricted to a few interested mathematicians and philosophers of science? This book is written in the conviction that what is going on in mathematics is important to a much larger group of people, in fact to everyone who is touched one way or another by the "mathematization" of modern culture. Mathematics is one of the primary ways in which modern technologically based culture understands itself and the world around it. One need only point to the digital revolution and the advent of the computer. Not only are these new technologies reshaping the world, but they are also reshaping the way in which we understand the world. And all these new technologies stand on a mathematical foundation.

Of course the "mathematization" of culture has been going on for thousands of years, at least from the times of the ancient Greeks. Mathematization involves more than just the practical uses of arithmetic, geometry, statistics, and so on. It involves what can only be called a culture, a way of looking at the world. Mathematics has had a major influence on what is meant by "truth," for example, or on the question, "What is thought?" Mathematics provides a good part of the cultural context for the worlds of science and technology. Much of that context lies not only in the explicit mathematics that is used, but also in the assumptions and worldview that mathematics brings along with it.

The importance of finding a way of talking about mathematics that is not obscured by the technical difficulty of the subject is perhaps best explained by an analogy with a similar discussion for physics and biology. Why should nonphysicists know something about quantum mechanics? The obvious reason is that this theory stands behind so much modern technology. However, there is another reason: quantum mechanics contains an implicit view of reality that is so strange, so at variance with the classical notions that have molded our intuition, that it forces us to reexamine our preconceptions. It forces us to look at the world with new eyes, so to speak, and that is always important. As we shall see, the way in which quantum mechanics makes us look at the world—a phenomenon called "complementarity"—has a great deal in common with the view of mathematics that is being proposed in these pages.

Similarly, it behooves the educated person to attempt to understand a little of modern genetics not only because it provides the basis for the biotechnology that is transforming the world, but also because it is based on a certain way of looking at human nature. This could be summarized by the phrase, "You are your DNA" or, more explicitly, "DNA is nothing less than a blueprint—or, more accurately, an algorithm or instruction manual—for building a living, breathing, thinking human being."[4] Molecular biology carries with it huge implications for our understanding of human nature. To what extent are human beings biological machines that carry their own genetic blueprints? It is vital that thoughtful people, scientists and nonscientists alike, find a way to address the metascientific questions that are implicit in these new scientific and technological advances. Otherwise society risks being carried mindlessly along on the accelerating tide of technological innovations. The question about whether a human being is mechanically determined by their blueprint of DNA has much in common with the question raised by our approach to mathematics, namely, "Is mathematical thought algorithmic?" or "Can a computer do mathematics?"

The same argument that can be made for the necessity to closely examine the assumptions of physics and molecular biology can be made for mathematics. Mathematics has given us the notion of "proof" and "algorithm." These abstract ideas have, in our age, been given a concrete technological embodiment in the form of the computer and the wave of information technology that is inundating our society today. These technological devices are having a significant impact on society at all levels. As in the case of quantum mechanics or molecular biology, it is not just the direct impact of information technology that is at issue, but also the impact of this technological revolution on our conception of human nature. How are we to think about consciousness, about creativity, about thought? Are we all biological computers with the brain as hardware and the "mind" defined to be software? Reflecting on the nature of mathematics will have a great deal to contribute to this crucial discussion.

The three areas of modern science that have been referred to above all raise questions that are interrelated. These questions involve, in one way or another, the intellectual models—metaphors if you will—that are implicit in the culture of modern science. These metaphors are at work today molding human beings' conceptions of certain fundamental human attributes. It is important to bring to conscious awareness the metascientific assumptions that are built into these models, so that people can make a reasonable assessment of these assumptions. Is a machine, even a sophisticated machine like a computer, a reasonable model for thinking about human beings? Most intelligent people hesitate even to consider these questions because they feel that the barrier of scientific expertise is too high. Thus, the argument is left to the "experts," but the fact is that the "experts" do not often stop to consider such questions for two reasons: first, they are too busy keeping up with the accelerating rate of scientific development in their field to consider "philosophical" questions; second, they are "insiders" to their fields and so have little inclination to look at their fields from the outside. In order to have a reasonable discussion about the worldview implicit in certain scientific disciplines, it would therefore be necessary to carry a dual perspective; to be inside and outside simultaneously. In the case of mathematics, this would involve assuming a perspective that arises from mathematical practice—from the actual "doing" of mathematics—as well as looking at mathematics as a whole as opposed to discussing some specific mathematical theory.

of, a creative light that I maintain often emerges out of ambiguity, on the other (this is itself an ambiguity!). My job is to demonstrate how mathematics transcends these two opposing views: to develop a picture of mathematics that includes the logical and the ambiguous, that situates itself equally in the development of vast deductive systems of the most intricate order and in the birth of the extraordinary leaps of creativity that have changed the world and our understanding of the world.

This is a book about mathematics, yet it is not your average mathematics book. Even though the book contains a great deal of mathematics, it does not systematically develop any particular mathematical subject. The subject is mathematics as a whole—its methodology and conclusions, but also its culture. The book puts forward a new vision of what mathematics is all about. It concerns itself not only with the culture of mathematics in its own right, but also with the place of mathematics in the larger scientific and general culture.

The perspective that is being developed here depends on finding the right way to think about mathematical rigor, that is, logical, deductive thought. Why is this way of thinking so attractive? In our response to reason, we are the true descendents of the Greek mathematicians and philosophers. For us, as for them, rational thought stands in contrast to a world that is all too often beset with chaos, confusion, and superstition. The "dream of reason" is the dream of order and predictability and, therefore, of the power to control the natural world. The means through which we attempt to implement that dream are mathematics, science, and technology. The desired end is the emergence of clarity and reason as organizational principles of the entire cosmos, a cosmos that of course includes the human mind. People who subscribe to this view of the world might think that it is the role of mathematics to eliminate ambiguity, contradiction, and paradox as impediments to the success of rationality. Such a view might well equate mathematics with its formal, deductive structure. This viewpoint is incomplete and simplistic. When applied to the world in general, it is mistaken and even dangerous. It is dangerous because it ignores one of the most basic aspects of human nature—in mathematics or elsewhere—our aesthetic dimension, our originality and ability to innovate. In this regard let us take note of what the famous musician, Leonard Bernstein, had to say: "ambiguity . . . is one of art's most potent aesthetic functions. The more ambiguous, the more expressive."[7] His words apply not only to music and art, but surprisingly also to science and mathematics. In mathematics, we could amend his remarks by saying, "the more ambiguous, the more potentially original and creative."

If one wishes to understand mathematics and plumb its depths, one must reevaluate one's position toward the ambiguous (as I shall define it in Chapter 1) and even the paradoxical. Understanding ambiguity and its role in mathematics will hint at a new kind of organizational principle for mathematics and science, a principle that includes classical logic but goes beyond it. This new principle will be *generative*—it will allow for the dynamic development of mathematics. As opposed to the static nature of logic with its absolute dichotomies, a generative principle will allow for the existence of mathematical creativity, be it in research or in individual acts of understanding. Thus "ambiguity" will force a reevaluation of the essence of mathematics.

Why is it important to reconsider mathematics? The reasons vary from those that are internal to the discipline itself to those that are external and affect the applications of mathematics to other fields. The internal reasons include developing a description of mathematics, a philosophy of mathematics if you will, that is consistent with mathematical practice and is not merely a set of a priori beliefs. Mathematics is a *human* activity; this is a triumph, not a constraint. As such, it is potentially accessible to just about everyone. Just as most people have the capacity to enjoy music, everyone has some capacity for mathematics appreciation. Yet most people are fearful and intimidated by mathematics. Why is that? Is it the mathematics itself that is so frightening? Or is it rather the way in which mathematics is viewed that is the problem?

Beyond the valid "internal" reasons to reconsider the nature of mathematics, even more compelling are the external reasons—the impact that mathematics has, one way or another, on just about every aspect of the modern world. Since mathematics is such a central discipline for our entire culture, reevaluating what mathematics is all about will have many implications for science and beyond, for example, for our conception of the nature of the human mind itself. Mathematics provided humanity with the ideal of reason and, therefore, a certain model of what thinking is or should be, even what a human being should be. Thus, we shall see that a close investigation of the history and practice of mathematics can tell us a great deal about issues that arise in philosophy, in education, in cognitive science, and in the sciences in general. Though I shall endeavor to remain within the boundaries of mathematics, the larger implications of what is being said will not be ignored.

Mathematics is one of the most profound creations of the human mind. For thousands of years, the content of mathematical theories seemed to tell us something profound about the nature of the natural world—something that could not be expressed in any way other than the mathematical. How many of the greatest minds in history, from Pythagoras to Galileo to Gauss to Einstein, have held that "God is a mathematician." This attitude reveals a reverence for mathematics that is occasioned by the sense that nature has a secret code that reveals her hidden order. The immediate evidence from the natural world may seem to be chaotic and without any inner regularity, but mathematics reveals that under the surface the world of nature has an unexpected simplicity—an extraordinary beauty and order. There is a mystery here that many of the great scientists have appreciated. How does mathematics, a product of the human intellect, manage to correspond so precisely to the intricacies of the natural world? What accounts for the "extraordinary effectiveness of mathematics"?

Beyond the content of mathematics, there is the *fact of mathematics*. What is mathematics? More than anything else, mathematics is a way of approaching the world that is absolutely unique. It cannot be reduced to some other subject that is more elementary in the way that it is claimed that chemistry can be reduced to physics. Mathematics is irreducible. Other subjects may use mathematics, may even be expressed in a totally mathematical form, but mathematics has no other subject that stands in relation to it in the way that it stands in relation to other subjects. Mathematics is a *way of knowing*—a unique way of knowing. When I wrote these words I intended to say "a unique *human* way of knowing." However, it now appears that human beings share a certain propensity for number with various animals.[8] One could make an argument that a tendency to see the world in a mathematical way is built into our developmental structure, hard-wired into our brains, perhaps implicit in elements of the DNA structure of our genes. Thus mathematics is one of the most basic elements of the natural world.

From its roots in our biology, human beings have developed mathematics as a vast cultural project that spans the ages and all civilizations. The nature of mathematics gives us a great deal of information, both direct and indirect, on what it means to be human. Considering mathematics in this way means looking not merely at the content of individual mathematical theories, but at mathematics as a whole. What does the nature of mathematics, viewed globally, tell us about human beings, the way they think, and the nature of the cultures they create? Of course, the latter, global point of view can only be seen clearly by virtue of the former. You can only speak about mathematics with reference to actual mathematical topics. Thus, this book contains a fair amount of actual mathematical content, some very elementary and some less so. The reader who finds some topic obscure is advised to skip it and continue reading. Every effort has been made to make this account self-contained, yet this is not a mathematics textbook—there is no systematic development of any large area of mathematics. The mathematics that is discussed is there for two reasons: first, because it is intrinsically interesting, and second, because it contributes to the discussion of the nature of mathematics in general. Thus, a subject may be introduced in one chapter and returned to in subsequent chapters.

It is not always appreciated that the story of mathematics is also a story about what it means to be human—the story of beings blessed (some might say cursed) with self-consciousness and, therefore, with the need to understand the natural world and themselves. Many people feel that such a human perspective on mathematics would demean it in some way, diminish its claim to be revealing absolute, objective truth. To anticipate the discussion in Chapter 8, I shall claim that mathematical truth exists, but is not to be found in the content of any particular theorem or set of theorems. The intuition that mathematics accesses the truth is correct, but not in the manner that it is usually understood. The truth is to be found more in the fact than in the content of mathematics. Thus it is consistent, in my view, to talk simultaneously about the truth of mathematics and about its contingency.

The truth of mathematics is to be found in its human dimension, not by avoiding this dimension. This human story involves people who find a way to transcend their limitations, about people who dare to do what appears to be impossible and *is* impossible by any reasonable standard. The impossible is rendered possible through acts of genius—this is the very definition of an act of genius, and mathematics boasts genius in abundance. In the aftermath of these acts of genius, what was once considered impossible is now so simple and obvious that we teach it to children in school. In this manner, and in many others, mathematics is a window on the human condition. As such, it is not reserved for the initiated, but is accessible to all those who have a fascination with exploring the common human potential.

We do not have to look very far to see the importance of mathematics in practically every aspect of contemporary life. To begin with, mathematics is the language of much of science. This statement has a double meaning. The normal meaning is that the natural world contains patterns or regularities that we call scientific laws and mathematics is a convenient language in which to express these laws. This would give mathematics a descriptive and predictive role. And yet, to many, there seems to be something deeper going on with respect to what has been called "the unreasonable effectiveness of mathematics in the natural sciences."[9] Certain of the basic constructs of science cannot, in principle, be separated from their mathematical formulation. An electron *is* its mathematical description via the Schrödinger equation. In this sense, we cannot see any deeper than the mathematics. This latter view is close to the one that holds that there exists a mathematical, Platonic substratum to the real world. We cannot get closer to reality than mathematics because the mathematical level *is* the deepest level of the real. It is this deeper level that has been alluded to by the brilliant thinkers that were mentioned above. This deeper level was also what I meant by calling mathematics irreducible.

Our contemporary civilization has been built upon a mathematical foundation. Computers, the Internet, CDs, and DVDs are all aspects of a digital revolution that is reshaping the world. All these technologies involve representing the things we see arise out of situations of ambiguity. Of course the creative process is intimately tied to the birth and the processing of mathematical ideas. Thus thinking about ideas as the fundamental building blocks of mathematics (as opposed to the logical structure, for example) pushes us toward a reevaluation of just what mathematics is all about. This section demonstrates that even something as problematic as a paradox can be the source of a productive idea. Furthermore, I go on to claim that some of the most profound ideas in mathematics arise out of situations that are characterized not by logical harmony but by a form of extreme conflict. I call the ideas that emerge out of these extreme situations "great ideas," and a good deal of the book involves a discussion of such seminal ideas.

The third section, "The Light and the Eye of the Beholder," considers the implications of the point of view that has been built up in the first two sections. One chapter is devoted to a discussion of the nature of mathematical truth. Is mathematics absolutely true in some objective sense? For that matter, what do we mean by "objectivity" in mathematics? Thinkers of every age have attested to the mystery that lies at the heart of the relationship between mathematics and truth. My "ambiguous" approach leads me to look at this mystery from a perspective that is a little unusual. Finally, I spend a concluding chapter discussing the fascinating and essential question of whether the computer is a reasonable model for the kind of mathematical activity that I have discussed in the book. Is mathematical thought algorithmic in nature? Is the mind of the mathematician a kind of software that is implemented on the hardware that we call the brain? Or is mathematical activity built on a fundamental and irreducible human creativity—a creativity that comes from a deep need that we human beings have to understand—to create meaning out of our lives and our environment? This drive for meaning is inevitably accompanied by conflict and struggle, the very ingredients that we shall find in situations of ambiguity.

SECTION I

THE LIGHT OF AMBIGUITY

WHAT IS THINKING? If we imagine thinking to be an ordered, linear, and logical progression, then the rigorous thinking that one finds in a mathematical proof or a computer program is the highest form of thinking. Is this the only way to think? More to the point, is this the way mathematicians think? In this section I investigate situations that seem to be at the opposite extreme from logical thought—I look for ambiguities in mathematics. Strangely enough, I find ambiguity everywhere, and not only ambiguity but also its close cousins contradiction and paradox. How strange it is that mathematics, the subject that appears to be the very paradigm of reason, and for this reason the model that other disciplines attempt to emulate, contains as an irreducible factor, the very things that reason ostensibly exists to eliminate from human discourse!

Ambiguity is not only present in mathematics, it is essential. Ambiguity, which implies the existence of multiple, conflicting frames of reference, is the environment that gives rise to new mathematical ideas. The creativity of mathematics does not come out of algorithmic thought; algorithms are born out of acts of creativity, and at the heart of a creative insight there is often a conflict—something problematic that doesn't follow from one's previous understanding. Now one might think that mathematics is characterized by the clarity and precision of its ideas and, therefore, that there is only one correct way to understand a given mathematical situation or concept. On the contrary, I maintain that what characterizes important ideas is precisely that they can be understood in multiple ways; this is the way to measure the richness of the idea.

Ambiguity is the central theme of this book. From beginning to end it is the single thread that unites the disparate subjects that are discussed. We each probably feel that we understand and are familiar with ambiguity. However, in our exploration of ambiguity in mathematics we may find that there is more to ambiguity than meets the eye. Ambiguity is very rich, and so each new aspect of ambiguity we encounter will teach us something not only about mathematics but also about the nature of ambiguity itself—at least about the way in which ambiguity is being used in this book. Since the whole book is, in a sense, a development of the meaning of "ambiguity in mathematics," I ask the reader not to prematurely close accounts with ambiguity.

and I never have been. I'm interested in *understanding*, which is quite a different thing."[3] Now, understanding is a difficult thing to talk about. For one thing, it contains a subjective element, whereas drawing logical inferences appears to be an objective task that even sophisticated machines might be capable of making. Nevertheless, if one wants to come close to plumbing the depths of mathematical practice, it will be necessary to begin by seeing beyond the formalist approach of equating mathematics with the trinity of definition-theorem-proof.

Logic is indispensable to mathematics. For one thing, logic stabilizes the world of mathematical results so that it presents itself to our minds in the conventional manner—as a body of permanent and absolute truths. However, logic is not the essence of mathematics nor can mathematics be reduced to logic. Mathematics transcends logic. Mathematics is one of the most profound areas of human creativity. Yet the statement that mathematics goes beyond logic needs to be supported. To do this, a number of characteristics of mathematics will be introduced that are clearly not derived from logic. These include a certain form of mathematical ambiguity as well as the related notions of contradiction and paradox.

"Ambiguity" is a central notion, so I shall spend a fair amount of time in explaining what I mean by ambiguity in mathematics. By ranging over a whole host of examples from mathematics and a few from other fields, I hope to show that ambiguity, as I use the term, is a phenomenon which is central to mathematical theory and practice. Ambiguity will give us a way to approach such questions as "What is the relationship between logic and mathematics?" "What is the nature of creativity in mathematics?" "What is meant by understanding in mathematics and what is its relationship to creativity?" Even the old chestnuts, "Is mathematics invented or discovered?" or "What accounts for the 'unreasonable effectiveness' of mathematics in the physical sciences?" Ambiguity will transform the mathematical landscape from the static to one that is dynamic and characterized by the play of ideas.

What I am attempting to develop is nothing less than a paradigm shift in our understanding of the nature of the mathematical enterprise. Once we begin to look at matters in this new "ambiguous" manner, many things suddenly appear in a new light.

These certainly include mathematical practice and the teaching and learning of mathematics. But this manner of looking at things has implications for how we view the scientific enterprise as a whole. These implications extend to the most fundamental of questions, such as "What is (mathematical) truth?" and "What is knowledge?"

With these heady reflections in the back of our minds, I now proceed to take up the basic notion of the meaning of ambiguity (for this book) and proceed to demonstrate its role in mathematics.

What Do I Mean by Ambiguity?

In this book, ambiguity is a key idea whose implications will take some time and effort to flesh out. For me the most elementary mathematical object, like the equation "1 + 1 = 2," for example, is ambiguous. What do I mean by this? I certainly do *not* mean that the statement "1 + 1 = 2" is unclear or incorrect. People often take ambiguity to be synonymous with vagueness or with incomprehensibility. Though this is a possible meaning, it is not the sense in which I shall use the term. What I am trying to accomplish by using the word ambiguous is to point to a certain metaphoric quality that is inherent in even quite simple mathematical situations. When we encounter "1 + 1 = 2," our first reaction is that the statement is clear and precise. We feel that we understand it completely and that there is nothing further to be said. But is that really true? The numbers "one" and "two" are in fact extremely deep and important ideas, as will be discussed in Chapter 5. They are basic to science and religion, to perception and cognition. "One" represents unity; "two" represents duality. What could be more fundamental? The equation also contains an equal sign. Again, in Chapter 5, I discuss various ways in which "equality" can be understood in mathematics. Equality is another very basic idea whose meaning only grows the more you think about it. Then we have the equation itself, which states that the fundamental concepts of unity and duality have a relationship with one another that we represent by "equality"—that there is unity in duality and duality in unity. This deeper structure that is implicit in the equation is typical of a situation of ambiguity. Thus even the most elementary mathematical expressions have a profundity that may not be apparent on the surface level. This profundity is directly related to what I am calling ambiguity.

The word "ambiguity" is actually being used for two main reasons. The first is that the ambiguous is commonly looked at as something to be contrasted with the logical. The second comes from one of the Oxford English Dictionary definitions of "ambiguity"—"admitting more than one interpretation or explanation: having a double meaning or reference." This notion of "double meaning" comes from the prefix "ambi," as can be seen in such words as "ambidextrous" or "ambivalent." However, the definition that I now put forward comes from a definition of creativity that was proposed by the writer Arthur Koestler.[4] He said that creativity arises in a situation where "a single situation or idea is perceived in two self-consistent but mutually incompatible frames of reference." I shall take the above to be the definition of ambiguity. To repeat:

> *Ambiguity involves a single situation or idea that is perceived in two self-consistent but mutually incompatible frames of reference.*

I hasten to add that putting such a precise definition at the beginning of Chapter 1 involves the risk that the reader will assume that ambiguity is now pinned down once and for all. On the contrary, ambiguity is one of those concepts, like "one," "two," and "equality," of which there is always more to say and learn. I am even tempted to say that "ambiguity" is not really a concept at all; it is more like a condition or context that *produces* concepts. If it is not a normal concept, how then do I go about describing it? My strategy is to start with the description above and give it substance by presenting a series of examples each of which will explore some dimension of ambiguity.

This book is an exploration of ambiguity in mathematics. Unfortunately mathematics is usually presented in a linear manner with the simple preceding the complex and assumptions before conclusions. I prefer to think both of mathematics and of this book as explorations. What is the nature of an exploration in mathematics? In the introduction to his textbook, *Transform Linear Algebra*, Frank Uhlig states:

> Linear Algebra is a circular subject. Studying Linear Algebra feels like exploring a city or a country for the first time. An overwhelming number of concepts, all intertwined and connected, are present in any first encounter with linear algebra. As with a new city, one has to start discovering slowly and deliberately. Of great help is that linear algebra is akin to geometry, and like geometry, many of its insights have been permanently there within us. We must only explore, look around, and awake our intuition with the reality of this mathematical place.[5]

What a poetic evocation of the spirit of learning and doing mathematics! I'm inviting the reader to enter into an exploration of mathematics in just this spirit. I shall look at mathematics through the lens of ambiguity. In so doing we shall be simultaneously investigating the nature of ambiguity itself. As Uhlig says, many of the basic insights are already there within us, but to discern them we shall have to put aside our habitual point of view and be open to considering a new viewpoint.

Self-Consistency, Incompatibility, and Creativity

The definition of ambiguity that I gave above involves a duality—there must be two frames of reference. Now, duality is a familiar idea in mathematics. For example, in projective geometry it is possible to interchange "points" and "lines" so that every statement about lines and points has a dual statement about points and lines. The statement, "Two lines define (meet at) a point" would have the dual statement, "Two points define (determine) a line." This kind of structural duality carries some, but not all of the meaning that I attribute to ambiguity.

Ambiguity, as the term is being used here, is not mere duality. The two frames of reference must be *mutually incompatible*, even though they are individually self-consistent. Yet, in spite of this incompatibility, there exists an over-riding unitary situation or idea. On the one hand, there is the harmony of consistency—things are in peaceful equilibrium. On the other, there is the disorder of incompatibility. Incompatibility is unacceptable in mathematics! It must be resolved! It is this need to resolve incompatibility that makes the situation of ambiguity so dynamic, so potentially creative. There are two perfectly harmonious ways of looking at the situation, yet they are in opposition to one another. So there is a need to resolve this unacceptable situation in order to restore equilibrium. The restoration of equilibrium can only come at a level that is, in a manner of speaking, higher than either of the original frames of reference. The equilibrium condition may not yet exist. It may only come into existence as a result of the need to reconcile the incompatibility of the original situation. Thus, a situation of ambiguity is a situation with creative possibilities.

Ambiguity may seem to be complicated, but its essence can be conveyed very simply. Here is an example of ambiguity. It's a joke—not very funny but with a mathematical connection—and it makes the point about the nature of ambiguity.

> A mathematician is flying non-stop from Edmonton to Frankfurt with Air Transat. The scheduled flying time is nine hours. Some time after taking off, the pilot announces that one engine had to be turned off due to mechanical failure: "Don't worry—we're safe. The only noticeable effect this will have for us is that our total flying time will be ten hours instead of nine." A few hours into the flight, the pilot informs the passengers that another engine had to be turned off due to mechanical failure: "But don't worry—we're still safe. Only our flying time will go up to twelve hours." Some time later, a third engine fails and has to be turned off. But the pilot reassures the passengers: "Don't worry—even with one engine, we're still perfectly safe. It just means that it will take sixteen hours total for this plane to arrive in Frankfurt." The mathematician remarks to his fellow passengers: "If the last engine breaks down, too, then we'll be in the air for twenty-four hours altogether!"

Here you have it—two conflicting frames of reference (one of them implicit) resulting in tension, and then a creative release, laughter. Of course in mathematics the release comes with the birth of a new idea or a new way of looking at the situation but the dynamics of a humorous situation is very similar. A joke is an example of ambiguity and creativity—you have to *get* a joke.

language uses verbs. Thus the dichotomy between matter and energy is built into language itself.[8]

How is the gap between these two to be bridged? The first and most obvious way would be to regard matter and energy as complementary. Thus one could regard matter and energy as indispensable aspects of the natural world and maintain that a complete description of nature would involve describing both domains and the laws that govern them. We would go on to describe the relationship between matter and energy. Thus a moving body possesses kinetic energy that is proportional to its mass and the square of its velocity. To look at things in this way would be to miss the radical insight behind Einstein's equation. $E = mc^2$ says that matter *is* energy. It says that these two mutually exclusive ways of describing reality are in fact one—that there is one reality that can be seen as energy when we look at it in one context and as matter when we look at it in another.

Thus, the equation is something that could be called a scientific metaphor. A literary metaphor like Shakespeare's "all the world's a stage, and all the men and women merely players" is a comparison between two different domains—it is really a kind of mapping from one of these domains, here ordinary life, to the other, here the stage. However, a metaphor requires more than a mere correspondence between different domains. "Getting" a metaphor requires an insight: it requires looking at the world in a new way. The power of this particular insight is extraordinary. Its consequence, the atomic bomb, is itself a metaphor for the power of the idea. This equation brings out the full implication of "ambiguity" as the term is being used here. There exist two frames of reference whose incompatibility generates enormous power. This power is then harnessed by the single idea that is represented by the equation $E = mc^2$.

Ambiguous Situations in Mathematics

Now let us move on to a more systematic exploration of ambiguity in mathematics. There will be a place for some fairly sophisticated mathematics, but I will begin with a number of elementary examples, very elementary indeed. The reason for including these examples is that they are accessible to everyone. Also, they are here to make the point that no mathematics is completely "trivial." Even elementary arithmetic, algebra, and geometry, when looked at from a fresh perspective, can manage to surprise you.

The Equations of Arithmetic

Let's return to the most basic of equations from arithmetic, something like "2 + 3 = 5." Where is the ambiguity here? I remember the way equations were explained in grade school through the metaphor of the balance. If you put a two- and a three-pound weight on one side of a scale and a five-pound weight on the other, then the two sides will balance. Equality, we were told, means balance. Now "balance" is a good way to think of equality, but is it the only way? From the balance metaphor we derive the idea that "2 + 3" and "5" are just two different ways to describe the same thing—that "2 + 3" and "5" are essentially identical and that the equality sign represents this identity. However "=" does not mean identical, as Bodanis pointed out in the paragraphs I quoted above. Thinking of equations as merely linking two otherwise identical quantities would not explain the power of equations to open up unsuspected relationships between things that were not necessarily connected a priori.

Where is the creative element in "2 + 3 = 5"? Where is the insight, the possibility for an aha! experience? In order to appreciate what is going on, we may have to listen to intelligent people who are less sophisticated than we are—children, for example. Various researchers in mathematics education (e.g., Kieren 1981) have pointed out children's propensity to understand the equality sign in operational terms; that is, "2 + 3 = 5" is understood as an action "2 added to 3 *makes* 5." The sum "2 + 3" is a process, a verb. Children learn what addition is about through the process of counting. Yet the right-hand side is an object, the number 5. What the equation "2 + 3 = 5" is doing is identifying a process with an object. This is similar to the moral of the Thurston story above, where the process of division was seen as a numerical object. To see that a process can be an object or, looking at it the other way around, that the object can be thought of as a process, entails a discontinuous leap—an act of understanding that is in essence a creative act. We all made this creative leap so long ago that we don't remember having done so. But it was an essential step in our development. And what was the essence of this act of understanding? It is that process and object are one ambiguous idea. Thus the ambiguity here is seeing that the two contexts of process and object are unified by this one idea that is captured symbolically by the equation "2 + 3 = 5." All the elements of ambiguity are present here: the two contexts that are in conflict until the conflict is resolved by an act of understanding. Subsequent to the act of understanding, what used to be a conflict becomes a flexible viewpoint where one is free to freely move between the contexts of number as object and number as process. I will return to this same ambiguity in a less elementary situation when I come to discuss infinite decimals.

The Square Root of Two

To our contemporary way of understanding things the square root of 2 is no mystery. It is a perfectly well-defined number. In what way, then, can $\sqrt{2}$ be called ambiguous? By our definition, ambiguity required "a single situation or idea"—precisely the fact that $\sqrt{2}$ is well defined. But it also required that $\sqrt{2}$ can be perceived in two self-consistent contexts which are somehow in conflict with one another.

This latter requirement can best be understood historically. In fact $\sqrt{2}$ has an interesting history. It appears, in Euclidean geometry, as a consequence of the Theorem of Pythagoras, as the length of the hypotenuse of a right-angled triangle with sides of unit length.

Thus, $\sqrt{2}$ existed for the Greeks as a concrete geometric object. On the other hand, they were able to prove that this (geometric number) was not rational, that is, it could not be expressed as the ratio of two integers, like 2/3 or 127/369. Such nonfractions came to be called irrational numbers, and the name "irrational" indicates the kind of emotional reaction that the demonstration of the existence of nonrational numbers produced.

Figure 1.1. $\sqrt{2}$ as a geometric object

There is no question that the demonstration that $\sqrt{2}$ was not rational precipitated a crisis.[9] I shall return to this crisis in a later chapter, but for now let's just say that a whole way of looking at the world, the philosophy of the secret Pythagorean society, was brought into question. Today we would say that $\sqrt{2}$ is an irrational number, but "an irrational number is no number at all . . . it is totally man-made," as Leopold Kronecker said, "and thus is of dubious significance philosophically."[10] So is $\sqrt{2}$ a number or not? We can all agree that it is a very different kind of number from the integers and the fractions, the numbers of arithmetic.

William Dunham makes a comment that is relevant here when he says that the irrationality of $\sqrt{2}$ is one instance of "a continuous feature of the history of mathematics . . . the prevailing tension between the geometric and the arithmetic."[11] There are two primordial sources of mathematics: counting, which leads to arithmetic and algebra, and measuring, which leads to geometry. Two self-consistent contexts, if you will. Initially these two domains were considered to be identical, but the $\sqrt{2}$ proof brought an inherent conflict between them out into the open. Rational numbers have a consistent meaning in both contexts, but in $\sqrt{2}$ we have a mathematical object that has a clear meaning in a geometric context but is problematic when considered as an arithmetic object, in this case a rational number. A number is a number is a number, to paraphrase Gertrude Stein, but is a geometric number really a number? At the very least there is a tension, an incompatibility, between the geometric and the arithmetic. It is this incompatibility that made $\sqrt{2}$ ambiguous for the Pythagoreans. This does not mean that it was viewed as vague or imprecise. The term ambiguity highlights the problematic aspects of $\sqrt{2}$ for the Greeks.

There are two possible reactions to this sort of ambiguous situation. One can abandon one of the seemingly inconsistent contexts or one can build a new context that is general enough to reconcile the conflict. Both reactions are interesting and can lead to new mathematics.[12] The Greeks chose the former and essentially abandoned algebra for geometry. Even so, the irrationality of $\sqrt{2}$ was a great blow to those, like the Pythagoreans, whose entire worldview was based on the rationality (in the sense of rational numbers or fractions) of the natural world (see Chapter 7). In fact, large portions of Euclidian geometry (the books on ratio and proportion) had been developed on the assumption that any two lengths are commensurable. That means that for any two line segments there is a (smaller) segment that divides into both segments evenly. This amounts to saying that the ratio of the lengths of the two segments is rational. Thus all these proofs (and also, it has been conjectured, the proof of the Pythagorean Theorem itself) that depended on this assumption had to be redone in a different way. This task was, in fact, accomplished successfully. In this activity one can see the need to resolve the incompatibility raised by the ambiguity and therefore the role of the ambiguity as a generator of mathematical activity.

It might be interesting to take a moment and discuss the theory of ratio and proportion. A ratio is the quotient of two numbers. Let's call them x and y. Today we would say that the ratio is the quotient, the number x/y. However the Greeks did not do this—in fact human beings did not do this for the next two millennia. The problem is in a way the same problem Thurston had with 134/29—is the *process* of division the same as the number that *results* from that process? In the case of commensurable numbers, $x = nz$ and $y = mz$ for some integers m and n, then x/y can be identified with n/m, but what does the quotient mean when x and y are incommensurable, like 1 and $\sqrt{2}$, for example? How can you call this kind of ratio a number? This was the major problem that necessitated a complete reworking of the

$1 = .999. . . .$

Now what is the meaning of these equations? What is the precise meaning of the "=" sign? It surely does not mean that the number 1 is identical to that which is meant by the notation .999. . . . There is a problem here, and the evidence is that, in my experience, most undergraduate math majors do not believe this statement. I remember putting this question, "does 1 = .999. . .?" to the students in a class on real analysis. Something about this expression made them nervous. They were not prepared to say that .999. . . is equal to 1, but they all agreed that it was "very close" to 1. How close? Some even said "infinitely close," but they were not absolutely sure what they meant by this. These students may be quite advanced in certain ways, but this statement is still an obstacle[14] for them. What is the obstacle? In my opinion it is the ambiguity contained in equating an infinite decimal to an integer.

The notation .999. . . stands for an infinite sum. Thus

$$.999\ldots = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \ldots$$

Now an infinite sum is a little more complicated than a finite sum, and this complexity is revealed by the fact that the notation is deliberately ambiguous. Thus this notation stands both for the *process* of adding this particular infinite sequence of fractions and for the *object*, the number that is the result of that process. As was the case of the equations of arithmetic, the two contexts (in the above definition of ambiguity) are again those of process and object. Now the number 1 is clearly a mathematical object, a number. Thus the equation 1 = .999. . . is confusing because it seems to say that a process is equal (identical?) to an object. This appears to be a category error. How can a process, a verb, be equal to an object, a noun. Verbs and nouns are "incompatible contexts" and thus the equation is ambiguous. Similarly, all infinite decimals are ambiguous. Students have a problem because they think of .999. . . only as a process. They imagine themselves actually adding up the series term by term and they "see" that this process never ends. So at any finite stage the sum is "very close" but not equal to 1. They don't see that this infinite process can be understood as a single number.

You can even go through a "proof" with them, something like:

Let $x = .999\ldots$.
Then $10x = 9.999\ldots$ (shifting the decimal point).
Thus $9x = 10x - x = 9.999\ldots - .999\ldots = 9$.
So $x = 1$.

The reaction is interesting. For the most part, the students will now agree that 1 is indeed equal to .999. . . . That is, they now accept it but, in my opinion, something of the old perplexity still remains. They have not resolved the ambiguity. They still do not "understand" the representation for infinite decimals. Understanding requires more than accepting the validity of a certain argument. It requires a creative act, which is what I mean when I refer to the resolution of an ambiguity.

I hasten to add that this ambiguity is a strength, not a weakness, of our way of writing decimals. To understand infinite decimals means to be able to move freely from one of these points of view to the other. That is, understanding involves the realization that there is "one single idea" that can be expressed as 1 or as .999. . ., that can be understood as the process of summing an infinite series or an endless process of successive approximation as well as a concrete object, a number. This kind of creative leap is required before one can say that one understands a real number as an infinite decimal.

Variables

One of the most basic aspects of mathematics involves the reduction of the infinite to the finite. Mathematics has been called the science of the infinite, yet mathematicians are human beings and therefore intrinsically restricted to the finite. Thus one of the great mysteries of mathematics is the manner in which the process of making the infinite finite occurs. This question will be examined in great detail in the discussion of infinity. For the moment, consider the notion of "variable." Most people are introduced to the idea of variable in high school algebra, where they learn to manipulate expressions such as "$3x + 2$." They are told that the "x" is not a number but can represent *any* number. In fact x is usually restricted to some particular set of numbers: natural numbers, integers, rational numbers, real or complex numbers. It may even be a subset of one of these sets of numbers. The domain of the variable may not even be specified explicitly but only inferred by the context. In this sense the notion of a variable is a little ambiguous.

However there is another and more serious way in which the idea of a variable is ambiguous. Let us suppose that we are talking about the positive integers. Then the expression $3x + 2$ actually stands for the whole set of numbers: $3(1) + 2 = 5$, $3(2) + 2 = 8$, $3(3) + 2 = 11$, 14, 17, 20, ... So $3x + 2$ is a short-hand for the whole set of numbers $\{5, 8, 11, \ldots\}$. However when we work with the expression $3x + 2$ we do not carry around the whole set of potential values in our head. We think of $3x + 2$ as some specific but unspecified element of that set. So we imagine x to have been chosen. It is some (one) specific number that can be written as $3x + 2$, but we know nothing about the value of x except that it is an integer. Thus we simultaneously think of x as general and specific. It is precisely this general/specific ambiguity that gives the notion of variable its importance in mathematics. An infinite set of possible values has been replaced by a finite set of values (here one value). It is true this one value is unspecified, but nevertheless something has been gained.

For example, consider the equation

$$3x + 2 = 8$$

and its solution

$$x = 2.$$

Does the "x" in "$3x + 2 = 8$" refer to *any* number or does it refer to the number 2? The answer is both and neither. At the beginning x could be anything. At the end x can only be 2. Of course at the beginning x (implicitly) can only be 2. Yet at the end we are saying that every number $x \neq 2$ is *not* a solution, so the equation is also about all numbers. Thus at every stage the x stands for *all* numbers but also for the *specific* number 2. We are required to carry along this ambiguity throughout the entire procedure of solving the equation. It begins with something that could be anything and ends with a specific number that could not be anything else. What an exercise in subtle mental gymnastics this is! How could this way of thinking be called *merely* mechanical? No wonder children have difficulty with algebra. The difficulty is the ambiguity. The resolution of the ambiguity, solving the equation, does not involve eliminating the double context but rather being able to keep the two contexts simultaneously in mind and working within that double context, jumping from one point of view to the other as the situation warrants.

A variable is general and specific at the same time. It is all values or it is a unspecified "typical" value. In that ambiguity lies its power. By not resolving the ambiguity until the end of the piece of mathematics one is able to use that ambiguity constructively. Thus when considering the function $f(x) = 3x^2 + 2$, we think of x as a typical real number. But we also think of the whole function as being identified with its parabolic graph. Then we can say that its derivative, for example, is the function $6x$. Again, we think of this in two ways: first, as a formula that is valid for all values of "x" (the derivative at $x = 2$ is 6 times 2 or 12); and second, as a specific (single) point on the graph where the slope of the tangent line is the specific number $6x$.

Without this double or ambiguous point of view, modern mathematics would never have been invented. Remember that Greek mathematics was geometric and not algebraic. Algebraic thought requires the use of the idea of variable. This was not as explicit in Greek thought as it would later become. Again, we can only speculate that it was the Greeks' reverence for clarity and harmony and their distrust and repugnance for ambiguity that prevented them from developing their mathematics in this direction.

The algebraic equation $3x + 2 = 8$ is ambiguous in yet another way. In solving this equation I am really making the following assertion: "Assuming that there exists a number x such that $3x + 2 = 8$, it follows that this number must be 2." Thus in setting out to solve an equation we have taken for granted that the solution exists. That is, the solution is both unknown and (implicitly) known at the same time. This ambiguity between being known and unknown is similar to the ambiguity of a variable that I mentioned earlier and is essential to equations.

I said above (with respect to the equation $3x + 2 = 8$) that we start by assuming that the solution exists and only then determine what it is. What if the solution does not exist? What hap-

Figure 1.2. Graph of "$f(x) = 3x^2 + 2$"

pens to the ambiguity? Consider the equation $x + 1 = 0$, and assume that there are only the nonnegative integers (0, 1, 2, 3, . . .) at our disposal. Then the solution $x = -1$ is not available, so there is no solution within the system we are working in. What happens? Remember that I said that ambiguous situations were dynamic; the two incompatible contexts may generate their own creative resolution. Here the incompatibility resides in the fact that the equation is a form that implicitly assumes a solution exists (all the terms in the equation belong to the known system of nonnegative integers), yet no solutions exist (within the system of nonnegative integers). The creative resolution of this dilemma *generates the required solutions*. You could say that the equation $x + 1 = 0$ brings the negative numbers into existence! Thus in order to give meaning to the equation $x + 1 = 0$ (and more generally $x + n = 0$) in a situation where only the nonnegative integers are available, we are forced to invent a new class of numbers. How exciting! Similarly the equation $x^2 + 1 = 0$ produces the complex numbers.[15]

There is power in this ambiguity even if the existence of the solution is not guaranteed. In fact, in this case we can see the generative power of ambiguity to creatively produce new ideas.

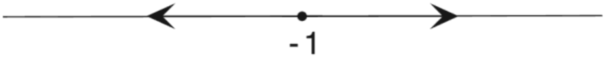

Figure 1.4. "$f(x) = 2x + 1$" generates a dynamical system

equation or dynamical system. Suppose the "generating function" is $f(x) = 2x + 1$. If the process has the value $x_0 = 1$ at time 0, then its value at time 1 would be $x_1 = f(x_0) = 2(1) + 1 = 3$, its value at time 2 would be $x_2 = f(x_1) = 7$, and so on. Thus the function could be considered as a "law" that governs how the process it represents evolves over time. The graph of the function is of little help to us when we think of a function in this way. What we want to know is, if we start with a certain value at time zero, what will happen "in the long run." From this point of view a completely different geometric picture is required. For the generating function $f(x) = 2x + 1$ it would look as shown in figure 1.4.

The picture contains the following information: the "dot" at -1 means that if the system starts at value $x = -1$, then it remains at this value at all future times; the right arrow means that if the initial value is greater than -1 then the future values of the process increase without bound; the left arrow that they decrease without bound if the initial value is less than -1.

So we now have a new way of thinking about the concept of function—a dynamic concept as compared with the versions of functions that were discussed earlier. We now need to reconcile these two definitions by integrating them into a new and more general concept.

At this stage we are thinking about a function as some sort of rule that applies to a whole set of numbers. However, there inevitably comes a time when we want to operate not just on individual numbers but on the rule itself. Thus, if $f(x) = x^2$ and $g(x) = 3x + 1$, we may wish to add $f(x)$ and $g(x)$ and thereby create a new function $h(x) = f(x) + g(x) = x^2 + 3x + 1$. Or we may wish to multiply them and create $k(x) = x^2(3x + 1) = 3x^3 + x^2$. Or again we may wish to consider the result of applying the first rule followed by the second rule $h(x) = g(f(x)) = 3x^2 + 1$. This last operation is called the "composition of functions f and g" and written $h = g \circ f$. It can also be done in the reverse order to obtain $f(g(x)) = (3x + 1)^2$. When we operate on functions in this way we are thinking of the function as one whole, unified object. We have made a function that began as something akin to a process, a process for operating on numbers, into an object that itself could be operated upon. At one level, one can add or multiply numbers; now, at a higher level, one can add, multiply, or compose functions. This is the process of abstraction at work. Abstraction consists essentially in the creation and utilization of ambiguity. The initial barrier to understanding, that a function can be considered simultaneously as process and object—as a rule that operates on numbers and as an object that is itself operated on by other processes—turns into the insight. That is, it is precisely the ambiguous way in which a function is viewed which is the insight.

At a higher level of abstraction one puts whole families of functions together to form function spaces, for example, all continuous functions defined on the interval of numbers between 0 and 1. Once a function is seen as a point in a larger space, we can talk about the distance between functions, the convergence of functions, functions of functions, and so on. This sort of dual representation is present in a great many mathematical situations.

Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus is one of the great theorems of mathematics. A consideration of this theorem will extend our discussion of ambiguity from the domain of concepts like variables and functions to include the domain of actual mathematical results. How, one might ask, can a mathematical theorem be ambiguous? The *essence* of this theorem is ambiguity; it is asserting that calculus is ambiguous!

Now "differential calculus" and "integral calculus" can be (and historically were) developed independently of one another. They appear, at first glance, to have nothing to do with one another. Integration is a generalization of the idea of area. A typical problem might be to calculate the area between the graph of the curve $y = x^2$ and the x-axis, between 0 and 1 (figure 1.5a). Differential calculus as developed by Newton and Leibniz was concerned with calculating the slope of tangent lines to curves or the related problem of instantaneous change in one variable with respect to another, velocity, for example, as shown in figure

Figure 1.5a. Area: $\int_0^1 x^2\,dx = 1/3$

Figure 1.5b. Slope of tangent: $(d/dx)(x^2)\,|_{x=1} = 2$

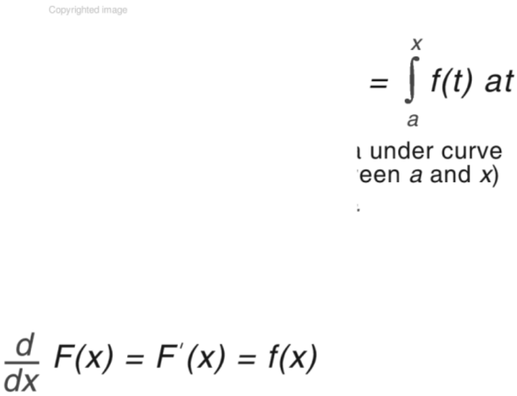

Figure 1.6. Fundamental theorem of calculus

1.5b. The Fundamental Theorem says that these processes are inverses of one another (when the functions involved are "reasonable" as in figure 1.6).

Now it may be possible to start with integration and then develop differentiation or vice versa, but the theorem says that, for functions of one variable, neither process is the more fundamental. Actually, the theorem says that there is in fact one process in calculus that is integration when it is looked at it in one way and differentiation when it is looked at in another. Another way of putting this is that without the Fundamental Theorem there would be two subjects: differential calculus and integral calculus. With it there is just the calculus, albeit with a multiple perspective. This multiple perspective is essential to an understanding of calculus.

How is this multiple perspective used? Well, since differentiating is easier than integrating, we can integrate by taking the inverse of the derivative, that is, by calculating the antiderivative. For example, since the derivative of the function $f(x) = x^2$ is the function $g(x) = 2x$, it follows that the integral of $2x$ is x^2. Whole lists of such antiderivatives may be established and then used to integrate elementary functions.

Fermat's Last Theorem

Perhaps the most famous mathematical problem in the last three hundred years involves the equation,

$$x^n + y^n = z^n$$

for $n \geq 2$. For $n = 2$ this equation represents the relationship between the lengths of the sides of a right-angled triangle according to the theorem of Pythagoras. Thus there exist many sets of solutions, including $x = 3$, $y = 4$, $z = 5$ or $x = 5$, $y = 12$, $z = 13$. The mathematician Pierre de Fermat (1601–1665) claimed that there were no integer solutions to this equation for $n > 2$ and moreover that that he had a "marvelous proof of this." Unfortunately the proof he was thinking of was never found. Building on the work of many talented mathematicians before him, the correct argument was finally obtained by Andrew Wiles in 1993.[20] It was a triumph of human ingenuity and creativity, and the entire story of the work on this conjecture makes fascinating reading for anyone who is interested in mathematics.

The proof hinges on the validity of a conjecture called the Taniyama-Shimura conjecture. This conjecture unifies the seemingly disparate worlds of elliptic curves and modular forms. To understand the power of ambiguity to revolutionize mathematics, one has but to read the comments on this conjecture by the Harvard mathematician Barry Mazur. He compared the conjecture to the Rosetta stone that contained Egyptian demotic, ancient Greek, and hieroglyphics. Because demotic and Greek were already understood, archaeologists could decipher hieroglyphics for the first time.

Mazur said,

> It is as if you know one language and this Rosetta stone is going to give you an intense understanding of the other language. But the Taniyama-Shimura conjecture is a Rosetta stone with a certain magical power. The conjecture has the

Number theory is an area of mathematics containing many unsolved problems which can sometimes be stated in language that is accessible even to nonspecialists. The only technical word we shall need is the notion of a prime number. *Primes* are positive integers greater than one that have no factors other than themselves and 1. For example the list of primes would begin with 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, and so on. The Greeks developed an argument (which is given in Chapter 5) to show that this list has no largest element, that is, that the number of primes is infinite. All primes with the exception of 2 are odd and, of course, the sum of two odd primes is an even number. But can all even numbers be generated in this way? Goldbach's Conjecture says that they can. It is an unsolved problem in number theory and one of the oldest unsolved problems in all of mathematics. No one has yet come up with a proof. In fact, in the year 2000 the British publisher Tony Faber put up a million-dollar prize for anyone who could come up with a solution before the year 2002. The prize was a way to generate publicity for the novel *Uncle Petros and Goldbach's Conjecture* by Apostolos Doxiadis. The prize was never awarded.

Goldbach's conjecture can be stated simply:

> Every even number greater than two can be written as the sum of two prime numbers. For example,
>
> $4 = 2 + 2$, $6 = 3 + 3$, $8 = 3 + 5$, $10 = 3 + 7 = 5 + 5$, $12 = 5 + 7$, $14 = 3 + 11 = 7 + 7, \ldots$.

In his fascinating book on number theory, Daniel Shanks divides unsolved problems into two categories: conjectures and open questions. A conjecture is a "proposition that has not been proven, but is favored by some serious evidence." For an "open question," on the other hand, the "evidence is not very convincing one way or another."[24] There is a great deal of evidence in favor of the validity of Goldbach and most mathematicians believe it to be true. For small values of n (small for a number theorist), $n \leq 6 \times 10^{17}$, the conjecture has been verified by computer.[25] In addition there is a heuristic argument (but not a proof) for the validity of the conjecture based on the formula for the statistical distribution of primes. Even at the level of rigorous proof there have been a number of results that go in the direction of the main conjecture. For example, in 1939 L. G. Schnirelmann proved that every number $n \geq 4$ can be written as the sum of at most 300,000 primes. This showed that the conjecture was true for a large but finite number of primes (instead of two). This result has been improved over time so that it is now known that every even number $n \geq 4$ can be written as the sum of at most six primes.[26] Another result, due to Chen Jingrun (1966), is that every sufficiently large even number n ($n \geq N$, for some N) can be written as the sum of two numbers, the first of which is a prime and the second is the product of two primes. There are many more results that go in the direction of Goldbach's Conjecture.

What is the relevance of unsolved problems to our discussion of ambiguity in mathematics? A well-formulated but unsolved problem has an intrinsic ambiguity both in the problem itself and also in the way one thinks about it or works on it. It may be true or false. It is even conceivable that it cannot be resolved one way or another.[27] While we are working on it we don't know the answer, so we must allow both of these possibilities to live in our minds at the same time. Of course we couldn't work on such a problem without having some intuition, based on some substantial evidence, about the validity of the statement. If, following Shanks, we call it a conjecture, then we are guessing that it is true. If we call the problem "open," then we allow for both possibilities. But whether we call it conjecture or open problem, there are always two possibilities—true or false. If we guess false, we must ask ourselves where we might look for a counterexample. If we guess true, we must ask why is it true, and where we would we look for a proof. Whatever we guess, there is always the possibility that we have guessed wrong. If we feel that the statement of the problem is true, then we are faced with another ambiguity. Is the proof accessible? Many conjectures are felt to be true, and yet one senses that a proof would require new ideas, major new developments in the subject that may not happen for many years. Who would risk wasting their careers and a good portion of their lives working on a problem that is not ripe for solution given the current state of development of the subject? So mathematical research is characterized by an instinct for the right problems: those that are significant yet accessible. These are another pair of conflicting or ambiguous characteristics. Too accessible implies the problem is most likely unimportant; too significant and it may be inaccessible.

The ambiguity of an unsolved problem is mitigated somewhat by the Platonic attitude of the working mathematician. That is, she feels that it is objectively either true or false and that the job of the mathematician is "merely" to discover which of these a priori conditions applies. Psychologically, this Platonic point of view brings the ambiguity of the situation into enough control so that researchers have confidence the correct solution exists independent of their efforts. It moves the problem from the domain of "ambiguity as vagueness" in which anything could happen to the sort of incompatibility that has been discussed in this chapter where there are two conflicting frameworks, true or false.

While the unsolved problem is unresolved, the ambiguity of the situation is there for all to see. Let's spend a few words comparing this situation to one in which the ambiguity of the mathematical situation is hidden from view. In the classroom, for example, the teacher and the student often stand on opposite sides of the ambiguity. In the teacher's perception of the situation, there is no ambiguity—the concept being discussed is clear and precise. For the student the concept is ambiguous in both senses of the word: it both is unclear and may contain various "meanings" that actually conflict with one another. However, and this is what is usually not appreciated, even for the teacher the concept retains its ambiguity. For in addition to its clarity, there is also (if the teacher actually has a deeper understanding of the concept) an openness and flexibility that allows the concept to be applied in a variety of circumstances.

Every situation of ambiguity admits a dual viewpoint that we could characterize as known versus unknown or as teacher versus student. In the case of the unsolved problem the "known" side is missing, so there is no disguising the ambiguity. In the teaching situation the teacher may well deny that the concept is ambiguous, but no one can do this for a problem that is unresolved. We really don't know the true state of affairs.

In fact, famous unsolved problems are often of great importance to the development of mathematics even if they remain unresolved. This is because the effort that is spent unraveling them often results in important developments in the subject. The main conjectures may remain unsolved, but other significant questions that arise in the course of the investigations often are solved. As in the case of the Goldbach Conjecture, different, weaker aspects of the main conjecture may be proved, and this leads to increased evidence for or against the main conjecture. The whole situation requires a state of mind that remains at once rigorous and flexible. It requires the ability on the part of the researcher to develop and sustain a state of ambiguity.

I cannot leave the topic of unsolved problems without commenting on what it tells us about the nature of mathematical research and about the art form that is called mathematics. What kind of person attacks such problems? Working on one of the great unsolved problems of mathematics is like embarking on a quest. The anthropologist Joseph Campbell[28] has written about the mythological Hero's Quest. In it the hero braves great perils in order to make some discovery that he brings back for the benefit of humankind. Working on a great mathematical conjecture is a kind of hero quest. What motivates people to spend their lives on such a quest? Why did Wiles spend seven years in his attic working on Fermat? The true motivation for such activity goes beyond fame and fortune—it must be found in the nature of the activity itself. This is another way by which examining mathematics has something to tell us about the nature of the human condition. It seems to me that the notion of the spiritual quest is the closest one will find to such an explanation. A spiritual quest is something that one is driven to do, driven from the deepest level of one's being. A spiritual quest has no rational explanation, or rather, the rational explanation, the adding up of the pluses and the minuses, always misses the mark. One is just so taken with the question, with the beauty and the excitement of the activity, that the effort and the sacrifice seem a small price to pay. A spiritual quest also has something about it that is self-validating and holds the promise of personal transformation. Its goals are both external and internal—a voyage of both discovery and self-discovery.

Any great quest demands courage. It is a voyage into the unknown with no guaranteed results. What is the nature of this courage? It is the courage to open oneself up to the ambiguity of the specific situation. The whole thing may end up as a vast waste of time; that is, the possibility of failure is inevitably present. To work so hard, in the face of possible failure, is what I mean by working with ambiguity. If we stop to think about it, this quality of ambiguity that one finds in the research environment is no different in kind from the ambiguity that is found in our personal lives. Our lives also have this quality of a quest, the attempt to resolve some fundamental but ill-posed question. In working on a mathematical conjecture, life's ambiguities solidify into a concrete problem. That is, the situation of doing research is isomorphic to some extent with the situation we face in our personal lives. This is one reason that working on mathematics is so satisfying. In resolving the mathematical problem we, for a while at least, resolve that larger, existential problem that is consciously or unconsciously always with us.

• • • • •

The above discussion should be borne in mind when we think about the learning of mathematics as students, teachers, or just people who are interested in mathematics. Learning something new entails entering into a situation of ambiguity. Situations of ambiguity are difficult by their very nature. Learners need support when they are encouraged to enter into new unexplored ambiguities. A new learning experience requires the learner to face the unknown, to face failure. Sticking with a true learning situation requires courage and teachers must respect the courage that students exhibit in facing these situations. Teachers should understand and sympathize with students' reluctance to enter into these murky waters. After all, the teacher's role as authority figure is often pleasing insofar as it enables the teacher to escape temporarily from their own ambiguities and vulnerability. Thus the value of learning potentially goes beyond the specific content or technique but in the largest sense is a lesson in life itself.

Ambiguity in Physics

Since physics is the science that is closest to mathematics, one might expect to find that the phenomenon of ambiguity is present in this domain as well. Physics involves an explicit duality, namely, the two dimensions of experiment and theory. One of not what one usually expects to find in physics. After all, one might ask, what is it *really*, a wave or a particle? We feel that it cannot be both. Yet there is one thing of which we can be sure—there is one electron. When we look at it in one way (in one frame of reference) we observe particle-like properties. When we look in another way we get wave-like properties. The electron is precisely the singular entity that emerges out of this fundamental ambiguity.

If complementarity and ambiguity refer to the same phenomenon, then why not call ambiguity complementarity? I maintain that even though these two ideas are referring to a similar phenomenon there are important differences. Complementarity refers to a situation where there is a duality—two contexts the "sum" of whose complementary aspects "adds up" to the entire actual situation. Ambiguity also involves dual concepts, but each context stands on its own, each context describes the entire situation, so to speak. Thus there would be the "particle" description of nature in which subatomic particles are classical objects with definite attributes. On the other hand there would be the "wave" description in which everything is a cloud of probabilities—what Werner Heisenberg called "tendencies for being," "potentia."

Moreover, an ambiguous situation not only boasts dual contexts but also emphasizes the incompatibility between these contexts. It is this "incompatibility" that most sharply differentiates "ambiguity" from "complementarity." It is this incompatibility that is at work when we read the anguished words of physicists who are trying to make sense of subatomic phenomena. They seem not to make sense! One senses that there exists an obstacle that must be overcome if one is to makes sense of this realm of reality. Using the idea of "ambiguity" brings to the fore the need for this "epistemological obstacle" to be overcome—the need for a new vision, the need for a creative leap.

This "incompatibility" gives the entire situation a dynamic aspect. It is like a force that pushes the situation toward a creative reconciliation of the incompatibility. Thus I prefer to think about the electron as an ambiguous object—not in any vague or mystical sense—but in the sense that the electron is both a particle and a wave and yet it cannot be both at the same time. When it is a particle, it is not a wave, and when it is a wave, it is not a particle. In fact the particle/wave ambiguity is so profound that its implications remain a subject of study and speculation.

Ambiguity may then refer to a phenomenon that is present in the external world of physical phenomena as well as in the interior, cognitive world. Which is primary? Does ambiguity (or complementarity) refer to a property of the natural world, and so it finds its way into the biology of our brains and from there into the world of mathematics? Or is ambiguity a feature of our thinking process, and so the conceptual structures that we create inevitably carry this feature? This is in itself a version of the mind/body problem. What is primary, mind or body?

The dominant view in modern cognitive science is that "mind" is a consequence of "brain." There have also been thinkers and traditions that say that "brain" is a consequence of "mind." The dominant Western tradition going back to Descartes is that there is a mind/body duality. I suggest that there is another possibility—an "ambiguous" possibility. I suggest that the mind/brain and subjective/objective situations are not merely dualities or complementarities but ambiguities. Calling them ambiguities makes all the difference because, whereas a duality may be seen to be a fixed and unchangeable aspect of reality, an ambiguity always allows for a higher-level unification. Thus one could say that there is one unified reality that looks subjective when we approach it in one way and objective when we approach it another (see Chapter 8). If reality itself has this ambiguous nature, then it is not so surprising to see the same ambiguous characteristics arising in both the "subjective" domains of mathematical and physical theory as well as in the "objective" domain of subatomic physics.

It is interesting for our discussion of mathematics that quantum mechanics is a completely mathematical theory. Actually it has two different mathematical formalisms, one discrete and the other continuous. That is, the theory of quantum mechanics is itself ambiguous. Now the two descriptions are mathematically isomorphic or equivalent. This does not mean, however, that there is nothing to be gained by having two different ways to look at the situation. On the contrary, given our previous discussion, one would expect that the subtlety of the phenomenon that we are trying to comprehend would require an ambiguous description.

Finally mathematics has something to learn from the world of quantum mechanics. This involves the normal, "formalist" view that mathematics starts off with "self-evident" ideas and builds up to very complex ones, that there is a movement from simplicity to complexity. In the world of quantum mechanics the elementary objects such as the electron and other subatomic particles are extremely subtle and complex entities. That is, it is conceivable that reality is complex all the way down. There may be a lesson here about mathematical objects. Are they not also complex all the way down? Is there any mathematical object that is "trivial" or "obvious" when viewed from *every* possible mathematical point of view? But more of this later on.

String Theory

String theory (and its generalization M-theory) is an exciting, relatively recent attempt to unify the two most fundamental physical theories of our time, general relativity and quantum mechanics. These "two theories underlying the tremendous progress of physics in the last hundred years . . . are mutually incompatible."[31] Thus the need for string theory arises out of the kind of ambiguous situation that I have been describing in this chapter. Both general relativity and quantum mechanics have been spectacularly successful in their respective domains. Their predictions have been experimentally verified to a very high degree of accuracy. Yet they are incompatible in situations in which both theories apply, for example, black holes and the "big bang." It is this context that created the need for a new theory that would unify the gravitational force with the other physical forces. String theory is the prime candidate for such a unifying theory. It is interesting at this stage of the discussion not only because of the ambiguous context in which it arises but because the theory itself incorporates ambiguity in a profound manner.

String theorists have a word for what I have been calling ambiguity—they call it *duality*. "Physicists use the term duality to describe theoretical models that appear to be different but nevertheless can be shown to describe exactly the same physics." There are "trivial" dualities and "nontrivial" dualities. The for-