What will count as mathematics in 2100?

Keith Devlin

What is mathematics today?

What *is* mathematics? That's one of the most basic questions in the philosophy of mathematics. The answer has changed several times throughout history.

Up to 500 B.C. or thereabouts, mathematics was—if it was anything to be given a name—the systematic use of numbers. This was the period of Egyptian, Babylonian, and early Chinese and Japanese mathematics. In those civilizations, mathematics consisted primarily of arithmetic. It was largely utilitarian, and very much of a cookbook variety. ("Do such and such to a number and you will get the answer.")

Modern mathematics, as an area of *study*, traces its lineage to the ancient Greeks of the period from around 500 B.C. to 300 A.D. From the perspective of what is classified as mathematics today, the ancient Greeks focused on properties of number and shape (geometry). [The word "mathematics" itself comes from the Greek for "that which is learnable." As always when interpreting one culture or age with another, it has to be acknowledged that things often appear quite different to those within a particular culture or age than when viewed from the other.]

It was with the Greeks that mathematics came into being as an identifiable discipline, and not just a collection of techniques for measuring, counting, and accounting. Greek interest in mathematics was not just utilitarian; they regarded mathematics as an intellectual pursuit having both aesthetic and religious elements. Around 500 B.C., Thales of Miletus (now part of Turkey) introduced the idea that the precisely stated assertions of mathematics could be logically proved by a formal argument. This innovation marked the birth of the theorem, the central focus of modern mathematics.

The next major change in the overall nature in mathematics (again from the perspective of looking back at the chain of development that led to today's subject) was when Isaac Newton (in England) and Gottfried Leibniz (in Germany) independently invented the calculus. Calculus is the study of continuous motion and change. Previous mathematics had been largely restricted to the static issues of counting, measuring, and describing shape. With the introduction of techniques to handle motion and change, mathematicians were able to study the motion of the planets and of falling bodies on earth, the workings of machinery, the flow of liquids, the expansion of gases, physical forces such as magnetism and electricity, flight, the growth of plants and animals, the spread of epidemics, the fluctuation of profits, and so on. After Newton and Leibniz, mathematics became the study of number, shape, motion, and change.

Most of the initial work involving calculus was directed toward the study of physics; indeed, many of the great mathematicians of the period are also regarded as physicists.

But from about the middle of the 18th century there was an increasing interest in the mathematics itself, not just its applications, as mathematicians sought to understand what lay behind the enormous power that the calculus gave to humankind. Here the old Greek tradition of formal proof came back into ascendancy, as a large part of present-day pure mathematics was developed. By the end of the 19th century, mathematics had become the study of number, shape, motion, change, and of the mathematical tools that are used in this study, together with a number of other topics, such as formal logic and the theory of probabilities. With the growth and diversification of the subject, it became quite difficult to say what mathematics is without writing a short essay.

In the 1980s, however, a definition of mathematics emerged on which most mathematicians now agree, and which captured the broad and increasing range of different branches of the subject: *mathematics is the science of patterns*. This definition does, admittedly, require some elaboration as to what exactly constitutes a pattern, but that aside it captures very well what the subject is about.

According to this new definition, what the mathematician does is examine abstract patterns—numerical patterns, patterns of shape, patterns of motion, patterns of behavior, voting patterns in a population, patterns of repeating chance events, and so on. Those patterns can be either real or imagined, visual or mental, static or dynamic, qualitative or quantitative, purely utilitarian or of little more than recreational interest. They can arise from the world around us, from the depths of space and time, or from the inner workings of the human mind. Different kinds of patterns give rise to different branches of mathematics. For example:

- Arithmetic and number theory study the patterns of number and counting.
- Geometry studies the patterns of shape.
- Calculus allows us to handle patterns of motion (including issues such as velocity and acceleration, polynomial motion, exponential motion, etc.).
- Logic studies patterns of reasoning.
- Probability theory deals with patterns of chance.
- Topology studies patterns of closeness and position.

and so forth.

It is mathematics viewed in this way that I will attempt to project one hundred years into the future, to the start of the 22nd century. But before I do that, I should note that, around 150 years ago, in the middle of the period when mathematics was growing in scope, the subject also changed in nature.

The last revolution in mathematics

For most of its history, mathematics was regarded as primarily about calculation or symbolic manipulation. Proficiency in mathematics was measured primarily in terms of an ability to carry out calculations or manipulate symbolic expressions to solve problems. In the middle of the 19th century, however, a revolution took place. Generally regarded as having its epicenter in the small university town of Göttingen in Germany, the revolution's leaders were the mathematicians Lejeune Dirichlet, Richard Dedekind, and Bernhard Riemann. In their new conception of the subject, the primary focus was not performing a calculation or computing an answer, but formulating and understanding abstract concepts and relationships. This represented a shift in emphasis from *doing* to *understanding*. For the Göttingen revolutionaries, mathematics was about "Thinking in concepts'" (*Denken in Begriffen).* Mathematical objects, which had been thought of as given primarily by formulas, came to be viewed rather as carriers of conceptual properties. Proving was no longer a matter of transforming terms in accordance with rules, but a process of logical deduction from concepts.

For example, one post-Göttingen concept is the modern notion of a function. Prior to the 19^{th} century, mathematicians were used to the fact that a formula such as $y = x^2 + 3x – 5$ specifies a rule that produces a new number (y) from any given number (x). Then along came Dirichlet who said, forget the formula and concentrate on what the function *does* in terms of input–output behavior. A *function*, according to Dirichlet, is any rule that produces new numbers from old. The rule does not have to be specified by an algebraic formula. In fact, there's no reason to restrict your attention to numbers. A function can be any rule that takes objects of one kind and produces new objects from them.

Mathematicians began to study the properties of abstract functions, specified not by some formula but by their behavior. For example, does the function have the property that when you present it with different starting values it always produces different answers? (Injectivity.)

This approach was particularly fruitful in the development of real analysis, where mathematicians studied the properties of continuity and differentiability of functions as abstract concepts in their own right.

Karl Weierstrass in Germany and Augustin Cauchy in France analyzed continuity and differentiability, finally coming up with today's famous epsilon-delta definitions. With Cauchy's contributions, in particular, mathematicians finally had a rigorous way to handle infinity, a concept that their predecessors had grappled with since the ancient Greek era. Riemann spoke of mathematics having reached "a turning point in the conception of the infinite."

In the 1850s, Riemann defined a complex function by *its property of differentiability*, rather than a formula, which he regarded as secondary. Karl Friedrich Gauss's residue classes were a forerunner of the approach—now standard—whereby a mathematical structure is defined as a set endowed with certain operations, whose behaviors are specified by axioms. Taking his lead from Gauss, Dedekind examined the new concepts of ring, field, and ideal—each of which was defined as a collection of objects endowed with certain operations.

Like most revolutions, the Göttingen one had its origins long before the main protagonists came on the scene. The Greeks had certainly shown an interest in mathematics as a conceptual endeavor, not just calculation, and in the 17^{th} century, Gottfried Leibniz thought deeply about both approaches. But for the most part, until the

Göttingen revolution, mathematics was viewed primarily as a collection of procedures for solving problems. To today's mathematicians, however, brought up entirely with the post-Göttingen conception of mathematics, what in the 19th century was a revolution is simply taken to be what mathematics is.

How will mathematicians one hundred years from now view their subject? Will it be more of the same? Will there be radically new branches of the subject? Or will there be another revolution, which changes the very nature of what is viewed as "mainstream mathematics"? These are the questions I want to address here.

How and why mathematics changes

The first thing to note is that, for all its abstraction, much of mathematics has been developed in response to the needs of society. For example, the needs of commerce and trade led to the development of techniques of arithmetic, and navigation and architecture gave rise to geometry and trigonometry. A great deal of the mathematics developed since the 17th century was created with applications in the physical world in mind. In particular, the invention of calculus was motivated in large part by the need to being precision to the study of the motion of the planets. In the physical domain for which it was developed, today's mathematics has been highly successful. That success depends in large part upon the deterministic nature of much of the physical universe, which makes it amenable to a mathematical approach.

While foretelling the future is always a perilous activity, most present-day scientists seem to agree that the next hundred years will see major advances in the life sciences—some of those sciences very new—such as biology, psychology, sociology, neuroscience, and the study of mind and consciousness. Some of these areas seem amenable to the application of current mathematical techniques. For example, many parts of biology already are highly mathematical. Other areas may yield to techniques that, while new, are not radically different from current mathematics. When it comes to the social and psychological world of people, however, we are in a (seemingly) highly nondeterministic realm that appears to rule out more than the occasional, fairly superficial use of mathematics *as we understand it today*. (The exception is where larger populations are concerned, when statistical techniques can capture the deterministic order that can emerge from the often nondeterministic actions of the individuals.) It is these areas that will, I believe, give rise to the development of new mathematics. But here's the rub. I suspect that this new mathematics will not look very much like today's mathematics. Whereas the Göttingen revolution changed the *nature* of mathematics, but left it *looking* on the surface much as it always had, I believe the next revolution will leave the fundamental *nature* of mathematics unchanged but will lead to something that *looks* very different on the surface.

The reason why this new mathematics will look different is precisely because it will be applied to domains having either a significant degree of nondeterminism or else such high complexity as to defy capture within a traditional mathematical framework in any form intelligible to the human mind. Handling such matters will require a new form of analysis that, *to present day eyes*, will look like a blend of mathematics with rigorous, logical—but *not mathematical*—reasoning (of the kind you can find today in, say, psychology or sociology — I'll give some examples presently).

This new form or reasoning—what I am suggesting will come to be viewed as a new form of *mathematics*—will result from the meeting of two approaches that can be witnessed today: top down (where attempts are made to apply current mathematical techniques to some human domain) and bottom up (where attempts are made to make human-science arguments more mathematical). At present, the gap between those two approaches is generally large. But a hundred years from now, it will, I believe, in many cases have been bridged.

Let me first illustrate my point with three examples from fields where both the bottom up and top down approaches have progressed quite far. All three involve quantifying uncertainties.

Bernoulli's utility concept

My first example is Daniel Bernoulli's work on risk assessment. The great 18th century Swiss mathematician, whose uncle Jacob did pioneering work in the mathematical theory of probability, set out to try to understand why people assess risk the way they do.

A modern-day example that shows how the human assessment of risk can differ from the mathematical analysis is the fear some people have about flying. Such individuals may know that the probability of being involved in a major airline accident is far less than being in a major auto accident. The issue is the nature of an airline crash and the importance they attach to such an event, however unlikely it may be. Fear caused by lightning is a similar phenomenon, where the tiny mathematical probability of being struck by a lightning bolt is far outweighed by the significance many individuals attach to the possibility of such an event.

It was this, essentially human, aspect of risk assessment that interested Daniel Bernoulli. To try to capture mathematically the way people actually assessed risk, he introduced the concept of *utility*.

Utility depends upon another notion of probability theory that preceded it: *expectation* (or *expected value*). Your *expectation* in, say, playing a certain game is a measure of what you can "expect" to win. It is the average amount you would win per game if you were to play repeatedly. To compute your expectation, you take the probability of each possible outcome and multiply it by the amount you would win in that case, and then add together all those amounts. By taking account of both the probabilities and the payoffs, the expectation measures the value to an individual of a particular risk or wager. The greater the expectation, the more attractive the risk.

For many examples, expectation seemed to work well enough. But there was a problem, and it was most dramatically illustrated by a tantalizing puzzle proposed by Daniel's cousin Nicolaus, commonly known as the Saint Petersburg Paradox. Here it is.

Suppose I challenge you to a game of repeated coin tosses. If you throw a head on the first toss, I pay you $2 and the game is over. If you throw a tail on the first throw and a head on the second, I pay you $4 and the game is over. If you throw two tails and then a head, I pay you $8 and the game is over. We continue in this fashion until you throw a head. Each time you throw a tail and the game continues, I double the amount you will win if and when you throw a head.

Now imagine that a friend comes along and offers to pay you $10 to take your place in the game. Would you accept or decline? What if he offered you $50? Or $100? In other words, how much do you judge the game to be worth to you?

The expectation of this game works out to be infinite, so in theory you should not give up your opportunity to play for any amount of money. But most people—even knowledgeable probability theorists—would be tempted to take a fairly low offer. Why is this?

This was the kind of problem with expectation that led Bernoulli to replace the highly mathematical concept of expectation (an example of what I am calling a top-down use of mathematics) by the far less formal and less precise concept of utility (which is very definitely bottom-up).

Utility is intended to measure the significance you attach to a particular outcome. As such, utility is very much an individual thing. It depends on the value a person puts on a particular event. Your utility and mine might differ.

At first glance, the move to replace the mathematically precise concept of expectation by the decidedly personal idea of utility might appear to render impossible any further mathematical analysis. Even for a single individual, it may well be impossible to assign specific numerical values to utility. Nevertheless, Bernoulli was able to make a meaningful, and definitely mathematical, observation about the concept. He wrote: "[The] utility resulting from any small increase in wealth will be inversely proportionate to the quantity of goods previously possessed."

Bernoulli's utility law explains why even moderately wealthy individuals will generally find it much more painful to lose half their fortune than the pleasure or benefit gained by doubling it. As a result, few of us are prepared to gamble half our wealth for the chance of doubling it. Only when we are truly able to declare "What have I got to lose?" are most of us prepared to take a big gamble.

For instance, suppose you and I each has a net worth of $10,000. I offer you a single toss of a coin. Heads and I give you $5,000; tails and you give me $5,000. The winner comes out with $15,000, the loser with $5,000. Since the payoffs are equal and the probability of each of us winning is 1/2, we each have an expectation of zero. In other words, according to expectation theory, it makes no difference to either of us whether we play or not. But few of us would play. We would almost certainly view it as taking an unacceptable risk. The 0.5 probability of losing $5,000 (half our wealth) far outweighs the 0.5 probability of winning the same amount.

Bernoulli’s concept of utility likewise explains the Saint Petersburg paradox. According to Bernoulli's law, once you reach the stage where your minimum winning represents a measurable gain *in your terms*, the benefit to be gained by playing longer starts to decrease. That determines the amount for which you would be prepared to sell your place in the game.

So much for expectation. In fact, a similar fate was to befall Bernoulli's utility concept in due course, when mathematicians and economists of a later generation collaborated with psychologists to look more closely at human behavior. But the fact remains that it was Bernoulli who first insisted that if you wanted to apply mathematics to real world problems that involve chance, and if you want the results of that analysis to be of real use, then you had to take account of the human factor. In so doing, he was approaching the issue in a bottom-up fashion, and thereby making one of the first steps toward what I am suggesting will eventually be classified as a fully-fledged branch of mathematics—something that will be done in university departments of mathematics and taught (*as mathematics*) to mathematics students.

Bayesian inference

My second example of what the 22nd century will view as mathematics addresses the question: How do you use inconclusive evidence to assess the probability that a certain event will occur? One method that has become increasingly popular in recent years depends on a mathematical theorem proved by an 18th century English Presbyterian minister by the name of Thomas Bayes. Curiously, Bayes’ theorem languished largely ignored and unused for over two centuries before statisticians, lawyers, medical researchers, software developers, and others started to use it in earnest during the 1990s.

Bayesian inference, as the method using Bayes’ theorem is called, is a step toward a new mathematics because it uses an honest-to-goodness mathematical formula (Bayes’ theorem) in order to improve—on the basis of evidence—the best (human) estimate that a particular event will take place. In the words of some statisticians, it’s “mathematics on top of common sense.” You start with an initial estimate of the probability that the event will occur and an estimate of the reliability of the evidence. The method then tells you how to combine those two figures—in a precise, mathematical way—to give a new estimate of the event’s probability in the light of the evidence.

In some highly constrained situations, both initial estimates may be entirely accurate, and in such cases Bayes’ method will give you the correct exact answer. In a more typical real-life situation, you don’t have exact figures, but as long as the initial estimates are reasonably good, then the method will give you a better estimate of the probability that the event of interest will occur. Thus, in the hands of an expert in the domain under consideration, someone who is able to assess all the available evidence reliably, Bayes' method can be a powerful tool.

Specifically, Bayes' theorem shows you how to calculate the probability of a certain hypothesis H, based on evidence E. Let P(H) be the probability that the hypothesis H is correct in the absence of any evidence—the *prior probability*. Let P(H|E) be the probability that H is correct given the evidence E. This is the revised estimate you want to calculate. Let P(E|H) be the probability that E would be found if H were correct. This is called the *likelihood*. To compute the new estimate, you first have to calculate P(H-wrong), the probability that H is false, and you have to calculate P(E|H-wrong), the probability that the evidence E would be found in the event that H were false.

Bayes' theorem says that:

$$P(H|E) = \frac{P(H) \times P(E|H)}{P(H) \times P(E|H) ++ P(H\text{-wrong}) \times P(E|H\text{-wrong})}.$$

A quantity such as P(H|E) is known as a *conditional probability*—the conditional probability of H occurring, given the evidence E.

Unscrupulous lawyers have been known to take advantage of the lack of mathematical sophistication among judges and juries by deliberately confusing the two conditional probabilities P(G|E), the probability that the defendant is guilty given the evidence, and P(E|G), the conditional probability that the evidence would be found assuming the defendant were guilty. Such misuse of probabilities is a real possibility in cases where scientific evidence such as DNA testing is involved, such as paternity suits and rape and murder cases. Prosecuting attorneys in such cases have been known to provide the court with a figure for P(E), whereas the figure relevant to deciding guilt is P(G|E), which, as Bayes' formula shows, is generally much lower than P(E). Unless there is other evidence that puts the defendant into the group of possible suspects, such use of P(E) is highly suspect, and indeed should perhaps be prohibited. The reason is that it ignores the initial low prior probability that a person chosen at random is guilty of the crime in question.

In addition to its use—or misuse—in court cases, Bayesian inference methods lie behind a number of new products on the market. For example, chemists make use of Bayesian methods to improve the resolution of nuclear magnetic resonance (NMR) spectrum data. When NMR spectroscopy is used to determine three-dimensional molecular structures, problems remain in translating the data into atomic coordinates. The data is usually insufficient to uniquely define a structure, and subjective choices in data treatment and parameter settings make it difficult to judge the precision of NMR structures. To overcome this problem, probabilistic methods are used to calculate structures from NMR data. The idea is to view structure determination as an inference problem, and use a Bayesian approach to derive a probability distribution that represents the calculated structure and its precision. This approach can improve the resolution of the data by several orders of magnitude.

Other recent uses of Bayesian inference are in the evaluation of new drugs and medical treatments, the analysis of human DNA to identify particular genes, analyzing police arrest data to see if any officers have been targeting one particular ethnic group, and counter-terrorism intelligence analysis.

At the moment, the uses of Bayesian methods are viewed as a combination of mathematics (in the traditional sense) and other forms of reasoning. What will change, I believe, is that the entire reasoning process will come to be viewed as, simply, *mathematics*, as I shall argue toward the end of this article.

Black-Scholes theory

My third example of what I think will be part of 22nd century mathematics is provided by the field of economics and finance.

The 1997 Nobel Prize for economics was awarded for the 1970 discovery of a mathematical formula. The prizewinners were Stanford University professor of finance (emeritus) Myron Scholes and economist Robert C. Merton of Harvard University. The prize would undoubtedly have been shared with a third person, Fischer Black, but for the latter's untimely death in 1995.

Discovered by Scholes and Black, and developed by Merton, the Black-Scholes formula tells investors what value to put on a financial derivative, such as a stock option. Use of the Black-Scholes formula is a clear example of how mathematics can be blended with other forms of reasoning. Human judgment is required both in providing numerical values to some of the formula's input variables and in deciding how much weight to attach to the derivative's value the formula provides.

When the Black-Scholes method was first introduced, the idea that you could use mathematics to (help) price derivatives was so revolutionary that Black and Scholes had difficulty publishing their work. When they first tried in 1970, Chicago University's *Journal of Political Economy* and Harvard's *Review of Economics and Statistics* both rejected the paper without even bothering to have it refereed. It was only in 1973, after some influential members of the Chicago faculty put pressure on the journal editors, that the *Journal of Political Economy* published the paper. (Black-Scholes 1973)

Industry was far less shortsighted than the ivory-towered editors at the University of Chicago and Harvard. Within six months of the publication of the Black-Scholes article, Texas Instruments had incorporated the new formula into their latest calculator, announcing the new feature with a half-page ad in *The Wall Street Journal*.

Modern risk management, including insurance, stock trading, and investment, rests upon the fact that you can use mathematics to predict the future well enough that you can make a wise decision as to where to put your money. When you take out insurance or purchase stock, the real commodity you are dealing with is risk. The underlying ethos in the financial markets is that the more risk you are prepared to take, the greater the potential rewards. Using mathematics can never remove the risk. But it can help to tell you just how much of a risk you are taking, and help you decide on a fair price.

What Black and Scholes did was find a way to determine the fair price to charge for a derivative such as a stock option. The idea with stock options is that you purchase an option to buy stock at an agreed price prior to some fixed later date. If the value of the stock rises above the agreed price before the option runs out, you buy the stock at the agreed lower price and thereby make a profit. If you want, you can simply sell the stock immediately and realize your profit. If the stock does not rise above the agreed price, then you don't have to buy it, but you lose the money you paid out to purchase the option in the first place.

What makes stock options attractive is that the purchaser knows in advance what the maximum loss is: the cost of the option. The potential profit is theoretically limitless: if the stock value rises dramatically before the option runs out, you stand to make a killing. Stock options are particularly attractive when they are for stock in a market which sees large, rapid fluctuations, such as the computer and software industries. Most of the many thousands of Silicon Valley millionaires became rich because they elected to take a portion of their initial salary in the form of stock options in their new company.

The question is, how do you decide a fair price to charge for an option on a particular stock? This is precisely the question that Scholes, Black, and Merton investigated back in the late 1960s. Black was a mathematical physicist with a recent doctorate from Harvard, who had left physics and was working for Arthur D. Little, the Boston-based management consulting firm. Scholes had just obtained a Ph.D. in finance from the University of Chicago. Merton had obtained a bachelor of science degree in mathematical engineering at New York's Columbia University, and had found a job as a teaching assistant in economics at MIT.

The three young researchers—all were still in their twenties—set about trying to find an answer using mathematics, exactly the way a physicist or an engineer approaches a problem. But would a mathematical approach work in the highly volatile world of options trading, which was just being developed at the time. (The Chicago Board Options Exchange opened in April 1973, just one month before the Black-Scholes paper appeared in print.) Many senior market traders thought such an approach could not possibly work, and that options trading was beyond mathematics. If that were the case, then options trading was an entirely wild gamble, strictly for the foolhardy.

The old guard were wrong. Mathematics could be applied. It was heavy duty mathematics at that, involving stochastic differential equations. The formula takes four input variables—duration of the option, prices, interest rates, and market volatility—and produces a price that should be charged for the option.

Not only did the new formula work, it transformed the market. When the Chicago Options Exchange first opened in 1973, less than 1,000 options were traded on the first day. By 1995, over a million options were changing hands each day.

So great was the role played by the Black-Scholes formula (and extensions due to Merton) in the growth of the new options market that, when the American stock market crashed in 1978, the influential business magazine *Forbes* put the blame squarely onto that one formula. Scholes himself has said that it was not so much the formula that was to blame, rather that market traders had not grown sufficiently sophisticated in how to use it.

At present the use of the Black–Scholes formula in assessing the value of stock options is viewed as a combination of (standard) mathematics and other forms of reasoning, but again I believe that in due course the entire reasoning process will be thought of as mathematics.

Mathematical theories of language

One feature of the examples I have presented so far is that the top down and bottom up approaches have to some extent already met. My next, and final set of examples are taken from a field where this has not yet happened: linguistics.

The first major attempt to develop a mathematical theory of ordinary language began with the publication of Noam Chomsky's seminal work *Syntactic Structures* in 1957.

Chomsky based his analysis on the axiomatic approach to mathematics, where the mathematician starts with an initial set of assumptions, or *axioms*, and then proceeds to deduce truths (*theorems*) from those axioms. He was inspired, in particular, by the dramatic advances that had been made in mathematical logic in the first half of the 20th century, and by the new branch of mathematics known as recursion theory.

Just as the logicians had been able to formulate axioms and rules that show how a mathematical proof is constructed as a chain of mathematical statements, so too Chomsky formulated rules that show how grammatical sentences are built up from words and phrases. For example, one such rule is that a sentence can be constructed by taking a determinate noun phrase (i.e., a noun phrase that starts with a determiner such as *the* or *a*) and following it by a verb phrase. For example, the sentence *The large black dog licked the tabby kitten* consists of the determinate noun phrase *The large black dog* followed by the verb phrase *licked the tabby kitten*. Using the letter S to stand for "sentence", DNP to denote "determinate noun phrase", and VP to denote "verb phrase", this rule may be written like this:

$$S \rightarrow \text{DNP VP}$$

This expression is read as "S arrows DNP VP," or more colloquially, "a sentence results from taking a determinate noun phrase and following it by a verb phrase." The sentences (S), determinate noun phrases (DNP), and verb phrases (VP) are examples of what are called *syntactic categories*.

The rule for generating determinate noun phrases (DNP) from noun phrases (NP) is

DNP $\rightarrow$ DET NP

where DET is given by the rule

DET $\rightarrow$ *the*, *a*

The first of the above two rules reads "a determinate noun phrase can be generated by taking a determiner and following it by a noun phrase." The second rule is an example of what is called a lexical rule, since it assigns particular words (i.e., items of the lexicon) to a syntactic category, namely the syntactic category DET of determiners. It says that either of the words *the* and *a* constitutes a determiner.

Thus, the determinate noun phrase *the large black dog* is generated by taking the determiner *the* and following it by the noun phrase *large black dog*.

Here are some further rules of syntax:

VP $\rightarrow$ V DNP

"A verb phrase results from taking a verb and following it by a determinate noun phrase."

NP $\rightarrow$ A NP

"A noun phrase results from taking an adjective and following it by a noun phrase." This rule has a circular property, in that you start with a noun phrase and the rule gives you another noun phrase. For example, the noun phrase *large black dog* results from combining the adjective *large* with the noun phrase *black dog*. The rule could then be applied again to give the noun phrase *old large black dog*, etc.

NP $\rightarrow$ N

"A noun is (itself) a noun phrase."

S $\rightarrow$ *If* S *then* S

"A sentence can consist of the word *If* followed by a sentence followed by the word *then* followed by another sentence." For example,

If John comes home then we will play chess.

On their own, these kinds of rules for combining phrases to give sentences produce stilted, machine-like, and often ungrammatical utterances of the kind produced by a robot or a space alien in a low budget science fiction movie. To obtain the correct verb forms, gender and plurality agreements, etc. to give a genuinely grammatical sentence, the strings of words generated by the initial composition rules have to be "massaged." Chomsky introduced *transformation rules* to perform this task.

Transformation rules take a word string produced by the generative grammar, such as *The large black dog lick the tabby kitten*, and turn it into a grammatical sentence such as *The large black dog licks the tabby kitten* or *The large black dog licked the tabby kitten*. Similarly, transformation rules generate other variants of the initial word string, such as the passive form *The tabby kitten is licked by the large black dog* or the question *Did the large black dog lick the tabby kitten?*

For example, the grammar-generated string

The large black dog lick the tabby kitten.

is transformed to the sentence

The large black dog licks the tabby kitten.

by the rule

$$\mathrm{DNP}_{\mathrm{sing}}\,\mathrm{V}_{\mathrm{stem}}\;\mathrm{DNP} \;\rightarrow\; \mathrm{DNP}_{\mathrm{sing}}\,\mathrm{V}_{\mathrm{stem}}\textit{s}\;\mathrm{DNP}$$

In words, this reads: "Starting with a grammar-generated string consisting of a singular determinate noun phrase, add the letter *s* to the verb stem." (*Note that this kind of rule uses the arrow to mean something different from its meaning in the generative grammar or the lexicon*.)

The same grammar-generated string

The large black dog lick the tabby kitten.

is transformed to the passive sentence

The tabby kitten is licked by the large black dog.

by the transformation rule

$$\mathrm{DNP}^1{}_{\mathrm{sing}}\,\mathrm{V}_{\mathrm{stem}}\;\mathrm{DNP}^2 \;\rightarrow\; \mathrm{DNP}^2\;\textit{is}\;\mathrm{V}_{\mathrm{stem}}\textit{ed by}\;\mathrm{DNP}^1{}_{\mathrm{sing}}$$

The superscripts 1 and 2 on the two DNPs are there simply to indicate that there are two different phrases involved, and to keep track of which goes where. This rule reads: "Given a grammar-generated string consisting of a singular determinate noun phrase, a verb stem, and a determinate noun phrase, to obtain the passive form, put the second noun phrase first, follow it by the word *is*, followed by the verb stem with the string *ed* appended, followed by the word *by*, followed in turn by the first noun phrase." This rule does not apply to irregular verbs, such as *break–broken*, which form their passives in a different manner.

Even with an extensive list of transformation rules, Chomsky's mathematical treatment of syntax provided at best a fairly crude *model* of syntactic structure, and in his subsequent work he adopted a different approach. The significance of his original work from my present perspective is that it was a first attempt to develop a mathematical description of linguistic structure in a top down fashion, by formulating formal, symbolic rules—a formal system—and then trying to modify the framework to provide a better fit to real language.

Chomsky's was not the only attempt to develop a mathematical model of language. The logician Richard Montague developed an elaborate mathematical model of linguistic meaning that came to be known as Montague semantics, inspired by the work on mathematical truth of his doctoral advisor Alfred Tarski. (See Montague 1974 for details.)

Grice's maxims

Both Chomsky and Montague took an approach that was predominantly top down. An example of a bottom-up approach to language was presented in a lecture given by the British philosopher and logician H. P. (Paul) Grice at Harvard University in 1967. Grice subsequently published his lecture under the title *Logic and Conversation*. (See Grice 1975.) In his talk, he formulated a set of "maxims" that participants in a conversation implicitly follow. It was a bold attempt to apply a mathematical approach to the structure of conversation, very much in the spirit of Euclid's formulation of axioms for plane geometry.

Grice was trying to analyze the structure any conversation must have, regardless of its topic and purpose. He began by observing that a conversation is a cooperative act, which the two participants enter into with a purpose. He tried to encapsulate the cooperative nature of conversation by what he called the Cooperative Principle:

> *Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.*

In other words, be cooperative.

Grice went on to derive more specific principles—his maxims—from the Cooperative Principle, by examining its consequences under four different headings: quantity, quality, relation, and manner. He illustrated these four headings by means of non-linguistic analogies:

Quantity. If you are assisting a friend to repair his car, your contribution should be neither more nor less than is required; for example, if your friend needs four screws at a particular moment, she expects you to hand her four, not two or six.

Quality. If you and a friend are making a cake, your contributions to this joint activity should be genuine and not spurious. If your friend says he needs the sugar, he does not expect you to hand him the salt.

Relation. Staying with the cake making scenario, your contribution at each stage should be appropriate to the immediate needs of the activity; for example, if your friend is mixing the ingredients, he does not expect to be handed a novel to read, even if it is a novel he would, at some other time, desire to read.

Manner. Whatever joint activity you are engaged in with a friend, your partner will expect you to make it clear what contribution you are making, and to execute your contribution with reasonable dispatch.

In terms of conversation, the category of *quantity* relates to the amount of information the speaker should provide. In this category, Grice formulated two maxims:

1. Make your contribution as informative as is required.
2. Do not make your contribution more informative than is required.

Under the category of quality, Grice listed three maxims, the second two being refinements of the first:

1. Try to make your contribution one that is true.
2. Do not say what you believe to be false.
3. Do not say that for which you lack adequate evidence.

Under the category *relation*, Grice gave just one maxim:

Be relevant.

However, Grice observed that it would take a great deal more study to come up with more specific maxims that stipulate what is required to be relevant at any particular stage in a conversation.

Finally, under the category of *manner*, Grice listed five maxims, a general one followed by four refinements, although he remarked that the list of refinements might be incomplete:

1. Be perspicuous.
2. Avoid obscurity of expression.
3. Avoid ambiguity.
4. Be brief.
5. Be orderly.

As Grice observed, his maxims are not laws that have to be followed. In that respect, they are not like mathematical axioms. If you want to perform an arithmetical calculation in a proper manner, you have to obey the rules of arithmetic. But it is possible to engage in a genuine and meaningful conversation and yet fail to observe one or more of the maxims Grice listed. The maxims seem more a matter of an obligation of some kind. In Grice's own words: "I would like to be able to think of the standard type of conversational practice not merely as something which all or most do *in fact* follow, but as something which it is *reasonable* for us to follow, which we *should not* abandon." [Emphasis as in the original.]

One of the more interesting parts of Grice's analysis is his discussion of the uses to which people may put his maxims in the course of an ordinary conversation. Indeed, it was this part of his work that makes it a contribution to a science of communication. In science, the real tests of a new theory come when the scientist (1) checks the theory against further evidence, (2) attempts to base explanations on the theory, and (3) tries to use the theory to make predictions that can then be tested.

Grice made successful use of his maxims in analyzing a widespread conversational phenomenon he called *conversational implicature*.

Conversational implicature

Conversational implicature occurs when a person says one thing and means something other than the literal meaning. For example, suppose Naomi says to Melissa, "I am cold" after Melissa has just entered the room and left the door wide open. Literally, Naomi has simply informed Melissa of her body temperature. But what she surely means is "Please close the door." Naomi's words do not actually say this; rather it is *implicated* by her words. Grice used the word "implicate" rather than "imply" for such cases since Naomi's words certainly do not *imply* the "close the door" meaning in any logical sense. Assuming Melissa understands Naomi's remark as a request to close the door, she does so because of cultural knowledge, not logic.

Conversational implicatures are ubiquitous in our everyday use of language. They can be intended by the speaker, or can be made by the listener. Traditional methods of analyzing language say virtually nothing about the way conversational implicature works. Grice used his maxims to analyze the phenomenon.

Although Grice makes no claim that people have any conscious awareness of his maxims, his discussion of conversational implicature establishes a strong case that the maxims capture part of the abstract structure of conversation. They enable the linguist to provide satisfactory, after-the-event explanations of a variety of conversational gambits.

According to Grice, a participant in a conversation, say Bill in conversation with Doris, may fail to fulfill a maxim in various ways, including the following.

(1) Bill may quietly and unostentatiously violate a maxim. In some cases, Bill will thereby mislead Doris.

(2) Bill may opt out from the operation both of the maxim and the Cooperative Principle, making it plain that he is unwilling to cooperate in the way the maxim requires. For example, he might say, "I cannot say more."

(3) Bill may be faced with a clash. For example, he may find it impossible to satisfy both the quantity maxim "Be as informative as required" and the quality maxim "Have adequate evidence for what you say."

(4) Bill may flout or blatantly fail to fulfill a maxim. Assuming that Bill could satisfy the maxim without violating another maxim, that he is not opting out, and that his failure to satisfy the maxim is so blatant that it is clear he is not trying to mislead, then Doris has to find a way to reconcile what Bill actually says with the assumption that he is observing the Cooperative Principle.

Case (4) is the one that Grice suggests most typically gives rise to a conversational implicature. For example, suppose Professor Alice Smith is writing a testimonial for her linguistics student Mark Jones, who is seeking an academic appointment at MIT. She writes a letter in which she praises Jones's well groomed appearance, his punctuality, his handwriting, and his prowess at tennis, but does not say anything about his ability as a student of linguistics. Clearly, Professor Smith is flouting the maxim "Be relevant." The implicature is that Professor Smith has nothing good to say about Jones's ability in linguistics, but is reluctant to put her opinion in writing.

Notice that Grice's analysis of conversation does not involve any use of mathematical notation. Nevertheless, it is inspired by axiomatic mathematics, and the methodology is very definitely mathematical.

Sociolinguistics

Another example of a bottom-up attempt to develop a "mathematical" analysis of language is provided by a seminal article published in 1972 called *On the Analyzability of Stories by Children* (Sacks 1972), in which the sociologist Harvey Sacks tried to understand the way people use and understand ordinary language in an everyday setting, in particular the way a speaker and a listener make use of their knowledge of social structure to communicate. According to Sacks, the particular choice of words used by a speaker in, say, a description is critically influenced by her knowledge of social structure, and the listener utilizes his knowledge of social structure in order to interpret, in the manner the speaker intended, the juxtaposition of these words.

The principal data Sacks examined consists of the first two sentences uttered by a small child asked to tell a story:

> *The baby cried. The mommy picked it up.*

As Sacks observes, when heard by a typical, competent speaker of English, the utterance is almost certainly heard as referring to a very small human (although the word *baby* has other meanings in everyday speech) and to that baby's mommy (even though there is no genitive in the second sentence, and it is certainly consistent for the mommy to be some other child's mother). Moreover it is the baby that the mother picks up (although the *it* in the second sentence could refer to some object other than the baby). Why do we almost certainly, and without seeming to give the matter any thought, choose this particular interpretation?

To continue, we are also likely to regard the second sentence as describing an action (the mommy picking up the baby) that follows, and is caused by, the action described by the first sentence (the baby crying). We do this even though there is no general rule to the effect the sentence order corresponds to temporal order or causality of events (even though it often does so).

Moreover, we may form this interpretation without knowing what baby or what mommy is being talked of.

Furthermore, we recognize these two sentences as constituting a “possible description” (Sacks’ terminology) of an ordered sequence of events. Indeed it seems to be in large part because we make such recognition that we understand the two sentences the way we do.

As Sacks noted, what leads us effortlessly, instantaneously, and almost invariably, to the interpretation we give to this simple discourse, is the speaker and listener’s shared knowledge of, and experience with, the social structure that pertains to (the subject matter of) this particular utterance. Specifically, it is our knowledge of the way mothers behave toward their babies in our culture that leads us to hear the two sentences the way we do. It is this underlying social structure that Sacks is after.

Sacks was the first to admit that the chosen example is extremely simple. But, he claimed, far from rendering his study trivial, this very simplicity makes his observations all the more striking. He observed: “the fine power of a culture ... does not, so to speak, merely fill brains in roughly the same way, it fills them so that they are alike in fine detail.”

It is that observation that the influence of culture on human behavior makes us *alike in fine detail* that makes the domain amenable to a mathematical (in the evolving sense I am talking about) analysis.

To begin his analysis, Sacks first of all introduced what he called *categories*. For example, the following are categories (of persons): *male*, *female*, *baby*, *mommy*. He then defined a (*membership*) *categorization device* to be a non-empty collection of categories that “go together” in some natural way, together with rules of application. For example, the categorization device *gender*. This device consists of the two categories *male* and *female*, together with the rule for applying these categories to (say) human populations. Other examples are the *family* categorization device, which consists of categories such as *baby*, *mommy*, *daddy*, etc. and the *stage-of-life* device, which consists of categories such as *baby*, *child*, *adult*, etc.

As examples of the rules of application that are part of a categorization device, Sacks gave the following. First, *the economy rule*:

(ER) A single category from any device can be referentially adequate.

For instance, the economy rule allows use of the phrase *the baby* to be referentially adequate.

Sacks' second rule, *the consistency rule*, says:

(CR) If some population of persons is being categorized, and if a category from some device has been used to categorize one member of the population, then that category, or other categories of the same device, may be used to categorize further members of the population.

For instance, if the device *family* has been used to refer to some baby by means of the category *baby*, then further persons may be referred to by other categories in the same device, such as *mommy* and *daddy*.

Associated with the consistency rule, Sacks formulated the following *hearer's maxim*:

(HM1) If two or more categories are used to categorize two or more members of some population, and those categories can be heard as categories from the same device, then hear them that way.

For instance it is in this way that in the two sentences under consideration, "baby" and "mommy" are heard as from the *family* device. But notice that this does not preclude our simultaneously hearing "baby" as from the *stage-of-life* device—indeed, as Sacks himself argued, this is probably what does occur.

This hearer's maxim does not fully capture what goes on in the example under consideration. For it is not just that "baby" and "mommy" are heard as belonging to the same category *family*; rather they are heard as referring to individuals in *the very same* family. Sacks explained this by observing that the device *family* is one that is, what he called *duplicatively organized*. In his own words:

> When such a device is used on a population, what is done is to take its categories, treat the set of categories as defining a unit, and place members of the population into cases of the unit. If a population is so treated and is then counted, one counts not numbers of daddies, numbers of mommies, and numbers of babies but numbers of families—numbers of "whole families", numbers of "families without fathers", etc. A population so treated is partitioned into cases of the unit, cases for which what properly holds is that the various persons partitioned into any case are "coincumbents" of that case.

The following hearer's maxim is associated with duplicatively organized devices:

(HM2) If some population has been categorized by means of a duplicatively organized device, and a member is presented with a categorized population which can be heard as coincumbents of a case of that device's unit, then hear it that way.

According to Sacks, it is this maxim that results in our hearing “the baby” and “the mommy” in our example as referring to individuals in the very same family.

Sacks’ next point was that the phrase *the baby* is in fact heard not just in terms of the *family* device but simultaneously as from the *stage-of-life* device. The reason is, he claimed, that *cry* is, in his terminology, a *category-bound* activity, being bound to the category *baby* in the *stage-of-life* device.

Sacks codified what it is that leads us to hear “baby” as from the *stage-of-life* device in addition to the *family* device, by means of a further hearer’s maxim:

(HM3) If a category-bound activity is asserted to have been done by a member of some category where, if that category is ambiguous (i.e., is a member of at least two devices), but where, at least for one of those devices, the asserted activity is category-bound to the given category, then hear at least the category from the device to which it is bound.

The final part of Sacks’ analysis that I shall consider here concerns the way that an observer describes a particular scene. For instance, if you were to observe a very small human crying, you would most likely describe what you saw as “A baby is crying,” or some minor variant thereof. You are far less likely to say “A person is crying” or, even if you could identify the gender of the baby as female, “A girl is crying'.” Again, if you subsequently saw a woman pick up that baby, and if that woman looked about the age to be the baby’s mother, you would probably describe what you saw as “It’s mother picked it up.” You are less likely to say “A woman picked it up.”

Sacks explained the first of these observations, the use of the phrase “the baby”, by means of the following viewer's maxim:

(VM1) If a member sees a category-bound activity being done by a member of a category to which the activity is bound, then see it that way.

Since the activity of crying is category-bound to the category *baby* in the *stage-of-life* device, this is the natural way to see and to describe the activity, whenever such a way of seeing and describing is possible.

Turning to the remaining set of observations, there are social norms that govern, or can be seen to govern, the actions of members of the society, and one such norm is that a mother will comfort her crying baby. This is a very powerful social norm, and society generally demonstrates strong disapproval for a mother who fails to conform to it. Sacks’ point was that in addition to governing behavior—where by “governing” we may mean nothing more than that the norm serves to describe a normal way of behaving—norms fulfill a further role; namely, viewers use norms to provide some of the orderliness of the activities they observe. In this case, we may capture such a use of a norm by means of a second viewer’s maxim:

(VM2) If one sees a pair of actions which can be related via a norm that provides for the second given the first, where the doers can be seen as members of the categories the norm provides as proper for that pair of actions, then (a) see that the doers are such members, and (b) see the second as done in conformity with the norm.

By means of VM2(a), the viewer sees the person who picks up the baby as the baby's mother, provided it is possible to see it thus, and moreover, by VM2(b), takes it that this action is performed in accordance with the norm that says that mothers comfort their crying babies.

Why this will be viewed as mathematics

Neither Grice's analysis nor Sacks' can be classified as mathematics. They are, however, clearly inspired by mathematics, and exhibit a definite mathematical flavor. They do, I believe, represent first attempts to develop, in a bottom-up fashion, analyses that may (and I think will) eventually lead to analyses that will in time (perhaps when they meet top-down approaches) be viewed as mathematics. Their approaches work because, for all the nondeterminism inherent in people, their behavior nevertheless exhibits regular, repeated, and for the most part predictable patterns. And that makes them amenable to a mathematical analysis, *providing only that you broaden your conception of mathematics to include the study of such patterns*. Today, we are not yet at the point of making such a leap. But as a result of the increasing production of such analyses, I believe that a hundred years from now, such reasoning will indeed be thought of as mathematics, just as probability theory is today, despite the fact that the ancient Greeks believed—and wrote—that a study of chance events was not within the realm of a mathematics.

The new mathematics I am speaking of is not yet here. At best, what we see today are blends of (contemporary) mathematics with other methods of reasoning and analysis, such as Bayesian reasoning or Black-Scholes analysis. More commonly, there is not even a meeting of the two approaches, let alone a blending. The top down approach involves the development of mathematical *models* (such as Chomksy's formal grammar) that reflect the domain of interest in a fairly crude fashion. The mathematical models typically ignore much of the complexity of the domains they are intended to model. Meanwhile, he bottom-up approach, which tries to reflect the full complexity of the domain (for example, Grice's maxims), can perhaps best be described as the adoption of a mathematical *approach* to an analysis, as opposed to an application of mathematics.

So why do I think that what we are seeing are the initial steps toward what will in due course be viewed as a *bona fide* part of mathematics? First, there is historical precedence. Mathematicians have always tried to extend their discipline to apply to new areas. The ancient Greeks tried to develop a "mathematical analysis" of language and reasoning (the Stoic and Aristotelian schools of Logic). Gottfried Leibniz in the 17th century and George Boole in the 19th further tried to develop a mathematics of language and reasoning, leading eventually to a rich branch of mathematics known as mathematical logic, developed in the first half of the 20th century. Then there are the examples of Bernoulli, Bayes, and Black–Scholes–Merton I considered earlier.

Mathematical logic eventually came to be regarded as a fully-fledged branch of mathematics. The work of Bernoulli and Bayes in probability theory is also generally viewed as mathematics. Black-Scholes theory is usually referred to as “financial mathematics”, a terminology that arguably reflects acceptance of it as a branch of mathematics. Certainly many college and university mathematics departments offer course called “Financial Mathematics” and count completion of such a course towards a mathematics degree.

On the other hand, statistics and computer science, for all their heavy dependence on mathematics, are generally classified as outside of mathematics. This despite the fact that, in a great many colleges and universities, both are taught (only) in the Mathematics Department.

Moreover, statisticians, in particular, are often seen in large numbers at major mathematics conferences.

All this seems to suggest a certain degree of indecision on the issue.

However, the case I am trying to make in this essay is based not so much on historical precedent, but rather on a highly pragmatic view of the discipline as a human activity carried out by a human community. I titled my article “What will count as mathematics in 2100?” As I see it, the question then is, who gets to say what is and what is not to be called “mathematics”?

One answer is to leave it to the profession to *define* the answer — a solution that mathematicians (of all people) might be expected to find attractive. This was done for several years by the International Mathematical Union, in connection with the publication of the (now abandoned) annual *World Directory of Mathematicians*. The general criterion of admissibility to the Directory was two articles reviewed in *Mathematical Reviews*, *Referativnyi Zhurnal*, or *Zentralblatt fuer Mathematik* over the preceding five years, or the publication of five papers reviewed in these journals at any time. That certainly provides a precise, profession-certified definition of what counts as a “mathematician”, and by extension yields a definition of what counts as mathematics. The problem is, the definition excludes the vast majority of people who earn their living as “professors of mathematics” at colleges and universities all over the world, and an even great number of people in industry, government, and elsewhere in society, who spend large parts of every day “doing math”.

Another way to provide an answer is to abstract one from mathematical practice. Mathematics is the science of patterns. The mathematician extracts—or abstracts—patterns either from the world or from a mental discipline (including mathematics itself) and studies them. What makes such a study mathematics, as opposed to some other discipline, is the high degree of abstraction of both the patterns and the way they are studied. Mathematical patterns are the structural skeletons that lie beneath the world. Such is their abstraction that the only tool available for their study is pure human reasoning. Except in an occasional, peripheral way, observation and experiment are of little use. When applied to the deterministic, physical world, mathematics leads to results that can be given with absolute certainty. When we try to use the mathematical method to study nonmathematical objects—in particular, people—however, the results become less certain. But, I would suggest, that is the nature of the *domain*; it is not caused by the *method* we use to analyze it. It is only a matter of time, I believe, before we become so familiar with mathematically-inspired analyses such as the examples I have presented here that the mathematical community — or at least a substantial portion thereof — begins to accept them as a regular part of the discipline. I believe we put ourselves on that path the moment we took on board the definition of mathematics as the science of patterns.

Contemporary mathematics may have declared its goal to be the formulation of precise definitions and axioms and the subsequent deduction of theorems, but that is a fairly recent phenomenon, and likely just a passing fad. It is also, I suggest, a foolish one that serves no one particularly well. In the last hundred years or so, mathematics has parted company with (and even distanced itself from) theoretical physics, statistics, and computer science, and even split internally into “pure” and “applied” mathematics, only to find that some of the most exciting and productive new developments within core mathematics (i.e., those parts that have not been cast out) have come from those other disciplines.

Of course, there is a difference between “mathematics pursued for its own sake” and mathematics carried out in the course of studying some other domain. Of course there is a difference between formal, proof-oriented mathematics based on axioms, and applications of mathematics such as the work of Chomksy or of Black and Scholes. And of course there will always be some individuals who prefer to focus entirely on the pursuit of axiom-theorem type mathematics. The question is, will that relatively small group be able to define what the word “mathematics” means?

I believe the answer is no. In my view, what counts as mathematics will be determined not (solely) by one particular organization within mathematics (such as the IMU). Rather, I believe it will be determined for the most part on sociological grounds, by society as a whole. What classifies as “mathematics” will be determined by what gets done in university “mathematics departments” and by what society expects from people it categorizes as “mathematicians.” For it is society as a whole that, one way or another, provides both the environment in which mathematics is done and the funds for its pursuit. Since the main importance of mathematics for humanity as a whole is the role it plays in human understanding, when the main objects of that understanding are *people*, as well as deterministic physical systems, as will be the case in the coming century, mathematics will change.

I should finish by observing that I am by no means the first to come to such a conclusion. For instance, the late Gian-Carlo Rota of MIT wrote (see Rota, Schwartz & Kac 1985):

> *Sometime, in a future that is knocking at our door, we shall have to retrain ourselves or our children to properly tell the truth. The exercise will be particularly painful in mathematics. The enrapturing discoveries of our field systematically conceal, like footprints erased in the sand, the analogical train of thought that is the authentic life of mathematics. Shocking as it may be to a conservative logician, the day will come when currently vague concepts such as motivation and purpose will be made formal and accepted as constituents of a revamped logic, where they will at last be allotted the equal status they deserve, side-by-side with axioms and theorems.*

If you take Rota's phrase "made formal and accepted as constituents of a revamped logic" as suggesting formalized mathematics in the sense currently understood in mathematics—as some have done—then Rota's suggestion seems doomed to fail, as I am sure Rota himself would have agreed. But if you interpret it in the light of a changing conception of what counts as mathematics, as I do and as I suspect Rota intended, then it predicts a rosy future for mathematics.

References

Fischer Black & Myron Scholes (1973), The pricing of options and corporate liabilities, *Journal of Political Economy*, Vol. 81, pp.637-59.

Noam Chomsky (1957), *Syntactic Structures*, The Hague/Paris: Mouton.

H. Paul Grice (1975), Logic and Conversation. In Peter Cole and Jerry L. Morgan (Eds.) *Syntax and Semantics, Vol. 3, Speech Acts*, New York: Academic Press.

Richard Montague (1974), The Proper Treatment of Quantification in Ordinary English. In Richmond Thomason (Ed.), *Formal Philosophy: Selected Papers of Richard Montague.* New Haven: Yale University Press.

Gian Carlo Rota, Jacob Schwartz & Mark Kac (Eds.) (1985), *Discrete Thoughts*, Basel: Birkhäuser

Harvey Sacks (1972), On the analyzability of stories by children. In: John Gumperz, & Del Hymes (Eds.) *Directions in Sociolinguistics: the Ethnography of Communication*. New York: Rinehart & Winston.