

Magic, Mystery, and Matrix

Edward Witten

In the twentieth century, the quest for deeper understanding of the laws of nature has largely revolved around the development of two great theories: namely, general relativity and quantum mechanics.

General relativity is, of course, Einstein's theory according to which gravitation results from the curvature of space and time; the mathematical framework is that of Riemannian geometry. While previously spacetime was understood as a fixed arena, given *ab initio*, in which physics unfolds, in general relativity spacetime evolves dynamically, according to the Einstein equations. Part of the problem of physics, according to this theory, is to determine, given the initial conditions as input, how spacetime will develop in the future.

The influence of general relativity in twentieth-century mathematics has been clear enough. Learning that Riemannian geometry is so central in physics gave a big boost to its growth as a mathematical subject; it developed into one of the most fruitful branches of mathematics, with applications in many other areas.

While in physics general relativity is used to understand the behavior of astronomical bodies and the universe as a whole, quantum mechanics is used primarily to understand atoms, molecules, and subatomic particles. Quantum theory has had

a much more complex history than general relativity, and in some sense most of its influence on mathematics belongs to the twenty-first century. The quantum theory of particles—which is more commonly called nonrelativistic quantum mechanics—was put in its modern form by 1925 and has greatly influenced the development of functional analysis, and other areas.

But the deeper part of quantum theory is the quantum theory of fields, which arises when one tries to combine quantum mechanics with special relativity (the precursor of general relativity, in which the speed of light is the same in every inertial frame but spacetime is still flat and given *ab initio*). This much more difficult theory, developed from the late 1920s to the present, encompasses most of what we know of the laws of physics, except gravity. In its seventy years there have been many milestones, ranging from the theory of “antimatter”, which emerged around 1930, to a more precise description of atoms, which quantum field theory provided by 1950, to the “standard model of particle physics” (governing the strong, weak, and electromagnetic interactions), which emerged by the early 1970s, to new predictions in our own time that one hopes to test in present and future accelerators.

Quantum field theory is a very rich subject for mathematics as well as physics. But its development in the last seventy years has been mainly by physicists, and it is still largely out of reach as a rigorous mathematical theory despite important efforts in constructive field theory. So most of its impact on mathematics has not yet been felt. Yet in many active areas of mathematics, problems are

Edward Witten is professor of physics at the Institute for Advanced Study. His e-mail address is witten@ias.edu. This article is based on the Josiah Willard Gibbs Lecture given at the Joint Meetings in Baltimore in January 1998. The author's work is supported in part by NSF Grant PHY-9513835.

studied that actually have their most natural setting in quantum field theory. Examples include Donaldson theory of four-manifolds, the Jones polynomial of knots and its generalizations, mirror symmetry of complex manifolds, elliptic cohomology, and many aspects of the study of affine Lie algebras.

To a certain extent these problems are studied piecemeal, with difficulty in understanding the relations among them, because their natural home in quantum field theory is not now part of the mathematical theory. To make a rough analogy (Figure

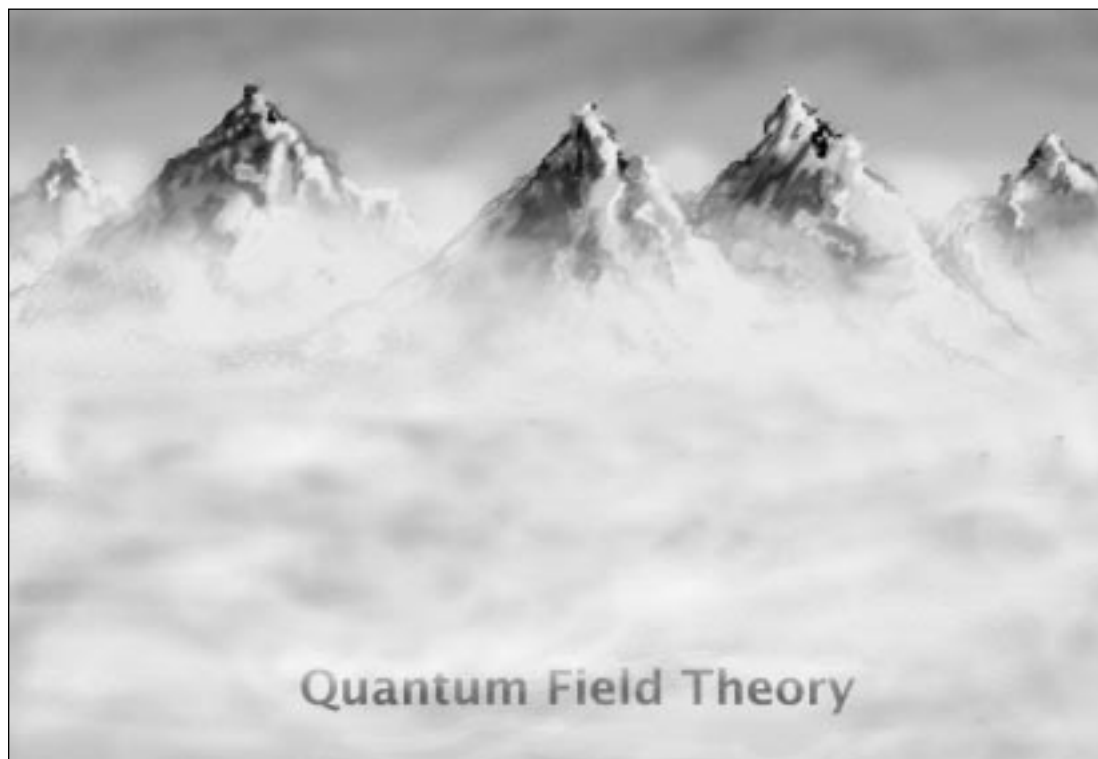


Figure 1.

1), one has here a vast mountain range, most of which is still covered with fog. Only the loftiest peaks, which reach above the clouds, are seen in the mathematical theories of today, and these splendid peaks are studied in isolation, because above the clouds they are isolated from one another. Still lost in the mist is the body of the range, with its quantum field theory bedrock and the great bulk of the mathematical treasures.

So there is one rather safe, though perhaps seemingly provocative, prediction about twenty-first century mathematics: trying to come to grips with quantum field theory will be one of the main themes.

The $1/r^2$ Singularity

To see a little further than this, we must discuss quantum mechanics in a little more depth. The origin and subsequent development of quantum mechanics depended a lot on the “inverse square law” of gravity and electricity. The gravitational forces between two masses M_1 and M_2 separated a distance r is

$$-\frac{GM_1M_2}{r^2}$$

(with G being Newton’s constant), and the electrical force between two charges q_1 and q_2 separated a distance r is likewise

$$\frac{q_1q_2}{r^2}.$$

For elementary particles one typically has $q_1q_2 > GM_1M_2$, which is why gravity can generally be ignored on an atomic scale and below, but for

astronomical bodies typically $GM_1M_2 > q_1q_2$, so gravity dominates on large scales.

Obviously, the inverse square law means that the force becomes infinite for $r \rightarrow 0$. This singularity did not cause great difficulties for Newton, since (for instance) the Moon was always at a safe distance from the Earth, far from $r = 0$. However, once the electron and atomic nucleus were discovered almost a century ago, the $1/r^2$ singularity did become a severe problem. A simple calculation based on nineteenth-century physics showed that, because of the strong force at small r , the electron should spiral into the nucleus in about 10^{-9} seconds. This was obviously not the case.

To cure this problem, quantum mechanics was invented. In quantum mechanics the position x and momentum p of a particle do not commute, but obey Heisenberg’s relation

$$[p, x] = -i\hbar,$$

with \hbar being Planck’s constant. This relation gives a sort of “fuzziness” to the electron and other particles. Because of this fuzziness, one never really gets to $r = 0$, and the problem is averted.

What I have just described is nonrelativistic quantum mechanics—or quantum mechanics of particles, as I called it before. This theory was developed by about 1925 and has long since been more or less assimilated mathematically. The whole theory of elliptic operators on manifolds is a kind of mathematical counterpart of nonrelativistic quantum mechanics; group representation theory is also a close cousin.

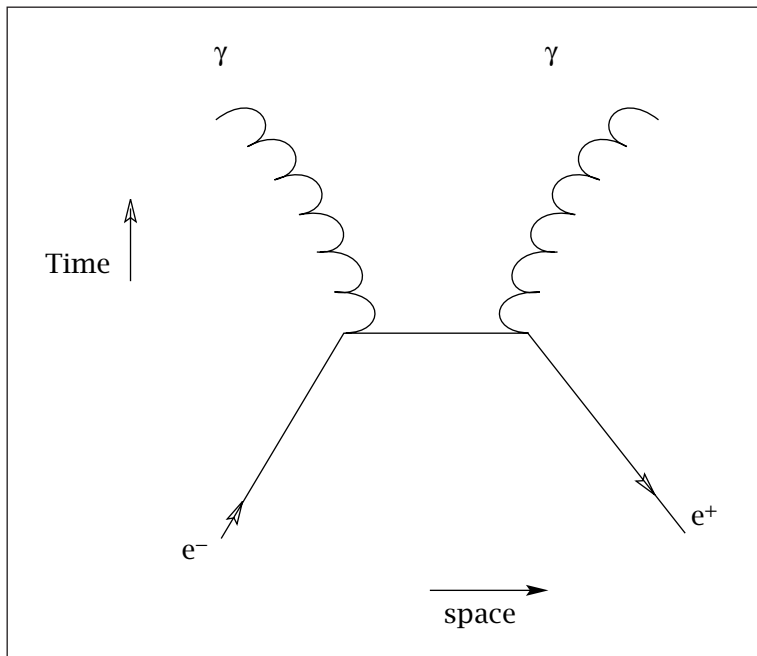


Figure 2. A spacetime history with an electron e^- and its antiparticle, the positron e^+ , annihilating into two photons or basic units of light (labeled by γ). If the direction of time is reversed, one gets a process in which incoming light waves combine to create a charged particle-antiparticle pair. Such phenomena occur all the time in quantum field theory.

Quantum Field Theory

But trying to allow for special relativity makes things much more challenging. In special relativity one cannot assume the “instantaneous action at a distance” that is implicit in the inverse square laws of gravity and electricity. Instead, the force must be mediated by a field, and consistency of the whole setup requires that an uncertainty relation analogous to the Heisenberg formula $[p, x] = -i\hbar$ must be applied to the field. Then things become much more complicated and much more interesting. From the uncertainty relation one deduces that the field comes in “quanta”, which are observed as particles of a new kind—photons in the case of the electromagnetic field. The more familiar particles, like the electron, must likewise be reinterpreted as quanta of a field. One soon learns that, like classical electromagnetic waves, these quanta can be created and annihilated (Figure 2). This leads to the concept of antimatter and the prediction of matter-antimatter creation and annihilation. By this time one is living in a world that is much more surprising and interesting and certainly much more challenging mathematically.

Though quantum mechanics was invented because of the $1/r^2$ singularity, it turned out that once special relativity was included, quantum mechanics did not automatically cure all of the problems associated with that singularity. Much of the development of physics since 1930 had to do with the $1/r^2$ problem in the light of quantum me-

chanics plus special relativity. Among the milestones, some of which I alluded to before, were the following:

- By about 1950, renormalization theory and quantum electrodynamics gave a much more precise theory of electrons and atoms.

- In 1967–73 nonabelian gauge theories were incorporated in the description of nature (giving the electroweak part of the standard model) to overcome the problem of the $1/r^2$ singularity in the case of the weak interactions.

- In 1973 asymptotic freedom of nonabelian gauge theories was discovered and used to overcome and fully tame the $1/r^2$ problem in the case of the nuclear force. This also completed the construction of the standard model.

Those last developments set the stage for a new kind of interaction between quantum field theory and geometry. Nonabelian gauge theories, supplemented in time by other ingredients that I have not yet mentioned, notably supersymmetry and string theory, led physicists to gradually ask new kinds of questions that involved geometrical concepts and techniques not previously used in physics. In time it was realized that things could be turned around and that the quantum field theory methods could be used to draw inferences about geometry. And so it is that although quantum field theory is a rather old subject, its mathematical influence is in many respects rather recent and still lies mainly in the future.

Quantum Gravity

The developments that I have already mentioned, leading to the standard model of particle physics, put most of the known phenomena in physics, except gravity, more or less under one roof. The main hurdle that remains is to include gravity, but this involves problems of a quite different nature. At first sight, gravity presents us with just another instance of the familiar $1/r^2$ singularity. Gravity and electricity are indeed very similar in many ways, but the relation between them is not nearly as straightforward as is suggested by the fact that in classical physics they are both governed by inverse square laws. Relativistically, for instance, the field equations of electromagnetism (Maxwell’s equations) are linear, while Einstein’s equations for the gravitational field are highly nonlinear. Quantum fuzziness, springing from the uncertainty relation $[p, x] = -i\hbar$, is apparently not enough to deal with the $1/r^2$ singularity in the gravitational force. Overcoming this problem—combining quantum mechanics and gravity—is probably the main obstacle to unifying the forces of nature.

Making sense of quantum gravity is essential as well for addressing many commonplace questions that one might well ask with no special training in physics. Astronomers, for example, see that the universe is expanding today, and as far as we can tell

this expansion began in an explosion, often called the big bang. But contemplation of the big bang may seem to present paradoxes. What started the clocks? What was there before the big bang? Gravity and quantum mechanics were both important near the big bang, so the answers must depend on how gravity and quantum mechanics work together.

Physicists learned rather unexpectedly, beginning in the early 1970s, that the problem of quantum gravity could be overcome by introducing a new sort of fuzziness. One replaces “point particles” by “strings”. Of course, the point particles and strings must both be treated quantum mechanically. Quantum effects are proportional to Planck’s constant \hbar , and stringy effects are proportional to a new constant α' (equal to approximately $(10^{-32} \text{ cm})^2$) that determines the size of strings. In this theory stringiness and quantum uncertainty both contribute to smearing things out; together they tame the $1/r^2$ singularity of gravity.

If string theory is correct, then α' is just as fundamental in physics as \hbar , and its effects are at least as interesting. The \hbar and α' deformations both involve fundamental new tools and ideas in geometry. About the \hbar deformation we have ample experience and fairly extensive nonrigorous knowledge concerning some of the geometrical applications, though, as I explained before, the mathematical development still lies largely in the future. The α' deformation is far more mysterious and challenging even for physicists, as the basic tools and concepts have not yet been unearthed. Seeking to do so is perhaps the most exciting adventure in theoretical physics for the next few decades. The mathematical questions posed by the \hbar deformation are at least beginning to be asked, though answers still lie mainly ahead, but the equally exciting mathematical questions associated with the α' deformation are for the most part not yet even being asked. The reason for this is simply that the basic prerequisite for understanding what the α' deformation is supposed to mean is a thorough familiarity with the \hbar deformation, and this is not yet available mathematically.

The idea of replacing point particles by strings sounds so naive that it may be hard to believe that it is truly fundamental. But in fact this naive-sounding step is probably as basic as introducing the complex numbers in mathematics. If the real and complex numbers are regarded as real vector spaces, one has $\dim_{\mathbf{R}}(\mathbf{R}) = 1$, $\dim_{\mathbf{R}}(\mathbf{C}) = 2$. The orbit of a point particle in spacetime (Figure 3) is one-dimensional and should be regarded as a real manifold, while the orbit of a string in spacetime is two-dimensional (over the reals) and should be regarded as a complex Riemann surface. Physics

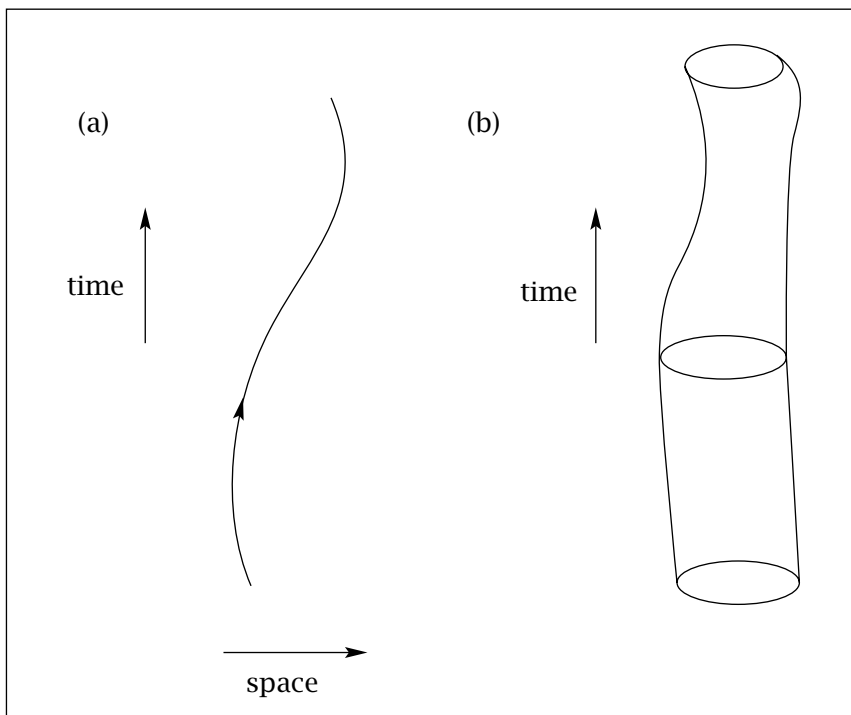


Figure 3. The spacetime orbit of a point particle (a) or a string (b) is a manifold of real dimension 1 or 2.

without strings is roughly analogous to mathematics without complex numbers.

The String Theories

The requirements of quantum mechanics plus special relativity are so tight that historically constructing string theories was very difficult. The conditions that must be obeyed are highly overdetermined. A vast effort went into the construction of string theories, and by the time the dust cleared in 1984–85 it was found that there were five of them. They differ by very general properties of the strings:

- In two theories (the Type IIA and Type IIB theories, which differ by whether there is invariance under reversal of the orientation of spacetime) the strings are closed and oriented and are electrical insulators.
- In two theories (the heterotic superstrings with gauge groups $SO(32)$ and $E_8 \times E_8$) the strings are closed, oriented, and superconducting.
- In the last case (Type I) the strings are unoriented and insulating and can have boundaries, in which case they carry electric charges on their boundaries.

Because there are so few string theories, the general framework of string theory makes certain general predictions that are out of reach without string theory:

1. *Gravity.* Each of the five string theories predicts gravity (plus quantum mechanics): that is, these theories predict a structure that looks just like general relativity at long distances, with corrections (unfortunately unmeasurably small in

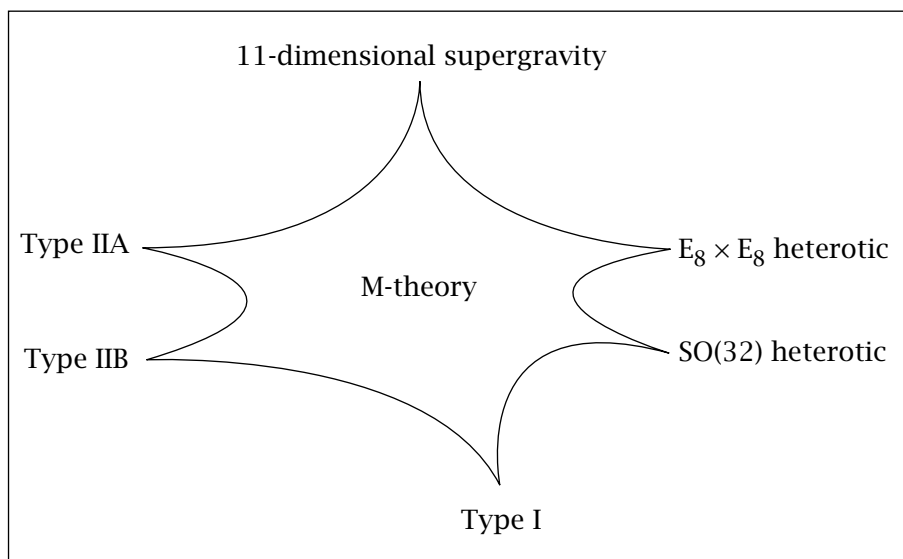


Figure 4. The five string theories—and a wild card, eleven-dimensional supergravity, that has proved to be important in getting a systematic understanding—are now understood as different limiting cases of a more comprehensive (and little understood) theory known as *M*-theory. The figure is meant to suggest a family of physical situations that are possible in *M*-theory. With some oversimplification one can think of the parameters in the figure as \hbar and α' .

practice) proportional to α' . This is very striking, since, as I have stressed, standard quantum field theory makes gravity impossible. It is the single most important reason for the intensive study of string theory in the last generation.

2. *Nonabelian Gauge Symmetry.* The second general prediction is nonabelian gauge symmetry (again with unmeasurably small corrections proportional to α'), which is of course the bread and butter of particle physics.

3. *Supersymmetry.* The last general prediction is “supersymmetry”, a subtle new kind of symmetry of elementary particles. We do not yet know if nature is supersymmetric, but there are hints (for instance, from accurate measurements of the low-energy gauge couplings) that it is. There is a good chance that we will know for certain from accelerator experiments within about a decade. The fact that we do not yet really know if it is right means that supersymmetry (which historically was discovered at least in part because of its role in string theory) is a genuine prediction of string theory, while gravity and nonabelian gauge symmetry (which were already known before they were seen to be consequences of string theory) might better be called postdictions.

To properly explain supersymmetry requires assuming some familiarity with quantum field theory and so is beyond our scope here. But as a very rough analogy, supersymmetric quantum theory is to ordinary quantum theory as differential forms on a manifold are to functions on a manifold. A very large fraction of geometrical applications of quantum field theory found in the eighties and nineties depend on supersymmetry. (Examples in-

clude the supersymmetric proofs of the positive energy theorem, the Atiyah-Singer index theorem, and the Morse inequalities, and the quantum field theory approaches to elliptic cohomology and to Donaldson theory.) This, along with its beauty and the boost its discovery would give to string theory, is yet another reason to hope that supersymmetry will be found! Surely, if supersymmetry is confirmed in accelerators, mathematical attention will be focussed on this fruitful branch of quantum field theory roughly as the discovery of general relativity focussed attention on Riemannian geometry.

Apart from the general predictions that I have stressed, string theory also leads in a simple way to elegant and qualitatively correct models that combine quantum gravity and the other known forces in nature, recovering the main features of the standard model. To improve

these constructions further, the most vital need is probably to understand the vanishing (or extreme smallness) of the cosmological constant (the energy density of the vacuum) after supersymmetry breaking. That remains out of reach.

Although physicists do not have any systematic understanding of the new geometrical ideas associated with the α' deformation, powerful methods using two-dimensional conformal field theory are available for exploring some of the associated phenomena. In the late 1980s and early 1990s much effort in string theory was focussed on describing some of these phenomena. An example is mirror symmetry, a relation between two spacetimes that are different in classical geometry but are equivalent for $\alpha' \neq 0$. This symmetry has attracted much mathematical interest because it has striking consequences, some of which can be extracted from their natural conformal field theory setting and stated in isolation. Closely related is the phenomenon of topology change. In general, in string theory the question, What is the topology of spacetime? does not make sense, because in general for $\alpha' \neq 0$ classical ideas in geometry are not valid. But in a suitable limit, upon varying a parameter, classical ideas may be a good approximation. It was found that one can perfectly well have a family of string theory states depending on a real parameter t that interpolate between two different spacetimes in the following sense. For $t \rightarrow \infty$, classical ideas in geometry are a good approximation, and one observes a spacetime X . For $t \rightarrow -\infty$, classical geometry is again a good approximation and one observes a different (and perhaps topologically

distinct) spacetime Y . Somewhere in between large positive t and large negative t one passes through a “stringy” region in which classical geometry is not a good description and the interpolation from X to Y takes place.

M-Theory

While understanding the new geometrical ideas that prevail for $\alpha' \neq 0$ remains in all likelihood a problem for the next century, the problem has lately been recast in a much wider context. For years the existence of five string theories, though it represented a dramatic narrowing of the possibilities that existed in prestring physics, posed a puzzle. It is rather strange to be told that there is a rich new framework for physics which unifies quantum mechanics and gravity and that in this new framework there are five possible theories. If one of those theories describes our universe, who lives in the other four worlds?

By learning something about what happens when α' and \hbar are *both* nonzero, we have learned a very satisfying answer to this question. The five string theories traditionally studied are different limiting cases of one richer and still little-understood theory. For $\hbar = 0$ these theories are really different, but with \hbar and α' both nonzero one can interpolate between them. The relation between them (Figure 4) is rather like the relation between the classical spacetimes X and Y mentioned two paragraphs ago. These are distinct in classical geometry—that is, for $\alpha' = 0$ —but for $\alpha' \neq 0$ they are two different limiting cases of a more subtle structure.

The richer theory, which has as limiting cases the five string theories studied in the last generation, has come to be called *M*-theory, where *M* stands for magic, mystery, or matrix, according to taste. The magic and mystery are clear enough, while “matrix” refers to a new noncommutativity, roughly analogous to $[p, x] = -i\hbar$ but very different, that seems to enter the theory. Physicists and mathematicians are likely to spend much of the next century trying to come to grips with this theory.

Reading Suggestions: For an introduction to quantum field theory and perturbative string theory, the reader might consult *Quantum Fields and Strings: A Course for Mathematicians*, P. Deligne, P. Etinghof, D. Freed, L. Jeffrey, D. Kazhdan, D. Morrison, and E. Witten, eds. (American Mathematical Society, to appear). A recent string theory text (for physicists) treating some of the nonperturbative as well as perturbative developments is *String Theory*, Vols. I and II, by Joseph Polchinski (Cambridge University Press, 1998).