Euclid's Windows and our Mirrors

Review of *Euclid's Window* **by** *Leonard Mlodinow*

This is a shallow book on deep matters, about which the author knows next to nothing. The concept of the book is appealing: a popular review of geometrical notions from Euclid to Einstein as background to contemporary string theory with comments on the related intellectual history and portraits of some principal figures: Descartes, Gauss, Riemann and Einstein. Unfortunately the author is indifferent to mathematics, has only approximate notions of European history, and no curiosity about individuals. Famous names serve only as tags for the cardboard figures that he paints. Disoriented by ideas and by individuals whose feelings and behavior are not those of late twentieth-century America, he attempts to hide his confusion by an incessant, sometimes tasteless, facetiousness, almost a nervous tick with him, by railing at or mocking his pretended dolts or villains, Kant, Gauss's father or Kronecker, or by maudlin attempts to turn his heros into victims. There would be little point in reviewing the book, were it not that the germ of an excellent monograph is there that, in competent and sensitive hands, could have been read with pleasure and profit by students, mature mathematicians, and curious laymen. As a member of the second group who knows scarcely more than the author about the material, I certainly found it an occasion to reflect on what I would have liked to learn from the book and, indeed, an occasion to discover more about the topics discussed, not from the book itself, but from more reliable sources.

String theory itself or, better and more broadly, the conceptual apparatus of much of modern theoretical physics, above all of relativity theory, statistical physics and quantum field theory, whether in its original form as quantum electrodynamics, or as the basis of the standard theory of weak and strong interactions, or as string theory, is mathematics, or seems to be, although often not mathematics of a kind with which those with a traditional training are very comfortable. Nonetheless many of us would like to acquire some genuine understanding of it and, for students especially, it is a legitimate object of curiosity or of more ambitious intellectual aspirations.

Mlodinow was trained as a physicist, and, at the level at which he is working, there is no reason, except perhaps his rather facile condemnation of Heisenberg, to fault his chapter on string theory, the culmination of the book. It is a brief rehearsal, larded with low humor, of the standard litany: the uncertainty principle; the difficulty of reconciling it with the differential geometry of relativity; particles and fields; Kaluza-Klein and the introduction of additional dimensions; the function of "strings" as carriers of multiple fields and particles; supersymmetry; and, finally, M-branes that permit the passage from one form of the theory to another.

In his introduction and in an epilogue Mlodinow expresses vividly and passionately his conviction that geometry is the legacy of Euclid and string theorists his heirs. Mathematicians, to whom this review is addressed, will recall that there is more in Euclid than geometry: Eudoxus's theory of proportion; the irrational; and primes. Since the last two are central to the modern theory of diophantine equations, there are other claims on the heritage, but they need no defense here. We are concerned with the geometry; with it alone we have our hands full.

A reservation that is more in need of expression is that, with their emphasis on string theory or, better, the geometrical consequences of quantum field theory, mathematicians are in danger of shortchanging themselves. These consequences, especially the dynamical methods – dynamical in the sense of dynamical systems – used to deduce them, methods discovered, I believe, largely by Witten, are of great appeal and undoubtedly very deep. They are certainly worthy of the careful attention of mathematicians; but, as a community, we should well be trying to address in a coherent way all dynamical questions, both analytic and geometric, raised by or related to renormalization in statistical mechanics and in field theory.

Although these questions as a whole lie athwart Mlodinow's concept, it is difficult when reading the last chapter of his book not to reflect on them and on the current relation between mathematics and physics. So after finishing with other aspects of the book, about which there is a good deal to be said, I shall return to these matters.

History and Biography

Euclid. The background metric, thus the general theory of relativity, is a feature of string theory that is not present in most other field theories. There are several evident milestones on the way from ancient mathematics to Einstein's theory: Euclid's account of plane geometry; Descartes's advocacy of coordinates to solve specific geometrical problems; the introduction of curvature by Gauss and the discovery of noneuclidean geometries; Riemann's conceptions of higher-dimensional geometries and his criterion for flatness; Einstein's equations for general relativity. To isolate these five developments, each a major moment in intellectual history, as the themes of a single essay on mathematics was brilliant; to realize the concept an enormous challenge, beyond me, beyond most readers of the Notices, and certainly far beyond the author, locked in the present, upon which for him all windows open, and dazzled by his own flippancy.

Euclid's *Elements* are, above all, a window on themselves and on Greek mathematics. Difficult to appreciate without commentary, they could never have been, in spite of tradition, suitable as independent reading for schoolboys. In Mlodinow's first chapter, the one on Euclid, the mathematics is given very short shrift; the author prefers trivial puzzles to real mathematics. He presents Euclid's five postulates, including the fifth, or parallel, postulate in Euclid's form (two lines falling on a given line in such a way that the sum of the interior angles on the same side is less than two right angles necessarily meet) and, in addition, in the form known as Playfair's axiom (a unique parallel to a given line can be drawn through any point), probably because Playfair's axiom is more familiar to him from high-school. If our concern is with Euclid as Einstein's predecessor, then it is Euclid's form that is pertinent, for it expresses the flatness. What the student or the layman needs from this chapter is an explanation of the fifth postulate's relation to flatness: to the basic property of a triangle that the sum of its interior angles is π and especially to the existence of similar figures, thus to what one might call a little pretentiously, imitating the current jargon, the conformal invariance of Euclidean geometry. Even the mature mathematician may enjoy recalling these deep, important, and yet elementary, logical relations, for not all of us have taken the time to think through the manifold concrete implications of noneuclidean geometry. It appears, however, that the author has not even read Heath's comments and does not appreciate how flatness manifests itself in the simple geometric facts that we know almost instinctively, so that with all the impudence of the ignorant, he can, later in the book, mock Proclus, who attempted, as other important mathematicians, like Legendre, were still doing centuries later, to prove the postulate, or Kant, whose philosophical imagination was unfortunately inadequate to the mathematical reality.

Otherwise the space in the first chapter is largely devoted to tales suitable for children, or sometimes not so suitable for children, as the author has a penchant for the lewd that he might better have held in check. He trots out the old war-horses Thales and Pythagoras and a new feminist favorite, Hypatia. Cajori, in his *A History of Mathematics*, observes that the most reliable information about Thales and Pythagoras is to be found in Proclus, who used as his source a no longer extant history by Eudemus, a pupil of Aristotle. Thales and Pythagoras belong to the sixth century BC, Eudemus to the fourth, and Proclus to the fifth century AD. Common sense suggests that there is considerable room for distortion, intentional or unintentional, in information that has been transmitted over a thousand years. This did not stop Cajori and many other historians of science from using it. Nor does it stop the present author, who even adds some gratuitous speculation of his own: could the merchant Thales have traded in the leather dildos for which Mlodinow claims Miletus was known?

Although Neugebauer writes in *The Exact Sciences in Antiquity*, "It seems to me evident, however, that the traditional stories of discoveries made by Thales or Pythagoras must be regarded as totally unhistorical", there may be a place in popular accounts for the myths attached to them, but not at the cost of completely neglecting the responsibility of introducing the reader, especially the young reader, to some serious notions of the history of science, or simply of history. I have not looked for a rigorous account of Thales, but there is a highly regarded account of Pythagoras, *Weisheit und Wissenschaft*, by the distinguished historian Walter Burkert in which the reality is separated from the myth, leaving little, if anything, of Pythagoras as a mathematician. One's first observation on reading this book is that it is almost as much of a challenge to discover something about the Greeks in the sixth century as to discover something about physics at the Planck length (10^{-33}cm., the characteristic length for string theory). A second is that most of us are much better off learning more about the accessible philosophers, as a start, Plato and Aristotle, or about Hellenistic mathematics, that the earlier mythical figures are well left to the specialists. The third is that one should not ask about the scientific or mathematical achievements of Pythagoras but of the Pythagoreans, whose relation to him is not immediately evident. Burkert's arguments are complex and difficult, but a key factor is, briefly and imprecisely, that for various reasons Plato and the Platonists ascribed ideas that were properly Platonic to Pythagoras, who in fact was a religious rather than a scientific figure.

Whether as mathematician, shaman, or purveyor to Ionian and Egyptian sex-shops, neither Thales nor Pythagoras belongs in this essay; nor does Hypatia. Since Hypatia is a figure from the late fourth century AD, it is easier to separate the myth from the reality, and there is an instructive monograph *Hypatia of Alexandria* by Maria Dzielska that does just this. Mlodinow refers to the book but there is no sign that he has read it. If he has, he ignored it!

Mlodinow's book is short, and the space is largely taken up with material that is irrelevant or false, and often both. Much of the reliable information about Hypatia comes from the letters of Synesius of Cyrene, bishop of Ptolemais. An Alexandrian philosopher, a late Platonist, and mathematician of some repute, the daughter of the mathematician Theon, Hypatia was renowned for her wisdom, erudition and virtue. Of some political influence in the city, an ally of the prefect Orestes, she was brutally murdered in 415 at, according to Dzielka, the age of sixty (this estimate is not yet reflected in standard references) by supporters of his rival, the bishop Cyril. Although now a feminist heroine, which brings with it its own distortions, Hypatia first achieved mythical status in the early eighteenth century in an essay of John Toland, for whom she was a club with which to beat the Catholic church. His lurid tale was elaborated by Gibbon, no friend of Christianity, in his unique style: ". . . her learned comments have elucidated the geometry of Apollonius . . . she taught . . . at Athens and Alexandria, the philosophy of Plato and Aristotle . . . In the bloom of beauty, and in the maturity of wisdom, the modest maid refused her lovers . . . Cyril beheld with a jealous eye the gorgeous train of horses and slaves who crowded the door of her academy . . . On a fatal day . . . Hypatia was torn from her chariot, stripped naked, . . . her flesh was scraped from her bones with sharp oyster-shells . . . the murder . . . an indelible stain on the character and religion of Cyril." This version, in which Hypatia is not an old maid but a young virgin, so that it is a tale not only of brutality but also of lust, is the version preferred by Mlodinow.

There is yet another component of the myth: Hypatia's death and the victory of Cyril mark the end of Greek civilization and the triumph of Christianity. Such dramatic simplification is right up Mlodinow's alley, who from this springboard leaps, as a transition from Euclid to Descartes, into a breezy tourist's account of Europe's descent into the Dark Ages and its resurrection from them, in which, in a characteristic display of ambiguity, the author wants to make Charlemagne out both a dunce and a statesman.

Descartes. Mlodinow would have done well to pass directly from Euclid to Descartes. Both Descartes and Gauss had a great deal of epistolary energy; so a genuine acquaintance with them as individuals is possible. The letters of Gauss especially are often quite candid. A good deal of their mathematics, perhaps all that of Descartes, is also readily accessible without any very exigent prerequisites. Yet Mlodinow relies on secondary, even tertiary, sources, so that his account, having already passed through several hands, is stale and insipid. Moreover, he exhibits a complete lack of historical imagination and sympathy, of any notion that men and women in other times and places might respond to surroundings familiar to them from birth differently than a late twentieth-century sight-seer from New York or Los Angeles. What is even more exasperating is that almost every sentence is infected by the itch to be jocose, to mock, or to create drama, so that a mendacious film covers everything. Without a much deeper and more detailed knowledge of various kinds of history than I possess, there is no question of recognizing each time exactly how veracity is sacrificed to effect. In some egregious instances, with which I was more than usually outraged, I have attempted to analyze his insinuations. It will be apparent that I am neither geometer nor physicist, and not a philosopher or a historian; I make no further apology for this.

Descartes was primarily a philosopher or natural scientist, and only incidentally a mathematician. So far as I can see, almost all the mathematics that we owe to him is in one appendix, *La géométrie*, to *Discours de la méthode*. Anyone who turns to this appendix will discover, perhaps to his surprise, that contrary to what Mlodinow states, Descartes does not employ the method we learned in school and "begin his analysis by turning the plane into a kind of graph". Not at all, Descartes is a much more exciting author, full, like Grothendieck and Galois, of philosophical enthusiasm for his methods. He begins by discussing the relation between the geometrical solution with ruler and compass of simple geometrical problems and the algebraic solution, goes on to a brilliant analysis of the curve determined by a generalized form of the problem of Pappus, an analysis that exploits oblique coordinates, not for all points but for a single one, and chosen not once and for all but adapted to the data of the problem. The analysis is incisive and elegant, well worth studying, and is followed by a discussion of curves in general, especially algebraic curves, and their classification, which is applied to his solution of the problem of Pappus. Descartes does not stop there, but the point should be clear: this is analytic geometry at a high conceptual but accessible mathematical level that could be communicated to a broad public by anyone with some enthusiasm for mathematics. He would of course have first to read Descartes, but there is no sign that Mlodinow regarded that as appropriate preparation.

Although adverse to controversy, even timid, Descartes, a pivotal figure in the transition from the theologically or confessionally organized society to the philosophically and scientifically open societies of the enlightenment, made every effort to ensure that his philosophy became a part of the curriculum both in the United Provinces where he made his home and in his native, Catholic France. In spite of the author's suggestion, his person was never in danger: with independence from Spain, the Inquisition had ceased in Holland and by the seventeenth century it had long been allowed to lapse in France. Atheism was nonetheless a serious charge. Raised by his opponent, the Calvinist theologian Voetius, it could, if given credence, have led to a proscription of his teachings in the Dutch universities and the Jesuit schools of France and the Spanish Netherlands, but not to the stake. The author knows this – as did no doubt Descartes – but leaves, once again for dramatic effect and at the cost of missing the real point, the reader with the contrary impression.

Gauss. It appears that, in contrast to many other mathematical achievements, the formal concept of noneuclidean geometry appeared only some time after a basic mathematical understanding of its properties. This is suggested by the descriptions of the work of Gauss and of earlier and later authors, Lambert in particular, that are found in Reichardt's *Gauß und die nicht-euklidische Geometrie* and by the documents included there. It was known what the properties must be, but their possibility, whether logical or in reference to the natural world, was not accepted. Modern mathematicians often learn about hyperbolic geometry quickly, almost in passing, in terms of the Poincaré model in the unit disk or the upper half-plane. Most of us have never learned how to argue in elementary geometry without Euclid's fifth postulate. What would we do if, without previous experience, we discovered that when the sum of the interior angles of just one triangle is less than π, as is possible when the parallel postulate is not admitted, then there is necessarily an upper bound for the area of all triangles, even a universal length? Would we conclude that such a geometry was totally irrelevant to the real world, indeed impossible? If we were not jaded by our education, we might better understand how even very perceptive philosophers could be misled by the evidence.

An intelligent, curious author would seize the occasion of presenting these notions to an audience to which they would reveal a new world and new insights, but not Mlodinow. What do we have as mathematics from him? Not the thoughts of Legendre, not the contributions of Lambert, not even the arguments of Gauss, taken from his reviews, from his letters, from his notes, nothing that suggests that the heart of the matter, expressed of course in terms of the difference between π and the sum of the interior angles of triangles, is whether the plane is curved. No, he does not even mention curvature in connection with noneuclidean geometry! There is a discussion, cluttered by references to the geography of Manhattan, of an attempt by Proclus to prove one form of the fifth postulate, but something that incorporated the perceptions of the late eighteenth or early nineteenth century would have been more useful. There is also a brief description of the Poincaré model, muddled by references to zebras, but any appreciation of an essential element of Gauss's thought, noneuclidean geometry as a genuine possibility for the space we see around us, is absent. The weakness of the Poincaré model as an expository device is that it puts us outside the noneuclidean space; the early mathematicians and philosophers were inside it.

Gauss's paper on the intrinsic curvature of surfaces, *Disquisitiones generales circa superficies curvas*, seems to have been inspired much less by his intermittent reflections on the fifth postulate than by the geodetic survey of Hannover. The paper is not only a classic of the mathematical canon but also elementary, not so elementary as noneuclidean geometry, but as the sequence Gauss, Riemann, Einstein begins with this paper, a serious essay would deal with it in a serious way and for a broad class of readers.

Having learned, as he claims, from Feynman that philosophy was "b.s.", Mlodinow feels free to amuse his readers by abusing Kant. He seems to have come away, perhaps because of the impoverished vocabulary, with a far too simple version of Feynman's dictum. It was not, I hope, encouraging us to scorn what we do not understand, and it was surely not to apply universally, especially not to the Enlightenment, in which Kant is an honored figure. As a corrective to the author's obscurantism and pretended contempt – since his views are plastic, shaped more by changing dramatic needs than by conviction, he has to concede some insight to Kant in his chapter on Einstein – I include some comments of Gauss, in which we see his views changing over the years, as he grows more certain of the existence of a noneuclidean geometry, and some mature comments of Einstein.

In a sharply critical 1816 review of an essay by J. C. Schwab on the theory of parallels of which, apparently, a large part is concerned with refuting Kant's notion that geometry is founded on intuition, Gauss writes,[1] "dass von diesen logischen Hilfsmitteln ... Gebrauch gemacht wird, hat wohl Kant nicht läugnen wollen, aber dass dieselben für sich nichts zu leisten vermögen, und nur taube Blüthen treiben, wenn nicht die befruchtende lebendige Anschaung des Gegenstandes überall waltet, kann wohl niemand verkennen, der mit dem Wesen der Geometrie vertraut ist." So, for whatever it is worth , Gauss seems here to be in complete agreement with Kant. In the 1832 letter to Wolfgang von Bolyai, he comments on the contrary,[2] "in der Unmöglichkeit (to decide *a priori* between euclidean and noneuclidean geometry) liegt der klarste Beweis, dass Kant Unrecht hatte. . . " So he has not come easily to the conclusion that, in this point, Kant was wrong. He also refers Bolyai to his brief 1831 essay in the Göttingsche Gelehrte Anzeigen on biquadratic residues and complex numbers, in which he remarks,[3] "Beide Bemerkungen (on spatial reflections and intuition) hat schon Kant gemacht, aber man begreift nicht, wie dieser scharfsinniger Philosoph in der ersteren einen Beweis für seine Meinung, daß der Raum nur Form unserer äußern Anschaung sei, zu finden glauben konnte,. . . "

Einstein's remarks appear in his *Reply to criticisms* at the end of the Schilpp volume *Albert Einstein, Philosopher-Scientist*. Excerpts will suffice: "you have not at all done justice to the really significant philosophical achievements of Kant"; "He, however, was misled by the erroneous opinion – difficult to avoid in his time – that Euclidean geometry is necessary to thinking. . . "; "I did not grow up in the Kantian tradition, but came to understand the truly valuable which is to be found in his doctrine, alongside of errors which today are quite obvious, only quite late."

Language is bound to time and place; translation, even by a skilled hand, entails choices and changes not merely its intonations but sometimes its sense. Since so much space has had to be devoted in this review to the issue of misrepresentation, I thought it best to let Gauss and Riemann speak for themselves. As a help to those unfamiliar with German, I add rough translations.

[1] "Kant hardly wanted to deny that use is made of these logical methods, but no-one familiar with the nature of geometry can fail to recognize that these alone can achieve nothing and produce nothing but barren blossoms if the living, fructifying perception of the object itself does not prevail."

[2] "the clearest proof that Kant was wrong lies in this impossibility"

[3] "Kant had already made both observations, but one does not understand how this perceptive philosopher was able to believe that he had found in the first a proof for his view that space is only a form of our external intuition. . ."

There is a grab-bag of doubtful tales about Gauss's family and childhood; Mlodinow, of course, retails a large number of them. He seems to be particularly incensed at Gauss's father, of whom he states, "Gauss was openly scornful. . . calling him 'domineering, uncouth, and unrefined'." and to be persuaded that Gauss's father was determined at all costs to make a navvy of him. Having dug a good many ditches in my own youth, I can assure the author, who seems to regard the occupation as the male equivalent of white slavery, that it was, when hand-shovels were still a common tool, a healthful outdoor activity that, practiced regularly in early life, does much to prevent later back problems. In any case, the one extant description of his father by Gauss in a letter to Minna Waldeck, later his second wife, suggests that the author has created the danger out of whole cloth:[4] "Mein Vater hat vielerlei Beschäftigungen getrieben. . . , da er nach und nach zu einer Art Wohlhabenheit gelangte. . . Mein Vater war ein vollkommen rechtschaffener, in mancher Rücksicht achtungswerter und wirklich geachteter Mann; aber in seinem Haus war er sehr herrisch, rauh und unfein . . . obwohl nie ein eigentliches Mißverhältnis entstanden ist, da ich früh von ihm ganz unabhängig wurde." As usual, Mlodinow takes only that part of the story that suits him and invents the rest. The reader who insists nonetheless on the "herrisch, rauh und unfein" and not on the virtues that Gauss ascribes to him should reflect on possible difficulties of the father's rise and on the nature of the social gap that separated him, two hundred years ago, from Gauss at the age of thirty-three and, above all, from Gauss's future wife, the daughter of a professor.

Riemann and Einstein. In two booklets published very early, the first in 1917, the second in 1922, *Über die spezielle und die allgemeine Relativitätstheorie*, and *Grundzüge der Relativitätstheorie*, the second better known in its English translation, *The meaning of relativity*, Einstein himself gave an account of the Gauss-Riemann-Einstein connection. If Mlodinow had not got off on the wrong foot with Euclid, Descartes and Gauss, he might have made the transition from Gauss to Riemann by, first of all, briefly describing the progress of coordinate geometry from Descartes to Gauss. Then, the Bolyai-Lobatchevsky noneuclidean geometry already at hand as an example, he could have continued with Gauss's theory of the intrinsic geometry of surfaces and their curvature, presenting at least some of the mathematics, especially the *theorema egregium* that the curvature is an isometric invariant and the formula that relates the difference between π and the sum of the interior angles of a triangle to the curvature. (Why he thinks the failure of the pythagorean theorem is the more significant feature of curved surfaces is not clear to me.) For the rest of the connection, he could have done worse than to

[4] "My father had various occupations. . . since he became over the course of time fairly well-off. . . my father was a completely upright, in many respects admirable and, indeed, admired man; but in his own home he was overbearing, rude, and coarse. . . although no real disagreement ever arose because I very soon became independent of him."

crib from Einstein, who explains briefly and cogently not only the physics but also the function of the mathematics. On his own, Mlodinow does not really get to the point.

In the first of the two booklets, Einstein explains only the basic physical principles and the consequences that can be deduced from them with simple arguments and simple mathematics: the special theory of relativity with its two postulates that all inertial frames have equal status and that the velocity of light is the same whether emitted by a body at rest or a body in uniform motion; the general theory of relativity, especially the equivalence principle (physical indistinguishability of a gravitational field and an accelerating reference frame) as well as the interpretation of space-time as a space with a Minkowski metric form in which all Gaussian coordinate systems are allowed. These principles lead, without any serious mathematics but also without precise numerical predictions, to the consequence that light will be bent in a gravitational field.

In the second, he presents the field equations, thus the differential equations for the metric form, which is now the field to be determined by the mass distribution or simultaneously with it. More sophisticated arguments from electromagnetism and the special theory allow the introduction of the energy-momentum tensor $T_{\mu\nu}$, which appears in the field equation in part as an expression of the distribution of mass; the Ricci tensor $R_{\mu\nu}$, a contraction of the Riemann tensor associated to the metric form $g_{\mu\nu}$, is an expression of the gravitational field. The field equations are a simple relation between the two,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -\kappa T_{\mu\nu}, \qquad R = g^{\alpha\beta}R_{\alpha\beta}$$

where κ is in essence Newton's constant.

With this equation the mathematics becomes markedly less elementary, but, with some explanation, accessible to a large number of people and inevitable if Riemann's contribution is to be appreciated. Although it would certainly be desirable to explain, as Einstein does, how Newton's customary law of gravitation follows from this equation and to describe how Einstein arrived at it, the cardinal point, where the mathematics anticipates the needs of physics, is the introduction of the Riemann tensor. It does not appear explicitly in Riemann's lecture, published as *Über die Hypothesen welche der Geometrie zu Grunde liegen* and intended for a broad audience, so that the mathematical detail is suppressed; at best it is possible to extract from the lecture the assertion that, to use our terminology, a Riemannian manifold is euclidean if and only if it is flat. The mathematics that was developed by his successors and that Einstein was able to exploit is implicit in this assertion but appears only in a paper on heat conduction published posthumously in Riemann's collected works. Because it was submitted in response to a prize theme proposed by the Academy in Paris, the paper is often referred to as the *Pariserarbeit*.

Once again, Mlodinow misses the point. Coming up to the plate against several of the great geometers of history, he strikes out each time. I could hardly believe my eyes, but it seems he is persuaded that the introduction of elliptic geometry was the principal achievement of the lecture.

It appears from the biography prepared by Dedekind and included in his collected works that Riemann, born in 1826, was very moved as a child by the stories that he heard from his father, a lieutenant during the Napoleonic wars and later pastor, of the[5] "unglückliche Schicksal Polens", partitioned at the Congress of Vienna and then oppressed by the first Tsar Nicolas. Entering university, Riemann chose at first to study theology, partly at the urging of his father, who was devoted to his vocation, but partly to secure his future so that he could contribute to the support of his family. Mlodinow, unmoved by parental sentiments or filial piety or by the plight of a hapless nation but always ready with a wisecrack, suggests that his choice was so that "he could pray for the downtrodden Poles".

It is well known that of the three possible topics proposed by Riemann for his qualifying lecture on the occasion of his Habilitation, Gauss chose, sometime in December, 1853, the third, on the foundations of geometry, the only one that Riemann did not have in a drawer fully prepared. There are two accounts of Riemann's reactions to Gauss's unexpected choice, one by Mlodinow, one by Riemann.

Mlodinow: "Riemann's next step was understandable – he spent several weeks having some kind of breakdown, staring at the walls, paralyzed by the pressure. Finally, when spring came, he pulled himself together and in seven weeks hammered out a lecture."

Riemann in a letter to his brother:[6] ". . . dass ich. . . mich wieder mit meiner Untersuchung über den Zusammenhang der physikalischen Grundgesetze beschäftigte und mich so darin vertiefte, dass ich, als mir das Thema zur Probevorlesung beim Colloquium gestellt war, nicht gleich wieder loskommen konnte. Ich ward nun bald darauf krank, theils wohl in folge zu vielen Grübelns, theils in Folge des vielen Stubensitzens bei dem schlechten Wetter; es stellte sich mein altes Uebel wieder mit grosser Hartnäckigkeit ein und ich kam dabei mit meinen Arbeiten nicht vom Fleck. Erst nach mehreren Wochen, als das Wetter besser wurde und ich wieder mehr Umgang suchte, ging es mit meiner Gesundheit besser. Für den Sommer habe ich nun eine Gartenwohnung gemiethet und habe seitdem gottlob über meine Gesundheit nicht zu klagen gehabt. Nachdem ich etwa vierzehn Tage nach Ostern

[5] "unhappy fate of Poland"

[6] "I was so deeply occupied with my investigation of the connection of the basic physical laws that I could not abandon it immediately when the topic of the qualifying lecture for the colloquium was proposed to me. Shortly thereafter I became ill, partly because of too much brooding, partly because I had kept too much to my room in the bad weather; my old ailment appeared again with great tenacity, so that my work stood stock-still. Only after several weeks, as the weather improved and I began to get about again, did my health improve. I rented a country place for the summer and since then, thank goodness, have had no complaints about my health. After I finished another paper that I could hardly avoid, I started, about fourteen days after Easter, to work zealously on the preparation of my qualifying lecture and was finished by Whitsuntide."

mit einer andern Arbeit, die ich nicht gut vermeiden konnte, fertig geworden war, ging ich nun eifrig an die Ausarbeitung meiner Probevorlesung und wurde um Pfingsten damit fertig."

Recalling that Pentecost falls seven weeks after Easter and subtracting fourteen days, we find that the preparations took only five weeks, but this discrepancy is of little importance. The others make for two accounts with quite different implications.

According to Pais, in his scientific biography *Subtle is the Lord*, Einstein, at the age of sixteen, troubled by the separation from his family, which had moved to Italy, and anxious at the prospect of military service, obtained, with the help of his family doctor, a medical certificate that released him from the Luitpold Gymnasium and allowed him to join his parents in Pavia. It was apparently not rare to leave the gymnasium before the Abitur. Thomas Mann, the novelist, left with two years to go, perhaps for similar reasons, as his recently widowed mother had moved from Lübeck to Munich. Einstein left early to avoid military service; Mann stayed only to the end of the Obersekunda, the stage required for a reduction of compulsory military service to one year. The account in Victor Klemperer's autobiography, *Curriculum vitae*, suggests that for many students it was even normal to stay only to this point and then to find a commercial position of some sort.

Einstein, who was, Pais stresses, an excellent student, had, however, no such intention. He resumed his studies elsewhere almost immediately. Nonetheless, this turn in his career gives Mlodinow a foot in the door: Einstein becomes, in a late twentieth-century word with inevitable modern connotations, a dropout. This is a pernicious notion.

As André Weil pointed out many years ago in an observation at the beginning of his essay, *The mathematics curriculum*, "The American student... suffers under some severe handicaps,... Apart from his lack of earlier training in mathematics..., he suffers chiefly from his lack of training in the fundamental skills – reading, writing, and speaking...". Unfortunately this is not less valid today than when written; indeed it may now be true of students in Europe as well. It has certainly always applied to North Americans. Although we sometimes do better than Weil foresaw (*L'avenir des mathématiques*), we are, almost without exception, handicapped all our lives because we could not begin serious thinking when our minds were fresh and free.

What is striking about the education of Gauss and Einstein is the conjunction of talent and timely opportunity. They were both encouraged very early: Gauss by his first teachers, Büttner and Bartels; Einstein by his uncle and by a family friend, Max Talmud. Both had excellent educational opportunities: Einstein because of his milieu; Gauss by chance. An Einstein from a home without books, without music, without intellectual conversation would almost certainly have been much less confident, less intellectually certain, and much more dependent; a Gauss without early freedom, without early, extensive knowledge of the eighteenth-century mathematical literature would not have discovered the implications of cyclotomy or proved the law of quadratic reciprocity so soon, if at all.

To forget this, to exaggerate his difficulties and to represent Einstein as an academically narrow, misunderstood or mistreated high-school dropout is a cruel disservice to any young reader or to any educator who swallows such falsehoods.

Mlodinow adds the occasional literary touch, not always carried off with the desired aplomb. The first lines of Blake's *Auguries of Innocence*,

> To see a world in a grain of sand
> And a heaven in a wild flower,
> Hold Infinity in the palm of your hand,
> And Eternity in an hour.

are fused, compressed to *the universe in a grain of sand*, and attributed to Keats. This is a fair measure of his scholarly care and, I suppose, of his literary culture. His own style is smooth enough; he adheres by and large to the usual conventions of contemporary American grammar, although there is a smattering of dangling participles and the occasional blooper. Confronted with an unexpected "like", he panics; the resulting "like you and I" might in earlier, happier times have been caught by his copy-editor.

The book is wretched; there is no group of readers, young or old, lay or professional, to whom I would care to recommend it. Nonetheless, there are several encomiums on the dust-jacket: from Edward Witten, the dean of string theorists, and from a number of authors of what appear to be popularizations of mathematics. They are all of the contrary opinion; they find that it is "written with grace and charm", "readable and entertaining", and so on. Perhaps the book is a hoax, written to expose the vanity of physicists, the fatuity of vulgarizers, the illiteracy of publishers, and the pedantry of at least one priggish mathematician. Would that this were so, for it is certainly thoroughly dishonest, but not to any purpose, rather simply because the author shrinks from nothing in his desperation to be "readable and entertaining".

The lesson to draw for those who have a genuine desire to learn something about mathematics and its history is that the most effective and the most entertaining strategy is to go directly to the sources, equipped with a competent, straightforward guide, say Kline's *Mathematical thought from ancient to modern times* or, for more specific topics, Buhler's *Gauss* and similar studies and, of course, whatever linguistic skills they can muster. To learn about current goals the sources are of little help, and it is up to mathematicians to acquire sufficient understanding of their own field to provide clear and honest introductions. Whether the subject is old mathematics or new, intellectual junk-food just undermines the constitution and corrupts the taste.

Mathematics and Physics

Although this heading is brief, it is far too sweeping. More established areas aside, the dynamics of renormalization by no means exhausts even those domains of mathematical physics in which fundamentally novel conceptual structures are called for. Nevertheless, if what is wanted is encouragement to a broader view of the relation between mathematics and physics than is suggested by *Euclid's Window* and similar books then it is a good place to begin. So, at the risk of seriously overstepping my limits, for the subject is large and my knowledge fragmentary and uncertain, I recall the dynamical questions that arise in statistical physics and in quantum field theories.

Thermodynamics and statistical mechanics. An historical approach, beginning with the statistical mechanics and even the thermodynamics, is the simplest and perhaps the most persuasive. Fortunately there is a very good book to draw on, Cyril Domb's *The Critical Point*, written by a specialist with wide knowledge and great experience. It would be a superb book, were it not for the high density of misprints, especially in the formulas, which are often a challenge to decipher. Even with this flaw, it can be highly recommended.

Few mathematicians are familiar with the notion of the critical point, although all are aware of the phase transition from the liquid to the gaseous state of a substance. It occurs for pairs of values $(T, P(T))$ of the temperature and pressure where the two states can coexist. It was Thomas Andrews, a calorimetrist in Belfast, who in 1869 first understood the nature of the thermodynamic phenomena at that pair of values for the temperature and pressure at which for ordinary substances, such as water, or in his case carbon dioxide, there ceases to be any difference between the liquid and the gaseous state. More precisely, for a fixed temperature below the critical temperature, as the pressure is increased there comes a point, $P = P(T)$, where a gas, rather than simply being compressed, starts to condense. This is the point at which there is a transition from gas to liquid. When the pressure is large enough, the gas is completely condensed and the substance entirely in the liquid phase, which upon further increase of the pressure continues to be slowly compressed. When, however, T reaches the critical temperature T_c, different for different substances, there ceases to be any sudden change at $P = P_c$, which can therefore only be defined as a limiting value. The liquid at pressures above P_c is not distinguishable from the gas at pressures below. For $T > T_c$, $P(T)$ is no longer even defined. So the curve of values $(T, P(T))$

ends at the critical point. What happens at the curve's other end is not pertinent here as the solid state is not considered.

There is a fascinating phenomenon, critical opalescence, associated with the critical point that the mathematical physicists among the reader's colleagues may or may not be able to describe. If not, I recommend the description in Michael Fisher's contribution to Lecture Notes in Physics, v. 186. Critical opalescence is a manifestation of a statistical mechanical feature of the critical point: the correlation length becomes infinite there. This shows itself in a less flamboyant way as singular behavior at the critical point of various thermodynamic parameters, compressibility or specific heats, although the disturbing influence of gravity renders the experiments difficult. The critical point also appears for magnets, investigated later by Pierre Curie, for which other thermodynamic parameters, susceptibility or spontaneous magnetization, are pertinent. What was understood only much later was the nature of the singularities. They are the focus of the mathematical interest.

The first theory of the critical point, now referred to as a mean-field theory, was proposed within a very few years by van der Waals in his thesis and was warmly greeted by Maxwell, for whom it seems it was an occasion to learn Dutch or Low German, as it was then called, just as earlier, Lobatchevsky seems to have been an occasion for Gauss to begin the study of Russian. The mean-field theories, for gases and for magnets, were in general highly regarded, so highly regarded indeed that almost no-one paid any attention to their experimental confirmation. The critical indices describing the singular behavior are not those predicted by van der Waals, but almost no-one noticed, at least not until the forties, when Onsager, exploiting the spinor covering of orthogonal groups, succeeded in calculating explicitly some critical indices for a planar model of magnetization, the Ising model, and discovered values different than those of mean-field theory.

Suddenly there was great interest in measuring and calculating critical indices, calculating them above all for planar models. A striking discovery was made, *universality*: the indices, although not those predicted by the mean-field theory, are equal – or appear to be for they are difficult to measure – for broad classes of materials or models. Then came the first glimpses, by L. P. Kadanoff, of a dynamical explanation. At the critical point, the material or the model becomes in a statistical sense self-similar, and the behavior of the critical indices is an expression of the dynamics of the action of dilation on the system.

The probabilistic content of statistical mechanics is determined by the Boltzmann statistical weight of each state, an exponential with a negative exponent directly proportional to its energy and inversely proportional to the temperature. The energy will usually be an extensive property that depends on the interactions defined by a finite number of parameters and by a finite number of local properties such as the magnetization. The basic idea of the dynamical transformation is that, because it is only the statistics that matter, small-scale fluctuations can be averaged and substantial chunks of the system, blocks, can be reinterpreted, after a change of scale, as small uniform pieces with well-defined local properties.

It appears to have been K. G. Wilson who turned this idea into an effective computational tool, the renormalization group, and it is probably his papers that it is most important that analysts read, for the success of the renormalization-group method is a result of a basic property of the associated (infinite-dimensional) dynamical system: at certain fixed points, the pertinent ones, there are only a finite number, usually one, two or three, of unstable directions (or, more precisely, a finite number, perhaps larger, of nonstable directions). All other directions are contracting and indeed most of them strongly so. A model, depending on parameters, temperature, pressure, or magnetic field, appears as a point in the space of the dynamical transformation. A critical point appears as the parameters are varied and the corresponding point traverses a stable manifold. Because the manifold is stable and because it is the transformation that determines the properties of the system, all questions can be referred to the fixed point in it. This is the explanation of universality.

To establish such a theory for even the simplest of planar models, percolation for example, is a daunting mathematical challenge – in my view central. For other planar models, it is not even clearly understood what the dynamical system might be; indeed on reflection it is clear that the very definition of the dynamical transformation relies on the property to be proved. So if there is a theory to be created, its construction will entail a delicate architecture of difficult theorems and subtle definitions. My guess is that there may be a lot to be learned from Wilson, who after all must have been able at least to isolate the expanding directions sufficiently to permit effective calculations, but this guess is not yet based on much knowledge of his papers.

Quantum field theories. In quantum field theory exactly the same dynamical structure of a finite-dimensional unstable manifold and a stable manifold of finite codimension plays a central role in the construction, by renormalization, of theories like quantum electrodynamics. Indeed it is sometimes possible to pass, by an analytic continuation in an appropriate parameter, from statistical mechanics to quantum field theories, but a direct approach to them is often more intuitively appealing.

The field theory is a much more complex object in which the algebra takes precedence over the analysis, most analytic problems being, apparently, so difficult that they are best left unacknowledged.

In statistical mechanics there is an underlying probability space, say (X, μ). A related space – there is a conditioning by time – appears in field theories. It is enormous. In addition to functions on X, whose expectations are the pertinent objects in statistical mechanics and which act on $L^2(\mu)$, there are in a field theory many other operators as well, to monitor the symmetries or to implement creation and annihilation of particles. So temporal evolution in a field theory is more easily grasped directly than as an analytic continuation of some stochastic process. Seen most simply, it arises from a constant creation and annihilation of particles, anti-particles and fields, most created only to last for a very short time and then to be destroyed again.

The difficulties of the theory lie in these processes and its appeal to mathematicians in the constructs necessary to surmount the difficulties. The theory is usually prescribed by a Lagrangian, $\mathcal{L}$, which can perhaps be thought of as a prescription for the probabilities with which the elementary processes of creation or annihilation occur, when, for example, an electron and a photon collide to produce an electron of different energy and momentum or an electron and a positron collide, annihilate each other, and produce a photon. The distinction between an elementary process and a compound process is to a large extent arbitrary. We do not observe the processes occurring, only their outcome, so that what for one theory is purely compound may for another, equivalent theory be partly compound, partly elementary; for example, two electrons exchanging a photon, and then emerging from this intimate encounter with changed momenta could be elementary or it could be the composition of two electron-photon interactions, or an electron could briefly separate into an electron and a photon, that then fused into a single electron. Infinities arise because the elementary processes that enter into a compound process, being unobserved, can occur at all energies and momenta; the usual relation between energy, mass and momentum is violated.

One way to attempt to surmount this difficulty is to allow only elementary processes with momenta and energy that are no larger than some prescribed constant Λ. This provides numbers $\mathcal{G}_*(\mathcal{L})$ that are finite, although perhaps inordinately large, for every conceivable process, thus, otherwise interpreted, for every conceivable scattering of an arbitrary number of incoming particles, with whatever momenta and energy they are allowed, as a collection of outgoing particles, although the distinction between ingoing and outgoing is to some extent arbitrary. In the numbers $\mathcal{G}_*(\mathcal{L})$, referred to as the amplitudes, the distinction between an elementary and a compound process is lost, except of course for $\mathcal{G}_{\prime}(\mathcal{L})$ which are just the amplitudes prescribed by the Lagrangian itself. To obtain a true theory, it is necessary to take Λ to infinity, but if the numbers $\mathcal{G}_*(\mathcal{L})$ are to have finite limits, it may be necessary to adjust simultaneously the parameters in the original Lagrangian, thus in the initial prescription, so that they go to infinity. Since the initial prescription was made on the basis of an arbitrary distinction between elementary and compound process, this is not so paradoxical as it at first appears.

The source of the success of this method is ultimately the same dynamic property as in statistical mechanics. For large Λ the transformation $\mathcal{L} \to \mathcal{L}'$ defined by requiring for all processes the equality of the amplitudes $\mathcal{G}_*(\mathcal{L}) = \mathcal{G}_{\prime}(\mathcal{L}')$ operates in essence only on an unstable subspace of very small dimension, other directions being strongly contracted. Moreover, a finite set of amplitudes $\mathcal{G}^{\rangle}(\mathcal{L})$, $1 \leq i \leq n$, can be chosen as coordinates on this space. Thus for large Λ, we can choose $\mathcal{L}_*$ in a fixed space of dimension n so that $\mathcal{G}_*^{\rangle}(\mathcal{L}_*)$ has any desired values for $i = 1, \ldots, n$. In the end, as $\Lambda \to \infty$, the Lagrangian $\mathcal{L}_*$ sails off backwards in the stable directions, but the amplitudes $\lim \mathcal{G}_*^{\rangle}(\mathcal{L}_*)$ and, more generally, because of the stability, all $\lim \mathcal{G}_*(\mathcal{L}_*)$ remain finite and define the theory. This is a crude geometric or dynamical description of what is in reality a very elaborate process: renormalization as it appears, for instance, in quantum electrodynamics. Nevertheless the dynamics is paramount.

The dynamics looks even more doubtful than in statistical mechanics because there is now a whole family of transformations, one for each Λ. That can be remedied. If μ is smaller than Λ, then it is relatively easy to find $\mathcal{L}_\mu$ such that $\mathcal{G}_\mu(\mathcal{L}_\mu) = \mathcal{G}_*(\mathcal{L}_*)$. The pertinent map is

$$R_\nu : \quad (\mathcal{L}, *) \to (\mathcal{L}_\mu, \mu), \qquad \nu = \frac{\mu}{*}.$$

These maps form a semigroup in ν, which is always less than 1. The presence of the second parameter is disagreeable but seems to be tolerable. There is now much more to be moved by the maps: the Hilbert space and all the attendant operators. It is not yet clear to me how this is done, but it is a process with which physicists appear to be at their ease. In addition, in the standard model of particle physics as well as in many of the geometric applications that appeal to mathematicians (see the lectures of Witten in the collection of surveys *Quantum fields and strings: a course for mathematicians*, two volumes that are indispensable when first trying to understand quantum field theory as a mathematical subject) it is gauge theories that occur and the renormalization and the dynamics have to respect the gauge invariance.

It appears – I have not understood the matter – that just as the heat equation can be used to establish various index theorems or fixed-point theorems by comparing traces near $t = 0$ and near $t = \infty$ where they have quite different analytic expressions, so does, by a comparison of calculations at low and high energies, the dynamics of quantum field theory, which moves from one to the other, allow the comparison of quite different topological invariants: the Donaldson invariants and the Seiberg-Witten invariants. My impression, but it is only an impression, is that a number of the applications to topology or to algebraic geometry involve similar devices. If so, that is perhaps one reason, but not the sole reason, for attempting to establish analytic foundations for the procedure.

String theory. In string theory, there are even more ingredients to the dynamics. Grossly oversimplifying, one can say that the particles are replaced by the states of a field theory on an interval, thus by the modes of vibration of a string in a space M of dimension D, at first arbitrary. This is only the beginning! There is even at this stage a good deal of implicit structure that reveals the special role of $D = 10$ and $D = 26$: conformal field theory and supersymmetry above all. Moreover, the Feynman diagram is thickened; rather than a graph with vertices and edges, it becomes a surface with marked points. Finally the Lagrangian, which could be thought of as simply a finite collection of numbers, one attached to each of the different types of vertex, is now described by a minkowskian metric of signature $(1, D - 1)$ on the space M, so that there appears to be an infinite number of free parameters. It has, however, been pointed out to me that the apparently free parameters are rather dynamical variables. As in the general theory of relativity, this background metric tensor is to be treated as a collection of fields, thus as a collection of dynamical variables, and, as a consequence, it is subject to a quantization. So there are no arbitrary parameters in the theory!

When discussing statistical mechanics, we emphasized the critical point, but the dynamical transformation bears on other matters as well. It will reflect, in particular, the abrupt change from a gas to a liquid at temperatures below the critical temperature or the possibility of spontaneous magnetization: very small changes in the imposed magnetic field entail very large differences in the induced magnetization, not merely in size but also in direction. The dynamics is moving points apart as $\nu \to 0$. Analogues in field theory are multiple vacua or, in string theory, the great variety of low-energy (small ν) limits. So the arbitrary parameters appear to resurface!

Certainly, in string theory the analytic problems that it is fair to regard as central mathematical, although perhaps not physical, issues in statistical mechanics recede – for the moment at least – into the background. They are not entirely unrelated to the problem of choosing among the vacua and thus of constructing a single distinguished physical theory rather than a family of theories, a problem that also seems to be in abeyance at the moment. Such matters are far over my head. The issues of most current appeal in mathematics, and to a lesser extent in physics, are algebraic or geometric, perhaps above all geometric: the transitions from one family of low-energy theories to another; or the possibility – another low-energy phenomenon – that different spaces M and different background metrics on them lead to the same theory (mirror symmetry and other dualities).