Proposal for a Book:

PATTERN AND PARADIGM

THE CONCEPTUAL BASIS OF CONTEMPORARY PARTICLE PHYSICS

Bruce Andrew Schumm

Associate Professor of Physics

University of California at Santa Cruz

PATTERN AND PARADIGM

Bruce Andrew Schumm

University of California, Santa Cruz

Book Overview

This is a book about particle physics. Certainly, anyone considering the representation or publication of this book has a pretty good idea of what particle physics is all about: humankind's best attempt at a scientific explanation of the most basic principles which underly the structure and operation of the universe in which we live.

The selling point of this book is the following: I don't think you'll find a book quite like this one anywhere in the realm of science trade literature, much less within the field of particle physics itself. This book dives into the subject matter of particle physics in a thoroughly unabashed way. Particle physics is an extremely rich and beautiful subject, but its full depth and import can only be appreciated in the context of a deep and uncompromising exposition. This book tackles the subject matter of conventional particle physics at just this level, which is something that no other previous book aimed at the general public has attempted to do.

But the book is far from a dry, grinding presentation of the gory detail. Quite the contrary, it's a lighthearted, often humorous, and uniformly spirited discourse which will compel a lay audience through the labyrinth of fascinating notions and conjectures that form today's Standard Model of particle physics. The combination of the book's conceptual rigor and its engaging style make this book a one-of-a-kind exposition on one of the most fascinating and profound threads of modern science. I think it has the potential to become tremendously popular.

Anyone interested in science will get real pleasure from this book. They'll get a kick out of its humorous elements, they'll be moved by its more profound passages, and they'll be uplifted by the deep understanding that they'll develop regarding the fundamental nature of the universe, and its abstract and mathematical underpinnings. At times, the reader will find the text as much a rhapsody about the nature of human thought processes as it is an exposition of one of the greatest achievements of those processes.

As the story of *Pattern and Paradigm* unfolds, the reader will be treated to a discussion of many of the compelling developments from the last century of physics. She will enjoy a reprise on the basic tenets, and consequences, of quantum mechanics. She will encounter the excitement of the identification and discovery of the fundamental building blocks of nature, the final chapter of which obliged, much to my delight, to unfold even as this discussion was being written. She will be treated to a surprisingly tactile description (including some easily performed demonstrations) of the necessary background from higher mathematics, and then to its synthesis with the tangible in terms of the enigmatic notions of 'internal' physical spaces. She will see how the behavior, and even the *necessity*, of causative agents – interactions – follow from the understanding of these developments. And then she will be presented with the Higgs field, and exposed to the unorthodox notion that the most characteristic quality of matter – mass – is now thought to be nothing more than an illusion, concocted by the swirling of eddy currents in the all-pervasive Higgs field. It is, all in all, a strange and wonderful set of notions that are woven together this narrative about today's Standard Model – a paradigm providing nothing less than a description of the fundamental basis for the full and diverse set of phenomena that color the world in which we live with an almost infinite array of hues.

Happily, all this can be done – and is done – with very little requirement of mathematical background on the part of the reader. There are a few very basic mathematical notions that that the reader will need to know, such as orders of magnitude, or the appreciation that if an object's wavelength λ is the quotient of the fundamental physical constant h and the object's energy p ($\lambda = h/p$), then the bigger p is, the smaller λ will be. The book will contain appendices that will enable readers who don't use much math in their everyday lives to refresh themselves in these two areas. The core of the book's mathematical content – the higher and more abstract mathematics introduced towards the middle of the book – lends itself particularly well to a purely descriptive presentation, and requires no traditional mathematical background whatsoever.

It really is just the time for a book like this to come out. Over the decade of the 1990's, a program of exacting studies done in Europe and the United States have validated the Standard Model of particle physics to a degree unimagined as the decade began (I know – I played a major role in these tests, and I can tell you that no one was as amazed as I as the studies played themselves out). On the other hand, though, for all our prowess and success, we have yet to mount an

experiment powerful enough to find the one remaining missing component of the Standard Model: the Higgs Boson. This book describes the Standard Model in detail, motivates and describes the Higgs boson, and discusses how the tests of the 1990's provided a profound verification of the Standard Model while simultaneously suggesting that the discovery of the Higgs lies just ahead of us. It describes why it is that we think the discovery of the Higgs is a much of a new beginning to our attempts to glean the ultimate order of nature as it is the final chapter in the story of our current Standard Model. And all of this is done in a way that exposes the deep connection between the ethereal world of the abstract mathematician and the concrete realm of the experimental physicist – with essentially no demand on the preparation of the reader. Particle physics *can* be understood by anyone with a deep interest, and this is the book – the only book – which can provide that understanding.

So, my pitch for this book, in a nutshell: if you think there are a lot of people out there who would really like to know, at last, what particle physics is truly all about, and have a deep enough interest to carefully read a few hundred pages on the subject, then publish this book. If you want that presentation to be provided by a physicist who has developed some unique insights over the last ten or so years of intensive involvement with the subject, then again, publish this book. And if you want that presentation to be given by an author and physicist who can as easily poke fun at himself and the world around him as he can elicit the profound revelations of our most fundamental science, then once again, publish this book. I believe that the spirit of this book, and the intrinsically compelling nature of its subject matter, will carry it forth splendidly.

Brief Market Analysis

At this time, there is simply no book in the literature which satisfies the role that *Pattern and Paradigm* will fulfill. The two books which come closest – Leon Lederman's *The God Particle* [Houghton Mifflin, 1993] and Brian Greene's *The Elegant Universe* [Norton, 1999] – have substantially different goals and subject matter than those of this book. Lederman's deservedly popular book paints a broad, often anecdotal portrait of particle physics, but does not attempt to dig in and provide the concrete discussion necessary to give the reader the sense that she has grasped the theory of the fundamental at its core. Greene's book is more ambitious in this

regard, but covers substantially different ground, including much science which at this time, while indeed elegant and promising, is still fairly speculative. This book covers in detail the established, and quantitatively confirmed, fundamental theory of the workings of nature. It is quite complementary to Greene's *Elegant Universe*, and in fact makes frequent reference to it.

The potential audience for this book is large. The success of both Lederman's book (several hundred thousand copies sold) and Greene's book (New York Times bestseller) stand as proof of the high level of interest in the subject. The degree of rigor of the book is perhaps somewhat greater than that of Lederman's *God Particle*, but is about equivalent to that of Greene's *Elegant Universe*, and is somewhat less than that of two other successful popular treatises on physics: Stephen Hawking's *A Brief History of Time* and Steven Weinberg's *The First Three Minutes*. In a different field, the degree of attention required of the reader of *Pattern and Paradigm* is about the same as that of *The Blind Watchmaker* [Norton, 1987], Richard Dawkins' best-selling defense of Darwinian evolution. Perhaps the greatest motivation for the generation of *Pattern and Paradigm* has been the innumerable times, both with physics students as well as in social interactions with the lay populace, that I have found myself wishing that I had such a book to recommend.

Anyone considering the representation or publication of this book might particularly want to take a look at Greene's recent bestseller. This book does for his branch of theoretical particle physics (quantum gravity) more or less precisely what mine purports to do for mainstream particle physics: it presents the ideas in an accessible and forthright manner, with little other theme or thesis than that of the subject matter of the field itself. For particle physics, as for its related (but more speculative) field of quantum gravity, the science itself is story enough for the reader, for they are both fields which drive, in complementary and rewarding ways, right to the heart of one of the most basic and enduring threads of human intellectual inquiry.

Furthermore, a number of recent developments make the timing of this book ideal. More so than ever before, particle physics seems to be 'in the air'; in recent years, awareness of the field has risen steadily, to the point where it stands as a hallmark of human advancement. In particular, since I began working on this book in June, 2000, a substantial amount of popular interest has been generated, in part by the my own research, in the potentially imminent discovery of the Higgs boson.

Numerous articles have appeared in the popular press about the possible unfolding of this profound new chapter in experimental physics. The Higgs boson is a central, as of yet undiscovered and unexplored component of the Standard Model of particle physics, playing no less a role than that of the source of mass in the universe. As such, the Higgs boson is treated in depth in the penultimate chapter of this book.

In addition, in the fairly likely case that the existence of the Higgs is definitively confirmed before the end of the decade, this book would likely enjoy a second surge in popularity. For anyone who really wanted to know what the Higgs boson is, why it's important, and why its study has the potential to reshape once again our deepest notions of the structure of the universe, *Pattern and Paradigm* is the book they would turn to. In this respect, as well as a number of others, *Pattern and Paradigm* would occupy a unique place in the literature.

Qualifications

A major portion of my research over the last decade or so has been devoted to the intense scrutiny of the Standard Model that has been made possible by the world's leading particle accelerators. I have been one of the leaders of a relatively small core of physicists who have performed what is to date the most precise test of the self-consistency of the Standard Model, and cast considerable light on the issue of its one remaining undiscovered ingredient: the Higgs boson. In this pursuit, I have developed a deep familiarity with the paradigm of the Standard Model, as well as some unique perspectives towards it. I have taught a very successful (albeit graduate level) course on the physics of the Standard Model, and fascinated hundreds of people, from acquaintances and family members to graduate students in physics, with tours of the Stanford Linear Accelerator Center. I have written a 3000 word article on precise tests of the Standard Model which has been published in the SLAC *Beam Line* – a periodical, published for a popular audience by the Stanford Linear Accelerator Center, with a readership of several thousand. I am including a copy of this article in the proposal, along with the book's introduction and two more lengthy sample chapters.

In general, I am a young, active, and internationally recognized experimental particle physicist. My contributions to the field have been voluminous and diverse, including numerous publications in several different subfields of particle physics, the design and implementation of advanced experimental technology, and an involved

and ongoing participation in the dialogue about the future directions of particle physics experimentation. Over the past five years, I have been on the faculty of the University of California (Santa Cruz campus), where, in addition to my research work, I have been a teacher of physics who is held in high regard at all levels, from basic introductory physics through rigorous graduate level courses. Based on both my research and pedagogical achievements, and at the recommendation of my colleagues, I was accelerated to the position of Associate Professor with Tenure after only four years as an Assistant Professor.

In summary, there is a huge gap in the informative literature of particle physics – one of the most central and profound achievements of the human intellect. Due to a combination of professional experience, writing ability, and pedagogical skill, I believe that I am one of the few physicists who is in a position to fill this gap.

Specifications

I estimate that the text, when completed, will stand very roughly at 150,000 words. Some degree of artistic support will be required, although all but a few of the approximately 50 figures are graphical in nature.

Bruce A. Schumm

Address: 329 Natural Sciences 2
University of California, Santa Cruz
Santa Cruz, CA 95064

Phone: (831) 459-3034
EMAIL: schumm@scipp.ucsc.edu

212 Almena St.
Santa Cruz, CA 95062

Phone: (831) 423-9062

Birth date: October 23, 1958

Education:

B.A.	1981	Haverford College, Physics and Mathematics; *Magna Cum Laude*
Ph.D.	1988	University of Chicago, Physics

Committees: Executive Board, SLAC User's Organization (SLUO) 1993-1996
Organizing Committee, RADCOR 2000 Conference

Honors: Elected to Phi Beta Kappa, 1980
LBL Outstanding Performance Award, 1994

Experience:

7/99-Present	Associate Professor of Physics	University of California, Santa Cruz
7/95-6/99	Assistant Professor of Physics	University of California, Santa Cruz
3/94-6/95	Staff Scientist	Trilling/Goldhaber and Chamberlain Groups Lawrence Berkeley Laboratory 1 Cyclotron Road Berkeley, CA 94720
7/88-3/94	Research Associate	
1983-1988	Research Assistant	Experimental High Energy Physics Group Department of Physics and the Enrico Fermi Institute University of Chicago Chicago, IL 60637

Bruce A. Schumm — Selected Publications

1. **Precision in Collision: Dissecting the Standard Model**
   SLAC Beam Line, Vol. 31, No. 2, Spring, 2001.
2. **Direct Measurement of A_b at the Z^0 Pole Using a Lepton Tag**
   (with the SLD Collaboration), SLAC–PUB–8951, July, 2001 (to be submitted to *Physical Review Letters*).
3. **Search for the Decay $B^0 \to \gamma\gamma$**
   (with the BaBar Collaboration), SLAC–PUB–8931, July, 2001 (submitted to *Physical Review Letters*).
4. **Linear Collider Physics Resource Book**
   (With the American Linear Collider Working Group), SLAC–R–570, May, 2001.
5. **A High Precision Measurement of the Left-Right Z Boson Cross Section Asymmetry**
   (With the SLD Collaboration), *Phys. Rev. Lett.* **84**, 5945 (2000).
6. **B Physics with a Giga-Z Sample**
   Proceedings of the 5^{th} Int. Workshop on Linear Colliders, Fermilab, Batavia, Illinois, October 24-28, 2000.
7. **Search for Charmelss Hadronic B Decays with the SLD Dectector**
   (With the SLD Collaboration), *Phys. Rev.* **D62**, 071101 (2000).
8. **Tracking and Vertexing at a Linear Collider**
   Proceedings of the 4^{th} Intl. Workshop on Linear Colliders, Sitges, Spain, April 28 – May 5, 1999.
9. **Precise Electroweak Measurements at the Z^0 Pole**
   Proceedings of the 19^{th} International Conference on Physics in Collision, Ann Arbor, MI, June 24–26, 1999.
10. **A Bipolar Front End Integrated Circuit for the BaBar Helium Drift Chamber**
    With D. Dorfan, E. Spencer (UC Santa Cruz) and D. Coupal, D. Nelson (SLAC), *Nucl. Instr. & Meth.* **A409**, 310 (1999).

List of Chapter Titles

Brief Outline

Chapter 1. Introduction

This relatively brief introductory chapter provides motivation for interest in the book, some general background discussion about particle physics, and a very cursory outline of the chapters that lie ahead.

Chapter 2. The True Movers and Shakers (The Forces of Nature)

This chapter introduces the four basic forces of nature – electromagnetism, gravity, and the strong and weak nuclear forces – differentiating them in terms of three characteristics: overall strength, range of influence, and the associated charge. The role that each force plays in the order of the universe is presented and discussed. The notion of the 'unification' of forces is introduced, in preparation for the eventual presentation of the Standard Model of particle physics, which is a unified theory of the electromagnetic and weak nuclear forces, and whose elucidation forms the thrust of this book.

Chapter 3. The Baby *and* the Bathwater (The Modern Physics Revolution)

This chapter provides the relevant background notions from early Modern Physics – Einstein's relativity and the quantum mechanics of the Copenhagen school. The chapter begins with a discussion of relativity, focussing on the equivalence of mass and energy. It then proceeds to introduce the notion of quantization and its original motivation in terms of attempts to understand the physics of glowing objects. A lengthy discussion of the distinction between particles and waves follows, with a detailed exposition on wave properties, and particularly *phase*, which plays a central role in the theoretical structure of the Standard Model. The chapter closes with thorough discussions of the origin and interpretation of Heisenberg's uncertainty principle, the notion of the wavefunction, and Schroedinger's wave mechanics.

Chapter 4. The Marriage of Relativity and Quantum Mechanics (Relativistic Quantum Field Theory)

In the decades immediately following the birth of modern physics, the difficult task of bringing together Einstein's relativity with quantum mechanics gave birth to a new framework – relativistic quantum field theory – into which all modern fundamental theories of causation are cast. This chapter presents the basic notions of relativistic quantum field theory, and describes how they once again fundamentally reshaped our view of the natural world. The chapter begins with the classical (19^{th} century) interpretation of causation in terms of force fields. It then goes on to introduce the modern notions of quantized fields, and the re-interpretation of force in terms of particle exchange. It discusses how quantum field theory generalizes the idea of 'force' to the idea of 'interaction', and introduces the essential notion of the 'minimal interaction vertex'. Feynman diagrams, constructed from the minimal interaction vertices and providing an intuitive representation of the set of possible modes of causation, are introduced. After the introduction of antimatter and its role in quantum field theory, the notion of virtual particles and the living vacuum is introduced. The chapter ends with a discussion of the triumph of the experimental confirmation of quantum field theory.

Chapter 5. Patterns in Nature (The Fundamental Building Blocks)

This chapter introduces the set of fundamental particles which constitute the ingredients from which the known universe is constituted. The connection between observed patterns and underlying structure, which played a central role in the theoretical development leading to the Standard Model, is first presented here in terms of the 'eightfold way' and the discovery of quarks. This is followed by the introduction of the leptons (electron-like particles) and the history of their discovery. Next comes a discussion of the 'November Revolution' and the establishment of the generational pattern, followed by the discussion of the role of the third generation and a digression on the origin of the dominance of matter over antimatter. The various force-carrying particles are then introduced, which will later (Chapter 8) be seen to arise from natural 'internal' symmetry patterns. The chapter ends with a compilation of these fundamental constituents, and a discussion of the 'particle zoo' and its codification.

Chapter 6. Mathematical Patterns (Lie Groups)

In this chapter, the book takes a turn into the abstract, introducing the notion of a mathematical group. A specific example – the group Mod(4) of arithmetic on a clock with only four hours – is presented as a simple example. Continuous 'Lie' groups, which will form the basis of the description of underlying natural symmetry patterns, are introduced. The re-ordering ('non-commuting') properties of Lie groups are introduced via a simple demonstration that can be done with a book or a small box as a prop. Finally, the specific Lie groups of relevance to the Standard Model are introduced.

Chapter 7. The World Within (Internal Symmetries)

This chapter presents the marriage between the previous chapter's abstract world of mathematical group theory and the concrete behavior of the natural world. The relation between symmetry groups and conserved (unalterable) physical quantities, formulated by the turn-of-the-century mathematician Emmy Noether (a friend of Einstein's), is presented. The connection between rotational symmetry and angular motion (angular momentum) is presented as a lead-in towards the notion of spin. The concept of an 'internal symmetry', and its ambiguous physical implication, is made in the context of the semi-real internal symmetry space of spin. Next follows the introduction of the fully abstract 'isospin' internal symmetry space, and its demonstrable connection to physical phenomena. The chapter closes with a reprise on the eightfold way, this time from the vantage point of internal symmetry. The notion of 'global' invariance is introduced in this closing discussion, to pave the way for the next chapter's introduction of the critical notion of 'local' invariance.

Chapter 8. Physics by Pure Thought (Gauge Theory)

This is the first of the two chapters which draw on the background of the previous chapters, pulling them all together into a description of the basis of the current fundamental theory of causation. The notion of quantum-mechanical phase is reintroduced, followed by a discussion of global phase invariance. At this point, the more powerful notion of local phase invariance and its associated 'gauge principle' is introduced. It is discussed how this simple principle leads to a complete description of causation via the introduction of the minimal interaction vertices from

which all interaction is constituted. It is discussed how this gauge principle is a natural goal of our millenia-long effort to impress the almost infinitely disparate world of natural phenomena into a single underlying ordering principle of tremendous physical simplicity and mathematical necessity. The chapter then goes on to introduce anew the two components of the Standard Model – the electromagnetic and weak nuclear interactions – in the context of the gauge principle and their associated internal symmetry spaces. This is followed by a discussion of the critical connection between the gauge principle and the formulation of self-consistent physical theories. The chapter closes with a presentation of the current theory (quantum chromodynamics) of the nuclear binding force.

Chapter 9. The Current Paradigm (Parity Violation, Hidden Symmetry, the Standard Model, and the Higgs Boson)

This is the second of the two chapters which pull together the essential threads from previous chapters. After the introduction of parity (mirror-reflection symmetry) violation, the beginning of the discussion of the unification of the weak nuclear and electromagnetic interactions is entered into. The book then goes on to ascribe the apparent 'weakness' of the weak nuclear force to the great mass of its corresponding exchange particles. After discussing the incompatibility between gauge theory and exchange-particle mass, the Higgs field is introduced along with the critical notion of spontaneously broken symmetry. With this done, and the very notion of mass completely reworked, the unification of the gauge theories of the electromagnetic and weak nuclear forces into the single electroweak theory of the Standard Model can proceed. With the Standard Model presented at last, the chapter continues with a discussion of the current status of the precise and exacting studies that have been done to test its predictions, which are being concluded in the United States and Europe as this is written. The chapter closes with a discussion of the light that these tests shed on the as-of-yet undiscovered Higgs boson.

Chapter 10. Onward and Inward (Parting Thoughts)

In this final chapter, a number of somewhat disparate threads are tied off. The chapter begins with a discussion of the deep impact that the field of particle physics has had on technology and on the economy as a whole. The chapter continues with a discussion of why the next decade holds great promise for further advances in particle physics, and why we may expect our view of the natural world to yet again be fundamentally reshaped. Finally, the book closes with some philosophical musings regarding the place of contemporary particle physics in the larger scheme of the evolution of the human intellect.

Talk of mysterie! Think of our life in nature – daily to be shown matter, to come into contact with it – rocks, trees, wind on our cheeks...

Henry David Thoreau

Copyright 2001 by Bruce A. Schumm

CHAPTER 1
INTRODUCTION

If you've gone to the trouble of opening the cover of this book, and are taking in these paragraphs with the thought in mind that you're going to read the book through, then you've probably got a pretty strong interest in science. You probably have a desire welling within you to know, at the deepest level permissible by the ever-incomplete state of human knowledge, what it is that drives forward the world in which you live.

If you *do* have that interest – that deep-seated curiosity to inform yourself of the surprisingly odd way in which particle physicists view the universe, and the stamina to read, think a bit, at times even re-read and think a bit more – then you might well find that the book you hold in your hands provides just the description of the fundamental workings of the nature that you've been looking for.

This is not, first and foremost, a story about the history of particle physics, or of the lives of its great protagonists. It's not a book of anecdotes about the culture and society of particle physicists. It *is* a book which presents, at a level well beyond the superficial, the conceptual ideas that underlie those physicists' view of the world – a view of the world, increasingly supported by the exacting scrutiny of experimental science, which is fundamentally simple and concise, yet at the same time mathematically elegant, deeply profound, and thoroughly fascinating.

Einstein himself held that any physical theory worthy of respect must be explicable to any clear-thinking person. This book represents my attempt to elucidate the current best-guess theory of particle physics – the paradigm of the Standard Model – for the interested public.

To be deeply interested, however, does not mean to be steeped in the formal content – mathematically or scientifically – of physics. In particular, I am expecting very little of the reader in terms of mathematical background. If need be, you can refresh yourself of the little mathematics required (orders of magnitude, and some very basic notions from algebra) in the appendices of this book. If you're not confident of your mathematical background, take a look at those appendices right now – I think you'll be reassured.

On the other hand, though, it's impossible to elicit the central notions of the contemporary theory of causation without an involved discussion of the mathematics which underlies it. In fact, the beautiful connection between the worlds of the mathematician and the physical scientist is one of the most profound and fascinating threads which we'll find ourselves following. The very fact that the abstractions of higher mathematics bear some relation to the physical world – in fact, seem to lie at the very heart of its order and operation – seems yet more astounding to me every time I find myself expounding upon it to some unsuspecting soul.

The mathematics that we'll discuss, though, is not the calculation-mired pursuit that confronts one in introductory college-level courses, but rather that of the abstract mathematician, whose tools are more those of logic and generalization than they are the vexing application of endless pages of arithmetic operation. It's a different sort of mathematics than most of us are used to – an evolved discipline which seems to bear little in common with the practical, everyday manipulations from which it sprang. More so than numbers and equations, this mathematics is a science of structure and, in particular, the patterns exhibited by that structure. And it is through these patterns that, in the latter half of the 20^{th} century, the deep connection between abstract mathematics and the basic organizing principles of the physical universe first came to be recognized.

Before the mathematics, though, we'll need to ground ourselves in the backdrop of physics from which the Standard Model sprang forth in the fertile period of the 1960's. Illustrating the danger of the use of the adjective 'modern' in labeling any human development, this grounding begins with the quaint old 'modern' physics – Einstein's special relativity and conventional quantum mechanics – of the early part of the 20^{th} century. A discussion of the contemporary framework of causation – the 'quantum field theory' of endearing Nobel Laureate Richard Feynman and others – then describes how the 19^{th} century notion of action-at-a-distance has been supplanted by the idea that physical forces are conveyed by the exchange of subatomic particles. The following chapter introduces the current array of fundamental (indivisible) subatomic particles, noting the repeating 'generational' patterns into which those particles fall, and setting the stage for the ensuing descriptive journey into the world of abstract mathematics.

The Standard Model itself comes to us in the second-to-last chapter of the book (the final chapter being somewhat of a verbal core-dump of musings and

philosophical conjecture) with the introduction of the chic notions of the Higgs field and spontaneously hidden symmetry. It is my most cherished hope that, in the context of the lengthy preceding discussion of the edifice on which this theory rests, you will at last appreciate this achievement for what I believe it to be: the coalescence of the millenia-long quest for an understanding of the fundamental nature of the universe into a paradigm whose success represents one of the most remarkable and astounding triumphs in the history of human intellectual endeavor. And, with almost equal longing, I pray that your interest may be piqued by the suggestion, presented in the final chapter, that the years ahead of us hold great promise for even further deepening of our understanding of the natural world.

And, lastly, it is also my hope that you enjoy the journey. While the experimental and theoretical threads leading into the development of the Standard Model have had a telling impact on the way that we live our day-to-day lives, it is really the satisfaction of innate human curiosity, and the attendant appreciation and awe of the workings of nature, which stand as the historical and continuing motivation for particle physics research. While numerous practical spin-offs have arisen directly from this research, to date its funding decisions have been based, by and large, on the sole criterion of whether or not the proposed program stands to clarify and deepen our understanding of the order of the natural world. And while the program has been profoundly successful, our success at returning its fruits back to its true benefactors – the citizens of the many nations of the world which support it through their governments' science programs – has been more limited. I am hoping that this book acts in some small measure to redress this deficiency.

Viewed from this perspective, the list of patrons is lengthy indeed, for particle physics research is a truly international and cooperative pursuit. The three traditional centers of the development of particle physics – Western Europe, Eastern Asia, and North America – have been roughly equal partners over the last 50 or so years. Increasingly, more and more nations and regions are being drawn into the effort, and all the continents (including Antarctica!) are now represented. It is a pursuit, like many others in science, where national allegiances play a relatively minor role, and collaborative professional and personal relationships spring forth quite independent of cultural background.

The support for particle physics research provided by this worldwide community

does need to be acknowledged, for it is not small. The annual global outlay for particle physics research is several billion US dollars per year; for example, the United States' contribution to this total is about $800 million per year. But perhaps, in the end, you'll agree that we do get our money's worth.

One final thing: as we move, very soon now, into our discussion of the conceptual basis of particle physics, one thing in particular bears keeping in mind. What follows is not fantasy, nor (for the most part) even speculation. Instead it is, to the best of our knowledge (and backed up, as we'll see, by some pretty impressive experimental verification), scientific fact. What it tells us about the world in which we live is, to whatever extent anything can be said to be such, an accurate and faithful reporesentation of physical reality. If it seems strange, or maybe even slightly perverse, so be it. The theory works, and as exacting tests performed in the decade of the 1990's demonstrated, it works remarkably well. It is as much of an absolute as any paradigm of the workings of nature could ever be.

But enough fanfare – it's time to get on with it.

To the extent that our world is an interesting place in which to live (in fact, is a place in which life itself is possible) it is because the world is rife with *causation* – the ability of things to influence other things. And the way that physical objects influence each other – be they contiguous neurons in the taxed brain of a struggling physics student, or celestial bodies sliding through the depths of outer space – is through the exertion of forces. So it's here that we'll begin, with a delineation and description of the 'four forces of nature'.

In fact, there are neither four of them, nor is the notion of 'force' quite the appropriate one for discussing the phenomenon of causation. But we do need to start somewhere, even if it's in the presentation of a point of view that we'll eventually need to substantially refine. Welcome to the world of particle physics!

Copyright 2001 by Bruce A. Schumm

CHAPTER 3

THE BABY *AND* THE BATH WATER

The Modern Physics Revolution

The mid-1800's was a very fruitful time in physics. Physicists of this era put forth a number of profound advances, including the development of thermodynamics and statistical physics (the physics of bulk systems of matter), and especially, the development and integration of the separate theories of electricity and magnetism into a single, extremely powerful theory of electromagnetism. Even in retrospect, the successes of that period seem monumental, and their impact on our outlook towards the natural world was profound. Scientists had developed experimental tools that could delve well beyond the realm of the five ordinary senses, and in doing so had been able to perceive the deep, underlying principles behind a large number of physical phenomena. The resulting theories were as profound for their mathematical sophistication as they were for the simplicity and economy of their expression. For example, the theory of electromagnetism that emerged can be stated in four simple equations – Maxwell's Equations – which can easily be silk-screened (and often are) on the back of a T-shirt. Of course, these are multi-variate differential equations, which would look impenetrable to someone untrained in higher mathematics. But mathematics is really nothing more than a language; once one knows how to speak this language, the expression of the *physical* content of the theory of electromagnetism is wonderfully simple.

To physical scientists of the late 1800's, these advances seemed so great and encompassing that many felt that a golden age had come and gone, never again to be rivaled. There simply didn't seem to be any questions left to answer whose scope was large enough to instigate the development of new physical theories of such a fundamental and overarching nature. There was a sense that the fundamental workings of nature had been successfully understood and described, and that with the principles developed, essentially any physical process could be explained.

In retrospect, of course, nothing could have been farther from the truth, and indeed the 20th century, the century of 'modern' physics, saw even more profound developments than those of the 19th. The science of the 19th century gave us

confidence that the workings of nature lie within the purview of rational thought and human exploration; the science of the 20th century showed us that, in many instances, this exploration leads our rational thinking into very strange and counter-intuitive realms. At the center of the revolution of modern physics lie the theories of relativity and quantum mechanics, both of which were brought to us in the first 30 years of the 20th century. It is not within our scope to give a complete description of either of these fields; however, we will need to expend some breath on both of them in order to introduce a few notions central to the development of particle physics.

I. EINSTEIN'S RELATIVITY

A full exposition, at this level, of Einstein's special and general theories of relativity would easily be itself as lengthy as this book – and a no less fascinating read. Such expositions do exist – one, in fact, penned by Einstein's own hand(***1); the recent book by Brian Greene mentioned in the previous chapter also has an extensive discussion of the principles and implications of Einstein's two theories of relativity. The first of these subjects – special relativity – is a particularly beautiful one, resting on two deceptively simple postulates: that 1) the speed of light as it moves through a vacuum will always be measured to be exactly 299,792,458 meters per second, independent of the motion of the person measuring that speed; and that 2) the laws of physics are the same for all observers, again independent of the relative motion of the observer. Another way that this second postulate can be stated is that there is no experiment one can do that will distinguish one observer's frame of reference as somehow being 'better' (at or close to being absolutely at rest) than that of any other observer.

From these two seemingly benign statements flow a host of utterly nonsensical but experimentally verifiable conclusions that completely rearrange one's sense of space and time – the very fabric into which the workings of the universe are inexorably stitched. Rulers shorten, time slows, the boundary between the notions of space and time erodes, matter and energy become interchangeable facets of a single physical quantity, 'mass-energy', and so on. It is beyond the scope of this book, but well within the scope of understanding of the reader for whom this book is intended, to describe and motivate these effects. Today, a sophomore physics major can reasonably be expected to obtain a deep and rigorous understanding of

special relativity. We will merely describe the few of the results of special relativity that are necessary for the discussion of particle physics.

To a particle physicist, special relativity is second nature. It is simply a tool of the trade. It is as familiar to him or her as, say, the Federal Tax Code is to a tax attorney. Luckily for the physicist, though, relativity is much less arbitrary and substantially easier to fathom than the tax code.

The most important thing we'll need to take away with us from our discussion of special relativity is the relationship between mass and energy. This relationship, as we'll see, is the following: they are essentially the same thing.

Prior to Einstein, the 'kinetic' energy E_K of an object in motion was thought to be just one half the product of the particle's mass m and the square of its speed v:

$$E_K = \frac{1}{2}mv^2.$$

That this quantity plays a central role in our description of the physical world is primarily due to the fact that energy is *conserved* (an extremely important and well-founded experimental fact), which means that no matter what happens, there's always the same amount of energy around before something happens as there is after it has happened. Thus, if our object moving with speed v collides with some other object or system of objects, whatever kinetic energy our original object gains or loses (by getting slowed or sped up by the collision) will be *lost or gained*, in some form or other, by the object or system of objects with which it collides.

Einstein modified this notion considerably, and the modification is encapsulated in what is certainly the most celebrated relation in all of physics (but which, by the way, is *not* known as 'Einstein's Equation' to card-carrying physicists):

$$E = mc^2$$

where c is just the well known, ever-unchanging speed of light – the velocity of the propagation of Maxwell's self-supporting electromagnetic field disturbances(***2). Remember that the '=' sign is nothing more than mathematical shorthand for the expression 'is the same as'. Forgetting about the factor c^2 for the moment, we see that this equation states quite bluntly that 1) energy is the same thing as mass; 2) mass is the same thing as energy. The factor of c^2 just tells us how *much* of what we

usually think of as energy is associated with the given amount of mass m. The fact that this proportionality constant, the speed of light, is so large (299,792,458 meters per second) means that the amount of energy associated with any given mass is shockingly large – for example, a single gram (one thirtieth of an ounce) of mass, if converted entirely into wall-plug style electrical energy, would approximately equal an entire day's output from a large (Gigawatt) modern-day power plant.

Thus, Einstein tells us that, in addition to the 'kinetic' energy of motion, we must also consider the 'potential' energy associated with the mass of a particle in motion. The particle's total energy is the *sum* of its mass-energy and kinetic energy (the latter of which, by the way, is only *approximated* by the simple form $\frac{1}{2}mv^2$). In a collision, *both* of these contributions to the particle's energy must be taken into account when doing the before/after energy balance accounting required by energy conservation. If conditions are right – say, if the particle is colliding with another particle that just happens to be its antimatter counterpart – the conversion of mass-energy into kinetic energy can be complete. Alternatively, and again if the conditions are right, the conversion of kinetic energy into mass-energy can also be complete. In this latter fashion, relatively light and relatively ordinary particles (such as electrons and positrons) can be hurled against each other with great opposing kinetic energy, only to have their kinetic energy be converted into the mass energy of an exotic, and very heavy, new particle – the kind of new particle which spirits the theory-developers to their blackboards, and the study of which leads to great advances in scientific understanding. This approach has been a central theme in much of experimental particle physics throughout its 50 or so year history.

The connection between mass and energy has become so familiar to particle physicists that they have taken on the habit of quoting the masses of their favorite particles exclusively in terms of the associated mass-energy. In fact, it's even worse than that. The unit (measuring-stick) of energy used by particle physicists is not the Joule or erg familiar to anyone who has taken introductory physics, but rather the 'electron-volt', or eV, which is a unit wholly inspired by the particle physicist's favorite toy – the particle accelerator.

The theory and design of modern particle accelerators is a subject perhaps worthy in and of itself of a few chapters of this book, but would take us a bit farther afield than we could comfortably go. For our purposes, it should suffice to

think of an accelerator as a very large and exceedingly expensive TV set. In a TV set, electrons are accelerated towards the screen by the electric force, where upon impact they make a certain color phosphor dot glow, which represents, say, one of many dots in an image of the face of your favorite tennis player as she winds up to serve. The magnitude of the accelerating electric force, times the distance over which the acceleration occurs, is known as the potential difference, or voltage. An electron accelerating through a given voltage obtains a well-defined amount of kinetic energy (energy of motion); if the voltage is exactly one volt (recall that wall-plug voltage in the United States is 120 volts), the electron obtains precisely one electron-volt (eV) of kinetic energy during the acceleration.

The essential point regarding the electron-volt unit of mass-energy is that an object with a mass-energy of one electron-volt, and without any kinetic energy to start with, would release precisely this much kinetic energy if it were annihilated by its antimatter counterpart.

Note that the electron-volt is a *general* expression for a particular amount of energy, and the energy of *any* object (not just electrons) can be expressed in electron-volts, whether that energy is possessed by the electron in the TV set, the ball that rockets at lightning speed over the net after the serve of a professional tennis player, or the caloric content of the chips and soda that you are mindlessly consuming as you watch the match.

In these terms, the electron, which is the lightest particle whose mass has been measured (although probably not the lightest known particle – more on this in Chapter 5), has a mass of about 511,000 electron-volts. Again, this is to say that, were you to somehow convert the electron's mass-energy to kinetic energy, the amount of kinetic energy released would be 511,000 eV. Similarly, the mass of a proton is about 938,000,000 eV. The heaviest known fundamental particle, the 'top quark' (to be introduced in Chapter 5) has a mass of about 175,000,000,000 eV(***3). This also is about the reach, or kinetic energy per beam particle, of today's most energetic particle accelerators. By way of comparison, the maximum energy of an electron in a TV set is a few thousand eV, so if we insist on thinking of particle accelerators as being like TV sets whose images somehow reflect the fundamental workings of the universe, then these are big TV sets indeed.

Although it has nothing to do with the subject matter of this book, we can't really take leave of our discussion of Einstein's Relativity without at least a mention

of the theory of general relativity. The formulation of this theory took 10 years of Einstein's professional life – from 1905 to 1915 – which in Einstein's own words was a difficult and uncertain period. In the end, though, it was well worth the effort, for what came out was a theory whose reshaping of the special-relativistic notion of space and time – already an almost inconceivable departure from common sense – was as radical and profound a departure from special relativity as the latter was from the classical, pre-relativistic, 'common sense' notions of space and time. The universe of general relativity is one in which the pedestrian, three dimensional world of our senses is twisted and distorted through higher dimensions, in ways not perceivable by our senses, as the curvature of a flat piece of paper into a sphere would not be perceivable to a two-dimensional person living on the sheet of paper. The curvature of spacetime is related to the distribution of mass/energy in the universe via a single equation (not reproduced here because its statement involves mathematical notation beyond our scope) properly known as 'Einstein's Equation'. Again, the general theory is readily verifiable; in fact, it is the only existing theory which correctly and quantitatively describes the observed behavior of the gravitational force. Since the Standard Model of particle physics is a theory formulated in the *flat* (uncurved) spacetime of special relativity, I will make no further mention of the general theory in this book.

II. QUANTIZATION: THE NEXT GREAT LEAP

The second component of the modern physics 'revolution' (Einstein's relativity being the first) was the development of quantum mechanics. Although no single figure played as central a role in the development of quantum mechanics as Einstein did in the development of relativity, one can point to two physicists (both German) whose work and ideas propelled the rapid crystallization of the quantum hypothesis into a fully rigorous and quantitative theory. These two physicists, Erwin Schroedinger (of the 'Schroedinger Wave Equation'), and Werner Heisenberg (of the 'Heisenberg Uncertainty Principle') were largely responsible for the synthesis of a number of disparate and somewhat qualitative notions into a concise and powerful theory during the few short years of the mid-1920's. The remainder of this chapter will be devoted to developing the concepts necessary for a discussion of their work. Of course, to mention only these two does a great disservice to numerous others, Einstein himself amongst them, whose willingness to interpret

certain physical phenomena in clever and radical ways laid the groundwork for the quantum renaissance of the 1920's.

The 'quantum' in quantum mechanics refers to the tendency for the properties which characterize a physical system to be restricted to a limited set of possibilities, or 'states'. When you excite a gas composed of a very large number of similar but free atoms, such as that contained in the tube of a neon light, the gas will begin to glow due to the emission of a very large number of very small flashes of light from the individual atoms in the gas. If you then shine this light through a prism, so that it is broken up into its constituent colors, you will not see the familiar 'rainbow' spectrum which passes continuously from a rich purple through green and yellow to red. Instead, you will notice (if you're careful enough) that the light is in fact composed of a limited number of discreet colors, or 'spectral lines'.

The individual atoms in the gas can not emit light of any color, but rather only of certain discreet, well defined colors. The atomic system is 'quantized', in the sense that we can assign a number (a 'quantum number') to each of these well-defined colors. This number simply acts to delineate, or 'quantify', which of the various possible colors we are seeing when we pick out one particular spectral line emerging from the prism. Note also that in emitting light which contributes to one of these 'permissible' spectral colors, the electrons in the neon atom pass instantaneously between two similarly permissible orbital states. The electrons can't spiral smoothly inward (through an unquantifiable continuum of *impermissible* orbital states) as the light is emitted, but must rather make a sudden jump – a 'quantum leap' – between two discreet *permissible* quantum – 'quantum states'. Since the permissible states are discreet (clearly separated from each other), they can be counted, and so a number (again, a *quantum number*) can be assigned to each permissible orbital state.

In fact, quantum mechanics tells us that all physical systems are quantized – not just atoms. However, in order for the effects of this quantization to be noticeable, the system under consideration must be small. For large systems (such as a piece of iron heated to glow white-hot, which when viewed through a prism appears to emit the full rainbow of possible colors) the possible states are still discreet, but there are so many states that are so closely spaced that you simply can't tell, even with the most sensitive experimental apparatus, that the allowable states of the system do not constitute a continuum of possibilities. For the excited gas of the

neon light, on the other hand, each atom emits its light independently of all the other atoms in the neon plasma, and so the system of interest is really just the single neon atom, which is small enough that quantum behavior is quite evident.

Quantum theory gives us an explicit notion of the meaning of the word 'small', i.e., of the physical scale at which quantum mechanical effects become appreciable. This scale, known as 'Planck's Constant' (h), is, along with the speed of light (c), one of a handful of fundamental quantities which characterize the overall workings of nature. In terms of the everyday, Planck's constant is indeed small, with a value of $h \simeq 6.6 \times 10^{-34}$ Joule-seconds. For those not familiar with common physical units, the 'Joule' is an everyday amount of energy – specifically, the amount of energy required to lift one kilogram (2.2 pounds) about one tenth of a meter (4 inches or so)(***4). The minuteness of Planck's constant – about 33 orders of magnitude (factors of ten) less than one – tells us that quantum mechanics operates on very small scales indeed(***5). Note that the units of Planck's constant are not Joules, but rather Joule-seconds; the significance of this will be discussed at length in later chapters. (Again, if you're unfamiliar with the 'exponential notation' used in the expression 6.6×10^{-34}, please see Appendix A.)

III. OF BALLS AND BEAMS: WAVES VS. PARTICLES

Even more than the odd property of quantization, though, quantum mechanics is concerned with the *wave-like* properties of matter. The window through which the great physicists of the early twentieth century first peered into the odd realm of the physics of very small scales was that of quantization, so it's a bit of a historical accident that the theory that was eventually settled upon to describe physics at this scale was given the name 'quantum mechanics'. A more appropriate name might well have been 'wave mechanics', although this perhaps doesn't have quite the zesty revolutionary connotation that the term 'quantum mechanics' seems to.

It's necessary at this point to clearly establish the notions of 'wave' and 'particle' as they exist in the mind of the physicist. We are motivated here to do this in order to fully appreciate the notion of 'wave-particle duality', which will be introduced shortly. In addition, though, we will introduce a number of properties of waves – particularly that of 'phase' – which shall play an essential role in our discussion of the Standard Model of particle physics in later chapters. So do read carefully!

Particles are hard and discreet, like miniature billiard balls. When two particles collide, they bounce off of each other and go off on their merry ways, with their directions and speeds forever altered. Perhaps more importantly, if someone asks you where a given particle is, you can develop a procedure (such as simply take a picture of it if it's big enough) to answer the question definitively. For a particle, the notion of 'position' is a sensible one. At any given time, the particle is 'localized' at a well-defined position that in principle can be determined by experiment.

Waves, on the other hand, are everything that particles are not. Bob a stick steadily up and down in a pond and a (circular) wave emanates forth, eventually filling the pond with evenly spaced undulations that travel outward from the stick. Exactly where in the pond is the wave? It's not a sensible question to ask. The wave is *not* localized to a single position in the pond; it's everywhere in the pond at once. In fact, the wave really isn't a 'thing' at all, is it? Certainly, the wave carries with it some energy, which can be substantial, as anyone who's been at the beach on a good day can tell you. But the wave itself is nothing more than the molecules in the body of water moving up and down in some organized way. The particles (in this case water molecules) seem to be real objects, taking up space and so forth, but the wave is nothing more than a description of the way in which those particles convey energy (from the bobbing stick or the wind or whatever) around.

Furthermore, if two sticks are bobbed up and down in the pond, the two waves that are created don't bounce off of each other, as particles would be expected to. Rather, they move through each other, *interfering* with one another as they do, but not changing their individual courses. Picture in your mind's eye two equal-sized waves that are passing through each other. If at some point both waves are peaking, then the peak is doubly high (constructive interference); on the other hand if a peak from one wave meets a trough from the other, they cancel each other out (destructive interference) and the pond surface is at the same height as it was before either wave arrived. Interference is an essential wave property, and is one of the features that distinguishes waves from particles. If two entities interfere, rather than bounce off of each other, then those entities must be waves.

In any regard, if waves are not characterized by position, then what are they characterized by? In a nutshell, four things: wavelength, frequency, amplitude and phase. There are other ways to characterize a given wave (such as the speed at which the wave travels through the water), but those other properties can always

be expressed in terms of the four listed above. The following discussion describes these four properties of waves.

Imagine yourself in a sailboat afloat on the ocean. You set anchor so that the boat is at rest, and climb the mast. Looking about you, you see an endless series of waves that are washing into the bow, passing along the boat and then rolling back away from its stern. You want to radio back to your friend on shore, who has a tendency to become sea-sick on really heavy days, the properties of the waves, so that he can decide whether or not to risk it and join you. He's got a lot of experience in this business, so he wants to know everything he can about the waves before he makes his decision.

The first thing that you notice is that the peaks and troughs of the ocean waves are separated by a well-defined distance – the *wavelength* – which is the same whether you're looking at the ripple closest to the boat, or the one a mile distant. You measure the wavelength and write it down. Secondly, you notice the rate at which the ship bobs up and down – the *frequency*, or number of wave peaks which pass by the boat per second. You measure this and write it down. Thirdly, you notice that there is a well-defined difference in height between the peaks and troughs. This difference, known as the *amplitude*, determines just how violently the ship bobs up and down in between successive wave-crests, and so you measure it carefully, for certainly your not-completely-sea-worthy companion will want to know this.

Finally, you recognize that if your friend is to know exactly when it is that the boat will be at the top of its bob, and when it will be at the bottom, you need to measure something about the *phase* of the wave – when your watch reads exactly noon, are you sitting on top of a wave crest, in a trough, or somewhere in between? How, you wonder, will you measure and convey this piece of information in a respectable quantitative fashion?

You are then struck by a sudden insight: riding waves is sort of like going around in a circle. See Figure 3.1: after going around the circle by 360 degrees, you're right back where you started from. You know that you've moved around the circle, but everything looks the same as if you just stayed in the same place. Similarly (see Figure 3.2), after bobbing down one crest and up to the top of the next wave crest, even though you know that you've bobbed up and down one wave, there's nothing you can see that verifies this. You're on top of the wave, just as you were before you

started bobbing. Just like going around the circle by 360 degrees, you're effectively back where you started: at the crest of one of an infinite sea of waves.

You now notice that at precisely noon the boat is exactly at the bottom of a trough, halfway between successive crests, and so you write down your phase – halfway between 0 degrees (the top of one crest) and 360 degrees (the top of the next crest). Since halfway between 0 and 360 degrees is 180 degrees, you write down you phase at noon as 180 degrees. Half-way around the circle puts you on its opposite side; halfway between crests of the undulating wave puts you on the part of the wave that opposes the crest, i.e., it puts you in the trough. There's a direct an exceedingly useful analogy between going endlessly around in circles, returning time and time again to the same point, and bobbing endlessly up and down on ocean waves, returning time and time again to the crest of an endless succession of indistinguishable waves.

From this observation of the waves *phase*, and given your record of the frequency (rate at which the wave crests wash by), your erstwhile friend should be able to figure out exactly when it is that he will find himself at a crest, trough, or anywhere in between. You congratulate yourself for being so clever and complete in your measurements, and radio them back to shore, and then head down below decks for a game of solitaire while awaiting your friend's decision.

After seven consecutive wins, with no one there to share in your triumph, the long-awaited answer comes back from your friend. The wavelength, frequency, and especially amplitude, all seem amenable to him. However, he can't believe that you went to all the trouble to measure and report the *phase* to him - how could it *possibly* make any difference to him whether he's in a trough at 1:01, on a peak at 2:07, or whatever? In fact, he's so discouraged by your apparent lack of critical thinking skills that he's decided to stay on shore and re-evaluate his commitment to his relationship with you.

Well, that's the way friends are sometimes. But, if you're interested in quantum mechanics, and particularly in the story of this book, then he's taught you a very important lesson. The overall phase of the system of waves just doesn't matter. And mechanics takes it one step further: no only does the phase not influence anything you might care about; in fact, in quantum mechanics, there's *no experiment* that you can do to determine the overall phase of a system with wave-like properties(***6). It's not even a observable aspect of a physical system. And since, as

we'll shortly see, *all* systems have wave-like properties, this is an important point of which to take heed.

Paradoxically, it is precisely this disconnect between overall phase and physical relevance which will give the notion of phase its central importance when we finally get around to discussing 'gauge theory' in Chapter 8. Dogged adherence to this principle – this meaninglessness of phase – coupled with its generalization to ensure consistency with the tenets of relativity and quantum field theory (next chapter's topic), lead directly to a profound reinterpretation of the fundamental nature of the interactions of matter. It's not that the phase of quantum mechanical systems becomes relevant in gauge theories, but rather that the very idea of the irrelevance of phase is suddenly understood to be, in and of itself, tremendously relevant. Gauge theory is a notion based on the relevance of irrelevance; within this oxymoronic inspiration lies what seems to be one of the most important intellectual advances in the storied history of particle physics.

IV. WAVE-PARTICLE DUALITY I: PARTICLES OF LIGHT

That light is a wave, and not particle-like, is obvious now that we've sharpened up our notions of what it takes to be a wave or a particle. If you shine two flashlight beams so that they intersect, they *don't* slam into each other and go crashing to the ground. Instead, they pass through each other essentially unaffected, as you would expect of a wave in good standing. Light can be focussed and reflected, and can 'diffract' around corners, filling up illuminating regions that are not in the direct path of the original light beam (this is why shadows of objects tend to have fuzzy edges). All of these properties can be shown to be a result of the essential wave-like property of interference.

As we have seen, waves are in essence an organized form of energy transfer through a 'medium'; in the case of water waves, the medium is just the water at the surface of the pond. What is the corresponding medium for light? The answer (thanks of course to Maxwell, the greatest of the many heros of the theory of electromagnetism) is that light is a self-supporting disturbance of electric and magnetic fields, which propagates through space at, not surprisingly, the speed of light $c = 2.997 \times 10^8$ meters per second, or about 186,000 miles per hour. These fields, when the disturbance reaches the human eye, can induce chemical reactions in the receptors of the human retina; ordinary 'color' is nothing more than the

brain's way of differentiating between different *frequencies* of oscillation (waving) of the field associated with the given ray of light. Brightness is also easily explained - it's simply the *amplitude* of that oscillation. Visible light falls roughly within the range of 4×10^{14} to 8×10^{14} oscillations per second, with the red end of the rainbow at lower frequencies, and the indigo end at higher frequencies. Directly below the lower-frequency end of the visible spectrum lies the realm of the infrared, while directly above lies that of the ultraviolet.

A very hot object – such as the filament of an electrical stove – begins to glow red as it heats up. Most of us also know that, were the filament made even hotter, its color would change from red-hot to white-hot. Whatever the color, though, if you put your hand anywhere in the vicinity of the object (without touching it, mind you), you can sense the heat radiating from the object. The light you see (and the infrared light that you don't see), as discussed above, are electromagnetic waves which efficiently transmit some of the thermal energy of the hot object to the surface of your hand. Experiments conducted in the nineteenth century provided accurate measurements of the 'spectrum' – the relative amount of energy contained in each range of color of radiated light – emitted by hot objects.

This very basic attribute of matter, reasoned the classical physicists, should be possible to understand, and so the physicists of the late 19^{th} century set about to explain it in terms of the classical theory of electromagnetism and the relatively advanced notions of the 'statistical' behavior of macroscopic systems. The problem was that the resulting classical 'blackbody'(***7) theory of the spectrum of the radiation of light from hot objects was in wild disagreement with observations. In fact, this classical theory predicted that an *infinite* amount of energy would be radiated at high frequencies – a failing that came to be known as the 'ultraviolet catastrophe'. This is clearly absurd, if for no other reason than the fact that it only took a finite amount of energy to heat the object up in the first place (remember that energy is conserved, so you can *never* wind up with more than you started with).

After much thought and re-examining of the assumptions that went into this classical blackbody theory, in the year 1900 a soon-to-be-famous German physicist by the name of Max Planck finally hit upon a hypothesis that, when incorporated into the blackbody theory, not only resolved the ultraviolet catastrophe, but also

produced a prediction for the spectrum that was in astounding agreement with experimental observations. Planck's hypothesis was that the energy E transferred by electromagnetic waves (light) of a given frequency (color) f must come in discreet, *quantized* packets of magnitude

$$E = hf$$

where h is, of course, Planck's constant. The bigger the frequency f, the more energy E it takes for the hot object to release a single quantum of light, which is the *minimum* amount of energy the object needs to give up in order to radiate at that frequency. For frequencies that are too high (too ultraviolet), this minimum energy requirement is just too demanding. Thus, instead of the prediction put forth by the classical theory of an *infinite* amount of energy radiated in ultraviolet light, the theory modified by the hypothesis of light *quanta* (flashes of light with a well-defined energy content of hf) led to *no* radiation for very large ultraviolet frequencies. Because it takes so much more energy to create an ultraviolet light quantum, no energy is radiated in the ultraviolet, and the ultraviolet catastrophe is resolved.

It's interesting to note that Planck himself did not consider the quantization of light to be that revolutionary – he just assumed that there was some detail about the way in which the light was emitted that no one quite understood. Five years later, however, in the same year as his publication of the special theory of relativity, Einstein published a paper on the 'photoelectric effect', by which electrons are knocked out of solid materials when certain such materials are illuminated with visible light. Based on the many detailed observations of this effect, performed and published by a number of experimental researchers, Einstein was able to convincingly argue that 1) this phenomena was explicable only if the quantization of light into individual packets of energy $E = hf$ was a fundamental property of light itself, and not just the emission process; and 2) these individual packets of light *behaved as if they were particles*, i.e., each released electron was knocked out of the material by a single packet, or quantum, of light, which, in order to knock the electron out of the solid, had to collide with the electron in the jarring manner of a particle. This radical idea was eventually accepted by the international physics community; soon afterward the American physicist Arthur Compton coined the term 'photon' for these indivisible particles of light. It's interesting to note that Einstein's 1921 Nobel Prize, which followed three years after Planck's own Nobel,

was awarded more for Einstein's explanation of the photoelectric effect than it was for his theories of relativity.

So, at the end of the day, we need to to put aside our calculations, switch off the voltage to our experimental apparatus, grab a cup of tea, and reflect for a moment on what we have learned. On the one hand, light is clearly a wave – the evidence for this comes directly from our everyday experience with flashlights, lenses, and especially these days, lasers. But on the other hand, when you look a little more carefully, at phenomena where Planck's nominally tiny constant sets the scale of activity, there's plenty of experimental evidence that demands that we treat a ray of light as if it's composed of a large number of light quanta *particles*. So, which is it – wave or particle, particle or wave? Which of these two apparently exclusive poles is representative of the fundamental nature of light? The answer, Einstein argued, is *both*. In some contexts, light behaves as if it is composed of little particles flowing along together. In other contexts, it behaves like a wave – something our intuition has a difficult time associating with a 'real' object that can bounce off of other things. Yet the wave *is* the particle, and vice versa. Light has a dual nature – part wave, part particle. This, of course, is the notion of wave-particle duality, a notion central to quantum mechanics in general, and particle physics in particular. Indeed, the photon, and particles like it, play a central role in the Standard Model of particle physics, and will figure heavily in the discussion to come.

Finally, if you have an electric stove, go down and turn it to high until it glows red. (If you have a gas stove, you can hold a needle in the flame until the needle turns red.) Take a good look at the light. Recall the discussion of this section – the explanation of the phenomenon you are observing, as simple and benign as it may seem, launched what is probably one of the greatest revolutions in the history of human thought.

V. WAVE-PARTICLE DUALITY II: MATTER WAVES

In 1924, a French prince and graduate student by the name of Louis-Victor de Broglie was stricken by a sudden insight. A common conviction of physicists is that, at its most fundamental level, nature prefers uniformity and simplicity over disparateness and complexity. Thus, he hypothesized that Einstein's conjecture of the dual nature of light represented merely the tip of an iceberg, and that the notion of wave-particle duality should extend to *all* of nature. If light, which our

common sense strongly suggests should be classified as a wave, can exhibit particle-like properties, then why can't matter particles, such as electrons, protons, atoms, molecules, dust, cats and dogs, and so forth, exhibit wave-like properties?

According to Einstein, the connection between light's wave-like property of frequency f and its particle-like property of energy of motion (kinetic energy) E is given by Planck's relation $E = hf$. Simply stated, de Broglie's hypothesis was that this relation should also hold for matter. If we look carefully enough, de Broglie argued, we should be able to convince ourselves that matter can exhibit wave-like properties, consistent with this relation between the object's energy of motion and its frequency.

In fact, this is not quite the form in which de Broglie's hypothesis is most commonly stated and most readily applied. Rather than being expressed in terms of kinetic energy (energy of motion) and frequency, de Broglie's hypothesis is usually expressed in terms of wavelength and momentum. We already know what wavelength is; throughout this book we shall studiously avoid drawing the mildly technical distinction between momentum and kinetic energy. They are correlated – whenever a particle's energy increases, so does its momentum. If you're not already schooled in the difference between these two, feel free to think of them as being one and the same, unless explicitly instructed otherwise.

In any regard, it's the more common form (in terms of momentum and wavelength) which will be of most use to us. Consider a bunch of wave crests, all separated from each other by some distance λ (the wavelength), washing by your boat with some speed v. With a little head scratching, you can convince yourself that the frequency f of this wave – the number of wave crests that pass the boat every second – is just

$$f = \frac{v}{\lambda}.$$

Note that the bigger the speed v, or the smaller the wavelength λ, the more crests per second will pass by the boat, and the higher the frequency that will be observed. So this relation makes sense.

Now, energy/momentum is proportional to frequency according to $E = hf$, but frequency is inversely proportional to wavelength according to $f = v/\lambda$, so energy/momentum is inversely proportional to wavelength. Thus, we can write (remember that we're avoiding the distinction between energy E and momentum

p)

$$p = \frac{h}{\lambda}$$

or, equivalently,

$$\lambda = \frac{h}{p}.$$

Whether or not you followed the twists and turns of this argument, take away the following message. If de Broglie is correct, matter particles should exhibit wave-like properties. The wavelength of the associated matter wave should be *inversely* proportional to the particles energy – the bigger the energy, the *smaller* the associated wavelength. The constant of proportionality is just Planck's constant h, which we know to be a very small number.

However, what's small to a human is not necessarily small to an atom. In fact, de Broglie's relation tells us that the wavelength of an electron that has been accelerated through 50 Volts of electrical potential (i.e., with an energy of 50 electron-Volts; recall our discussion in the section on Einstein's relativity) is about 2 Angstroms, or about 2×10^{-10} meters. This, it turns out, is roughly the spacing between the atoms of a crystal. It also turns out that the regular, uniform spacing of atoms in a crystal lattice are ideal for studying the effects of interference – the phenomenon which is so uniquely characteristic of wave-like behavior.

In 1927, the American physicists Clinton Davisson and Lester Germer were studying the reflection of 50 electron-volt electrons off of a plug of pure nickel. A minor failure of their apparatus led to a contamination of the nickel sample, and so they had to re-purify the sample by heating it up. Although they didn't immediately recognize it, they had crystallized the nickel sample in the process. When they again began studying the reflection of the 50 eV electrons, they noticed a very peculiar effect - the reflection was much weaker in general, although for particular angles of reflection, it was very pronounced. Davisson quickly realized that such behavior would be expected if the nickel sample had crystallized, and *if the electrons were exhibiting wave-like behavior* – if the electron waves reflected from each of the individual, regularly-spaced atoms in the crystal were *interfering* with one another to form the pronounced reflection pattern that was observed. This would only work if the wavelength associated with the electrons' wave-like behavior was roughly that of the spacing between the nickel atoms in the crystal lattice. As

we have just seen, this is precisely the prediction of de Broglie's hypothesis, and so, naturally, both de Broglie (1929) and Davisson (1939) eventually found themselves in the company of the Swedish King, receiving their respective Nobel prizes.

So, waves (such as light) have particle-like properties, and matter particles (such as electrons) have wave-like properties. The concept of wave-particle duality is, as de Broglie conjectured, universal. Quite generally, *any* corporal object has associated with it a wavelength, which can easily be calculated via de Broglie's hypothesis. For instance, I estimate the wavelength of this book, traveling with the momentum required to propel it in a perfect arc on its way to the trash can in the corner of your room, to be about 10^{-35} meters.

As a sidelight, this is a development with a very practical application. Have you ever looked through a conventional microscope and seen an atom or a molecule? It's impossible, even with the best possible microscope. The reason, it turns out, is that you can't see any feature of the sample under study with a size smaller than the wavelength of the wave you're using to illuminate the sample. Visible light has a wavelength of between 4×10^{-7} and 7×10^{-7} meters, while atoms and molecules are about 10^{-10} meters across – way too small to be seen with visible light. On the other hand, as we've just seen, a relatively languid 50 eV electron has a wavelength of about 2×10^{-10} meters – increase the energy a little (electron beams in TV sets are typically a few thousand eV), and develop a way to focus electron beams (say, with magnetic fields) and now you have a hope of doing ultra-precise microscopy. Thus, the field of electron microscopy is born.

But why stop there? What happens if, instead of a 1,000 eV electron, you bombard a sample with, say, 100 million eV electrons from a particle accelerator? The corresponding electron wavelength will be much smaller (recall that wavelength is *inversely* proportional to momentum/energy) – in fact, it will be about 10^{-15} meters. Correspondingly, experiments done by Robert Hofstadter on the campus of Stanford University in the late 1950's were the first to 'see' the that the proton was not point-like, but in fact had a measurable radius – of about 10^{-15} meters. Buoyed by this success (which was further bolstered by Hofstadter's receipt of the 1961 Nobel Prize in physics), a much larger accelerator was assembled – the Stanford Linear Accelerator Center, or SLAC – on Stanford University land. The accelerator began operation with electron energies of 10 billion eV, and corresponding electron wavelengths of 10^{-17} meters. This allowed physicists to peer very deeply inside of

the proton, leading to the discovery of its internal constituents, known as 'quarks' (to be introduced in some detail in Chapter 5), in the late 1960's. Needless to say, this whole operation has kept the King of Sweden somewhat engaged over the years.

VI. HEISENBERG'S UNCERTAINTY PRINCIPLE

In the late 1970's, when I was in college, a popular piece of graffiti, found in most bathrooms in college science buildings, was the following: 'Heisenberg may have been here' (anon.). The purpose of this section is to provide the background necessary to understand this quip. Beyond that, it is purely a question of taste as to whether or not you consider it clever enough to merit the desecration of so many bathroom walls.

The necessity of the uncertainty principle follows directly from de Broglie's assertion that matter should possess wave-like properties. Consider an object whose momentum/energy is known *exactly* - not just 'very precisely', but really, truly, *exactly*. Then, according to de Broglie, its wave-like nature (or, in the lingo of quantum mechanics, its 'wavefunction') should be characterized by a wave of wavelength $\lambda = h/p$. Again, λ is the wavelength (say, in meters, yards, or whatever), p is the object's momentum (or energy, if you prefer not to worry about the distinction), and h is, as always, Planck's constant.

As we discussed before, though, a wave is not 'localized' – there is no point in space to which you can point and exclaim 'see - the wave is right there'. Recall the pilot bobbing up and down in the boat – the wave of wavelength λ extended as far in front and behind as the eye could see. The wave was characterized by its wavelength, frequency, amplitude, and (somewhat irrelevant) phase, but *not* its position, for it had no definable position. Thus, if de Broglie is correct, reasoned Heisenberg, a particle with a precisely known momentum p must exhibit the properties of a wave with wavelength $\lambda = h/p$, which is completely un-localized: *if the momentum of an object is exactly known, than absolutely nothing can be known about its position.* The exact value in meters of the wavelength λ is not material to the discussion. The point is simply that if the particle's wave-like properties correspond to a definite wavelength λ, then the particle behaves like a pure wave, which is completely un-localized (at any given time, the undulations of a pure wave

extend infinitely far forward and backward in space), and so nothing whatsoever can be said about its position.

The above is not quite the Heisenberg uncertainty principle, but is rather just a special case of it, for what if the momentum is *not* known precisely, but is known instead to lie within some range. In other words, the object's momentum is known to some degree, but there's some *uncertainty* about what its momentum really is. For example, say you have a machine that produces objects that have momentum between 0.99 and 1.01 as measured in kilogram-meters per second (kg-m/s), the standard yardstick for measuring momentum. You know that the machine produces objects with momentum of about 1.00 kg-m/s, but for any given object, you will be uncertain of the momentum by about 0.01 kg-m/s. Now, in this case de Broglie is not so incisive - he is *not* able to tell us that the wave-like properties (wavefunction) of any given object from the machine will be be characterized by a wave of a single wavelength $\lambda = h/p$. Instead, the de Broglie relation tells us that the wave function will be composed of the combination of a number of independent waves, with wavelengths varying between $h/(0.99)$ and $h/(1.01)$. If there's a range of momenta p at play, then there must be a similar range of wavelengths λ at play.

What's meant here by the expression 'combination of different waves' is fairly easily described. Think of two kids on opposite sides of a pond, exciting waves of slightly different frequencies by bobbing sticks up and down in the pond at slightly different rates. From each stick, a wave will emanate outward; when these waves meet, they will add together, or combine, to form a more complicated wave form.

Anyone who's ever played around with the mathematics of adding together waves of different but closely spaced wavelengths can perhaps anticipate where this argument is leading. Others will just have to take our word for it, unless they have the resources and energy to play around with it themselves (having a computer and knowing how to program it to graph various combinations of sine waves is helpful in this regard). In any case, when you add a few waves together, with wavelengths that vary, say, between $h/(0.99)$ and $h/(1.01)$ meters, the result does begin to become localized. Even though each individual wave is completely unlocalized, undulating infinitely far backwards and forwards in space, the *sum* of the waves of closely-spaced wavelengths does exhibit some localization, i.e., regions where all the waves add together surrounded by regions where all the waves cancel each other out. The more waves you add together with wavelengths between $h/(0.99)$

and $h/(1.01)$ meters, with correspondingly closer-spaced wavelengths, the more pronounced this localization becomes. In fact, there are really an *infinite* number of possible wavelengths between $h/(0.99)$ and $h/(1.01)$ meters; when you allow your mathematics to reflect this, the localization becomes complete – the particle is confined to a single region of space, which is the only region of space where the (infinite number of) waves contributing to the overall wave-like property (wave function) of the particle do not cancel each other out when added together.

The accompanying diagrams of Figure 3.3 illustrate this point – as the number of waves added together with wavelengths between $h/(0.99)$ and $h/(1.01)$ meters increases (with correspondingly less difference between the individual wavelengths of the waves being added together), the resulting wave function becomes more completely localized.

The essential point of the uncertainty principle is as follows. Even when you correctly consider the possibility that the de Broglie wavelength of the object can be one of an infinite number of possibilities between $h/(0.99)$ and $h/(1.01)$ meters, and the localization of the object is complete, this only means that the object is known (at any given point in time) to be in a single *region* of space, not at a single *point* in space. The last of the diagrams in Figure 3.3 showing the complete localization which derives from adding together an infinite number of infinitely closely (in wavelength) spaced waves, demonstrates this. In this case – with the most complete localization allowed for this range of momenta (0.99 to 1.01 kg-m/s) – the object's position is not described by a single, well-defined point, but rather the small *region* of space defined by the extent of the hump in the object's wave function. The particle's in there somewhere, but within those bounds, the particles position is uncertain – not merely as a practical matter, but *fundamentally*.

In other words, all you can say about the position of the particle is that it lies somewhere within the hump shown in that last diagram. If you *measure* the particle's position by, say, bouncing light off of it, you will indeed find it is at some well-defined point within the hump, but the only prediction you can make ahead of time (before the measurement) is that the measured position will be somewhere inside the hump. The particle's position (before you disturb its wavefunction by trying to measure it) is uncertain to that degree.

So, even with this complete localization, allowed by the fact that for the full wavefunction you're adding together an infinite number of waves with wavelength

between h/(0.99) and h/(1.01) meters, there's *still* uncertainty in the position (location) of the object. And now the critical point: the property that determines the size of the region of localization of the object, and thus the uncertainty in the objects location, is the *magnitude of the uncertainty in the object's momentum.* The *less* uncertain the momentum, i.e., the better the momentum is known, the *larger* the region of localization, i.e., the wider the hump in the last plot shown in the figure. Perhaps this point is made more clearly by simply looking at Figure 3.4, which shows the localization achieved by adding together an infinite number of waves with infinitely- closely-spaced wavelengths between h/(0.995) and h/(1.005) meters – a range (uncertainty) in momentum only half that of Figure 3.3. You can see that the corresponding range (uncertainty) in position is *double* that of Figure 3.3.

In fact, this all fits in nicely with our discussion above for the case that the momentum is known exactly – in this case, the uncertainty in momentum is infinitely small (zero), so the uncertainty in position is infinitely large, i.e., nothing at all can be said about the position.

Since when the uncertainty in the momentum grows the uncertainty in position shrinks, and vice versa, then their *product* – the result of multiplying the two uncertainties together – might reasonably be expected to stay the same, regardless of how big the uncertainty of either is. This, in fact, is the quantitative message of the uncertainty principle. If we let Δp represent the uncertainty in the object's momentum (about 0.01 kg-m/s in our example), and Δx represent the uncertainty in its position, then Heisenberg's arguments tell us that (via the application of somewhat advanced mathematics not appropriate for our level of discussion) (***8)

$$\Delta p \Delta x = \frac{h}{4\pi}$$

where $\pi = 3.14159...$ is, as usual, the magical ration between the circumference and diameter of a circle. In particular, if Δp becomes zero (no uncertainty on the momentum, i.e. momentum known exactly), then Δx must become infinity in order to compensate, and nothing at all is known about the position x. Of course, on the other hand, if the position is known exactly, than nothing whatsoever can be said about the momentum of the object.

The uncertainty principle is one of the profound revelations of quantum mechanics, and it ripples throughout the field as a fundamental tenet that can not be

violated. Any result of the full theory of quantum mechanics must be demonstrably consistent with the uncertainty principle.

Perhaps more interestingly, though, it is a revelation that seems to address questions of deep philosophical import. With the advent of the uncertainty principle, determinism – the notion that the laws of nature set forth an inextricable course of events from which no deviation is possible – becomes indefensible. According to the uncertainty principle, the exact course of events is *fundamentally* unknowable. There is always some uncertainty in the physical properties of any given object, and not even nature herself knows how this uncertainty will resolve itself the next time the object makes its influence known – say, by the interaction with another object via one of the four forces. It's not just a matter of building a better instrument for determining these properties – the exact value is simply unknowable, even in principle. Many have gone on to conjecture that Heisenberg's uncertainty principle is the very source of human free will, although this remains to be demonstrated. In any regard, Heisenberg's principle itself is on such solid empirical footing that it is now accepted as one of the fundamental aspects of the natural world.

Another interesting and philosophically important notion which follows from the uncertainty principle is the loss of distinction between the observer and the observed.

To say that an object whose momentum is known to be within a certain range will have a corresponding (according to the uncertainty principle) uncertainty in its position is *not* to say that the object's position can never be determined with greater precision. Consider an object whose momentum is known rather precisely, so that the uncertainty in its position at any given time is as large as, say, one meter. You want to measure the position of this object more accurately, so you place a chamber of some special gas in the path of the particle. During the course of the particle's motion through the gas, it collides with a gas molecule, causing the struck molecule to emit a flash of light. Now, at the exact time that you observe the flash of light, you know that the object is in the vicinity of that particular gas molecule. You now know the position of the object to great precision – to about 10^{-10} meters, which is the size of a gas molecule. However, the uncertainty principle can not be violated, and so your knowledge of the momentum of the particle is now correspondingly much worse.

The point is this: in order to observe the particle, *it must interact with something*

from the system of the observer (in our case, the observer's system is the gas, the apparatus that detects the light flash, the electronics that amplifies the signal from the apparatus, and the human or computer that records the amplified signal). That interaction itself *necessarily* influences the object – alters the *wave function* which governs the object's physical properties, in such a way that the uncertainty principle is never violated. One can never precisely and simultaneously determine both the position and momentum of an object, for in determining one, the process of observation always changes the other to some new, and indeterminate, value. The physical system of the observed object and the system of the observer become thus intertwined in the process. If you want to make an observation of some physical system, you can't just consider the properties of that system in isolation; to really understand what's going on, you have to consider the properties of the observed system *and* the properties of the observing system, and how they interact. The true system under consideration must always be the *combination* of the system of the observed with that of the observer.

In quantum mechanics, physical quantities whose uncertainties are linked via the uncertainty principle are known as 'conjugates'. Position and momentum are conjugate, of course, but there are also a number of other conjugate quantities. Energy and time are conjugate – for example, the certainty with which you can determine the mass/energy of an unstable particle is inversely related, according to the uncertainty principle, to the amount of time it takes the unstable particle to decay into stable byproducts. Angular position and angular momentum (to be discussed in the next chapter) are also conjugates.

Finally, back to our little joke. Since any everyday object such as, say, Heisenberg himself, is known to have a momentum within certain reasonable bounds, then there's a corresponding uncertainty in its (Heisenberg's) position. So, where exactly is Heisenberg? It's not clear - there's a measure of uncertainty in his position. With some probability, albeit vanishingly small for most locations, he could be anywhere. And so maybe, just maybe, he could appear spontaneously in the bathroom of the science building of a small college in suburban Philadelphia. And if he actually *had*, let me tell you, we would have shown him a good time – you can be certain of that!

VII. A MASTER FORMULA FOR QUANTUM MECHANICS: THE SCHROEDINGER EQUATION

What we have described above are some aspects of the quantum mechanical behavior of *free* particles – objects executing an unwavering, linear motion through space, absent of any influence of the forces of nature. Of course, nature is only interesting to the extent that things can influence each other through these forces, so our theory of the quantum behavior of objects must be broadened. This was first accomplished in 1926 by a 40-year-old Austrian physicist at the University of Berlin named Erwin Schroedinger.

Schroedinger did not find it most natural to introduce forces into the newly emerging quantum theory directly in terms of their 'push' and 'pull' on the object being described, but rather in terms of the 'potential energy' associated with the force doing the pushing or pulling. We have discussed and made frequent use of the notion of energy and momentum, without making the distinction between the two. We'll continue to ignore the distinction, but we do need to generalize our notion of energy somewhat. Any object in motion has momentum, or 'energy of motion', associated with that motion. The more (faster) the motion, the greater the energy/momentum associated with that motion. You may recall that in the beginning of this chapter we gave the name 'kinetic energy' to this energy of motion.

On the other hand, consider a car out of gear (in neutral) coasting up a hill, and ignore the friction in the wheel bearings and between the tires and road. As the car climbs the hill, it loses speed. Its motion slows, and so it loses kinetic energy. But remember – energy is a *conserved* quantity. The car must have as much energy at the bottom of the hill, when it was moving fast, as it does at the point on the hill when it stops and begins rolling backwards down the hill again.

So, what happens to the energy that was in the form of kinetic energy (energy of motion) at the bottom of the hill? The answer is that it gets converted to gravitational *potential* energy as the car coasts uphill, slowing as it rises against the pull of gravity. At exactly the point at which the car reaches its maximum height and begins to coast back down the hill, *all* of the kinetic energy has been converted to gravitational potential energy, and the amount of potential energy possessed by the car at that point is exactly equal to the kinetic energy the car possessed before it started coasting up the hill.

The reason that this gravitational energy is called 'potential' energy is that it has the *potential* to do work on the car, restoring the car's original kinetic energy. This is clearly demonstrated by the fact that the car, when it reaches the bottom of the hill after coasting back down, has the same speed as it did just before it started up the hill (ignoring the effects of friction, of course). The gravitational potential energy has fully realized its 'potential' to convert itself to 'kinetic' energy of motion. Of course, if we had slipped blocks behind the wheels of the car just as it reached its maximum height on the hill, we would have locked that potential energy in place. But it would still be there, a very real aspect of the car's physical state, just waiting for some soul to remove the blocks so that it could exercise its potential to expend itself in the noble task of creating kinetic energy.

We can say one more thing about potential energy: it depends upon the location of the car. In the language of mathematics, we would say that the potential energy *is a function of* the car's location. In this case, the location of the car at any point in its coast is just the distance x along the road that the car has traveled from the bottom of the hill. At the bottom of the hill, the potential energy is zero – no kinetic energy has yet been transmuted to potential energy. As the car rolls up the hill, at each successively higher point the speed, and thus the kinetic energy, is less. This is precisely because the value of the gravitational potential energy is correspondingly greater at each successively higher point. The higher the car is, the more potential there is for gravity to do work on the car as it eventually drifts back down to the bottom of the hill.

Mathematically, again, if we denote the value of the (gravitational) potential energy by V, then we would write $V(x)$ to represent this potential energy function. The expression '$V(x)$' is nothing more than mathematical shorthand for the statement 'the potential energy V has a well defined value for any given location x of the object, which I'll tell you if you let me know what position x you're interested in'. In our example, of course, the potential energy is gravitational potential energy and the object is the car. However, in general, the potential energy could be that due to any of (or any combination of) the four forces, and the object anything which bears some amount of the charge appropriate for that force (recall that *mass* is the charge that catches the fancy of the gravitational force)(***9).

What follows, at last, is the full mathematical expression of the Schroedinger Equation, as you would see it in a physics textbook. To be candid, this will be

a restricted form of the Schroedinger Equation, for the case that the potential energy function $V(x)$ does not vary with time – in our analogy, this would mean that the slope which the car climbs is fixed, and not undulating up and down as it would, say, during a major earthquake. In addition, this form of the Schroedinger Equation is appropriate for one dimensional motion only – the car is restricted to move forward and backward along the road, and can't veer off the road into the second dimension (no four-wheeling, please!), or rise towards the heavens into the third dimension(***10). Nevertheless, what follows is a very useful relation, and one that physicists have spent much time working with over the years. The author is fully cognizant of the fact that this equation may have little meaning to readers of this book:

$$-\frac{h^2}{8\pi^2 m}\frac{d^2}{dx^2}\psi(x) + V(x)\psi(x) = E\psi(x).$$

This is what's known as a 'differential equation', which one typically studies in college only after having mastered calculus, so it's a bit beyond the scope of this book, for which the author admits the possibility that even high school mathematics has fallen by the understandable wayside of neglect through disuse. So, some explanation is in order.

The expression '$\psi(x)$' is nothing more or less than the celebrated wave function. It will be discussed in a little further detail below. The number 'E' is just the total energy (kinetic plus potential) of the particle. Remember that energy – really *total* energy – is conserved. So, E is just some fixed number – it doesn't depend on x (where you are in space), or on time.

This differential equation contains in it everything that can possibly be expressed about a given physical situation involving an object, of mass m, under the influence of some force, whose potential energy function is $V(x)$. If you think of this differential equation as a game that physicists play, then the way they would win the game is by finding all possible wave functions $\psi(x)$ that solve this equation, *given the mass m of the object under study, and the particular function* $V(x)$ *which describes the exact way in which the forces at play influence the object.* It's really the function $V(x)$ that specifies the physical situation at hand – an electron under the electromagnetic influence of a proton in a hydrogen atom has a certain $V(x)$; a car coasting up and down the hills of a country road has a different $V(x)$, etc.

What, exactly, does the wavefunction $\psi(x)$ represent? Remember what a 'function' is: $\psi(x)$ tells us that, associated with any position x, there is some specific number, which is of course just the value of the object's wave function at that point in space. Since the Schroedinger equation represents all that can be stated about the particular situation at hand, then the wave function $\psi(x)$ has somehow magically coded within it all that can possibly be known about the physical state of the particle, and incorporates all the tenets and constraints of quantum mechanics, which are somehow contained within the Schroedinger Equation and its rules of solution. The wave function $\psi(x)$ is an extremely economical encoding of this physical information. Although $\psi(x)$ itself has no physical meaning, any physical property of the object can be determined once $\psi(x)$ is known. If you want to know the probability of finding the object at any point in space, you simply perform a specific procedure on $\psi(x)$ – in this, case, just squaring (multiplying it by itself once) the value of $\psi(x)$ at that particular point in space. If you want to know the object's energy, you perform a different procedure (this one involves taking some derivatives, i.e., doing a little calculus). If you want to know the object's speed and direction of motion (velocity), there's a procedure for finding that, and so forth.

A few closing notes are in order. First, if the particle is perfectly free, i.e., *not* under the influence of any force, then there is no potential energy $V(x)$, or, in other words, $V(x) = 0$, and you can forget about that term in the Schroedinger Equation. It turns out that, in this case, the solutions $\psi(x)$ to the Schroedinger equation are nothing more than the pure waves with which we began our discussion of quantum mechanics, with a wavelength given by $\lambda = h/p$ – just the original conjecture put forth by de Broglie which started this whole mess. So, the Schroedinger Equation does just what we wanted it to – it allows us to generalize de Broglie's hypothesis to the case where the particle is not free, but is rather 'waving' around (quantum mechanically) under the influence of a force.

Second, since this theory is called 'quantum mechanics' (although, as mentioned above, it may well be better described as 'wave mechanics'), we'd better discuss where the notion of 'quantization' fits into the picture.

When we do put in the forces via the Schroedinger Equation, and require that the solutions $\psi(x)$ meet certain obvious standards of decency (such as not being infinitely large anywhere in space), we find that, for our given $V(x)$, *solutions to the Schroedinger Equation exist only for certain values of the total energy E.*

This general property is true for *any* potential energy function $V(x)$(***11); the particular allowable values of E depend, of course, on $V(x)$, but the fact that only certain, *discreet* values of E work is a general property of solutions $\psi(x)$ of the Schroedinger Equation. Thus, the Schroedinger theory is consistent with observed quantum behavior, such as the fact that only certain colors (energies) of light can be emitted by a given atom – those energies correspond to the change of the quantum state of the atom between two of the allowable states (energies) of the system. The energy of the emitted light photon is just the difference between the energies of states before and after the change (energy conservation again!). Since only certain energies are allowable for the states, then only certain energies (colors) are allowable for the emitted light.

By the way, the emitted colors are characteristic of the type of atom that's emitting (or absorbing) the light. So, for example, this is the way the elemental make-up of stars can be inferred by analyzing their light as it reaches earth.

Finally, it's interesting to note that the Schroedinger Equation consists of three 'terms': two to the left of the equals sign (separated by the '+' sign), and one to the right of the equals sign. The first term is nothing more than the mathematical representation of the procedure which, once you know $\psi(x)$, tells you how to determine the kinetic energy that the particle has at any location x. If, for a given $V(x)$, you plug one of the functions $\psi(x)$ that solves the Schroedinger Equation into this first term, what you will wind up with after executing the procedure implied by the mathematical symbols is the value of the function $\psi(x)$ again, but multiplied by a number, which number is just the value of the kinetic energy that the particle has when it's at the location x. Similarly, it's easy to see that the second term, to the right of the '+' sign, is just the *potential* energy times the value of the wavefunction at the location x. The third term, to the right of the '=' sign, is just the total energy times the wave function $\psi(x)$ at x.

Thus, if we look at the factors that multiply the wavefunction in the Schroedinger Equation, we find that to the left of the '=' sign we have the sum of the kinetic plus potential energies at the point x, while to the right of the '=' sign we have the total energy. Thus, the Schroedinger Equation is nothing more than the wave-mechanical statement that the sum of the kinetic and potential energies at any given point is just equal to the total energy, i.e., the Schroedinger Equation is simply the quantum mechanical version of the notion of energy conservation. From

this quantum-mechanical formulation of the notion of energy conservation arises the full set of constraints that shape the possible quantum mechanical wave functions. This again illustrates the central importance of the notion of energy conservation – a recurrent theme throughout the branches of physics(***12). Stay tuned, for in Chapter 7, with the help of Einstein's mathematician friend Emmy Noether, we'll actually be able to gain some insight regarding the fundamental physical origin of this indispensable principle.

VIII. PARTING WORDS

This has been a meaty chapter, wherein we have endured the introduction of a long sequence of new and not necessarily intuitive concepts. In order to fix up a few problems that nagged physicists around the turn of the twentieth century, we had to introduce several notions which fundamentally reshaped our view of what's really going on around us. To get rid of what appeared to be a little slurry at the bottom of the tub, we had to slosh our theoretical basin so violently that the baby – the prized 'common sense' notions of classical physics – went right out the window with the bathwater.

It is a testimony to the essential role played by quantum mechanics that its development spawned a number of entirely new fields in physics, including atomic and molecular physics, the physics of solids (which includes the physics of semiconductors and micro-electronic devices), and of course particle physics. It was with the development of these fields that the phenomenon of specialization arose, and scientists who contributed broadly across several fields of physics became more and more rare (but not extinct – the Italian-American physicist Enrico Fermi being a particularly notable exception).

Beyond this point, our discussion must respect this trend, and we'll need to leave the exploration of these other rich fields for another time. We can do this without regret though, for what lies ahead on our chosen path is certainly worth the journey. Indeed, our very next destination – quantum field theory – is nothing less than a further development of basic quantum theory, considered by some to be as profound a leap forward in our understanding of the fundamental workings of nature as the original quantum theory itself.

****

(***1) See Einstein's book 'Relativity – the Special and General Theory'.

(***2) It's no coincidence that Einstein originally formulated special relativity to address apparent logical inconsistencies in Maxwell's theory of electromagnetism having to do with how observers in different reference frames (i.e., in motion relative to each other) would have to divide up observed electromagnetic forces between their electric and magnetic contributions. In fact, Einstein's original paper on special relativity, published in the German journal *Annalen der Physik*, was entitled 'Zur Elektrodynamik bewegter Körper', or 'On the Electrodynamics of Moving Bodies'. Einstein demonstrated that the logical inconsistencies were not a failing of Maxwell's theory, but rather of the common-sense notions of space and time.

(***3) To avoid writing numbers which are too big and cumbersome, we need a shorthand. One thousand eV will be denoted '1 keV' (kilo-electron-Volt), one million '1 MeV' (mega-electron-Volt), one billion '1 GeV' (giga-electron-Volt), and one trillion '1 TeV' (tera-electron-Volt).

(***4) Recall that, just above, we discussed another unit of energy - the electron-Volt (eV). One eV is about 1.6×10^{-19} Joules - quite small. But individual charged particles (electrons, protons) are pretty small themselves, so you wouldn't necessarily expect them to obtain energies which are on the everyday scale of the Joule.

(***5) A wonderful book, *Mr. Tompkins in Wonderland*, somewhat whimsically speculates as to what the world would be like if Planck's constant were much larger – close to 1 – and the speed of light were much smaller – say, a few meters per second. The book was written by the Russian-American physicist George Gamow, who was one of the more prominent contributors to the development of the quantum theory of particle interactions. After a successful international career, Gamow spent the last years of his professional life at the University of Colorado at Boulder. From the robustness of the prose in *Mr. Tompkins*, one can speculate that he enjoyed the skiing.

(***6) Most precisely, if two or more wave-like systems are interfering, then the *relative* phases of the interfering systems are important factors in determining the physical properties of the system. However, in this case the *overall* system is the full many-body system, containing all the interfering sub-systems, and the phase of this *overall* system can have no physical relevance.

(***7) The term 'blackbody' refers to the fact that the calculation of the color spectrum of hot objects is most easily done for materials which, when cold, absorb

all electromagnetic radiation (light) which hits them, i.e., for objects which are perfectly *black* before they are heated to temperatures at which they begin to glow.

(***8) In fact, strictly speaking the '=' sign in the expression of the uncertainty principle should really be a '$\geq$' sign – $h/(4\pi)$ is really the *minimum* possible value of the uncertainty product; the true value depends in a very technical way on exactly how the localization is achieved. For most physical systems, though, the uncertainty product is within a factor of two or so of this minimum.

(***9) Strictly speaking, we can define a potential energy for any force which has the property that the net amount of work – the difference between the work done by the force on the object and the amount of work done by the object against the force – is exactly zero around any path which begins and ends at the same place (think of moving along such a closed path on the slope of a hill – whenever you are going downhill, gravity is doing work on you, while when you are going uphill, you are doing work against gravity; when you get back to where you began on the closed path, the total amount of work you did against gravity minus the amount of work gravity did on you is exactly zero). It can be shown mathematically that, for any such force, if you know object's potential energy at any point in space, you can determine the force (extent and direction of the 'push' or 'pull' exerted by the force on the object) at any point point in space. All four of the fundamental forces have the above property, and so can be incorporated in Schroedinger's quantum mechanics via their potential energy functions.

(***10) In fact, to be *totally* candid, what follows is really the 'time-independent' one dimensional Schroedinger Equation. The time dependent part of the Schroedinger Equation, i.e., the part that tells you how the wave function varies with passing *time* rather than with location, has been factored out into a separate equation, which concerns us even less than the time-independent equation presented here. Rest assured, however, that the time-*dependent* factor of the full Schroedinger Equation is much simpler, mathematically, than the time-*independent* factor presented here, so you are getting your full (over)dose of the quantum theory.

(***11) Strictly speaking, this property holds if and only if the object is *bound*, i.e., restricted to move in a certain well-defined region in space, by the force represented by the potential energy function $V(x)$. An obvious example of such an object is the electron bound electromagnetically to the proton in a Hydrogen atom.

(***12) The notion of energy conservation came along surprisingly late, given

its central importance throughout physics. This is because, unlike momentum (also a conserved quantity), energy can come in many different forms – kinetic energy (of both translation and rotation), potential energy (of numerous different forms), heat, radiation, etc. The notion of energy conservation seems to have emerged simultaneously with the development of the modern theory of heat in the mid 1800's. Amongst its earliest proponents were a German physician by the name of Julius Robert Maier, and a British experimentalist named James Prescott Joule, whose professional roots lay in the business of brewing beer. As a further digression, it has been claimed by those in the thick of it that the development of quantum mechanics itself was greatly abetted by the consumption of the Danish label Carlsberg beer. Let this serve as a warning as to the potentially disruptive effects of alcoholic beverages.

Copyright 2001 by Bruce A. Schumm

CHAPTER 6

MATHEMATICAL PATTERNS

Lie Groups

Much has been said about the ever-growing and almost miraculous interconnectedness between the abstract field of pure mathematics and the real-world study of natural phenomena. That the former provides such an essential tool for the pursuit of the latter is a continual source of inspiration and wonder for those who reflect upon it. The ancient Greeks, whose developments in abstract mathematics retain a solid standing in the body of modern mathematical knowledge, and who also put considerable effort into the interpretation of the natural world, made only limited headway in the application of simple mathematical principles towards the description of nature. The connection is just not that obvious.

It was with the rise of the Western University system in the early renaissance that the connection between the abstract world of numbers and the infinitely complex and rich world of natural phenomena became firmly established. In the middle of the 14^{th} century, individuals at the Merton College of Oxford University, and the Universities of Paris and Bologna, developed the first rigorous descriptions of motion, including quantitative notions of speed and (uniform) acceleration. The development reached a pinnacle, of course, in Newton's 17^{th} century formulation of calculus, in concert with his application of calculus to the description of celestial motion.

Mathematics has hardly stood still since the time of Newton, and as its development has led in successively more arcane and abstract directions, its application to the natural world has become all that more remarkable. The specific example of group theory, and its application to the fundamental description of natural processes, provides an opportunity to convey a sense of the nature of this profound connection, and to understand how the musings of abstract mathematicians, sequestered away in their ivory-clad garrets, has provided an essential component of the modern understanding of natural phenomena.

In and of itself, group theory is one of the most gratifying topics in the wide world of abstract mathematics. As we'll see quite shortly, it's relatively easy to set forth the precise requirements that a set of objects must satisfy in order to

be deemed a *group*, in the strict mathematical sense. The delineation and study of these abstract mathematical entities has entertained a lot of top mathematical talent over the years, particularly in the latter part of the 19^{th} century, during which a much of the theory of finite groups (groups composed of a finite number of objects) was laid out, once and for all. Our interests, however, do not lie within the theory of finite groups, but rather with that of *infinite*, or more specifically, *continuous* groups. In fact, our interests are even more narrow than that, focussing on a specific class of continuous groups known as Lie Groups(***1), first defined and studied in the 1870's by the Norwegian mathematician Sophus Lie.

Lie groups lie at the heart of a surprisingly large number of descriptions of physical phenomena, and enjoy broad application throughout the fields of natural science. In particle physics, Lie groups play such a central role that it is impossible to proceed further without their introduction. To be perfectly honest, it wasn't until *after* Gell-Mann's introduction of the eight-fold way – a 'straightforward' application of Lie groups to the categorization of the particle zoo (see the previous chapter) that physicists were made to realize that they had begun to speak the language of group theory. Nowadays, though, the connection is firmly established, and physicists have benefited greatly from the abstract mathematicians' exhaustive treatment of Lie groups.

I. AN EXERCISE IN ABSTRACTION: MATHEMATICAL GROUPS

To a mathematician, a 'group' is any set of objects with an associated rule, or 'operation', which combines pairs of objects in the set. The obvious examples of operations, and in fact the ones from which the more general notion of 'operation' was abstracted, are addition and multiplication. For example, if x and y are two numbers (say, the monthly bills for the two separate phone lines in your house), then $z = x + y$ is the combination of x and y which tells you by exactly how much you are obligated to enrich the phone company each month. The operation of addition takes the pair of numbers x and y and combines them, yielding a third number, z.

The nice thing about addition and multiplication is that, without any external reference, most of us know how to take any given numbers x and y and combine them into the result z. To a mathematician, however, the group's operation could

be absolutely *any* rule which combines the objects in the set, as long as the operation satisfies four specific requirements (see below) – even if the only way to delineate that rule is to write out a complete table of pairs of numbers x and y and their combined result z. Generically, the process of operation is often represented by the symbol '$*$', i.e., the expression '$x * y$' denotes the combination of the objects x and y according to the rules of whatever operation you have chosen to associate with the set, of which multiplication and addition are just two of the many possibilities.

Not every set and operation on elements within the set comprise a group, however. In order to form a group, the set and operation must satisfy a set of four defining criteria, known as 'axioms'.

First, the set of objects must be 'closed' with regard to the operation. This is nothing more than a short-hand way to say that if x and y are objects in the group, then the result z of their combination (under the operation associated with the group) must also be an object in the group. For example, the set of whole numbers 0, 1, 2, 3, 4, ... are 'closed' under the operation of addition: the sum of any two whole numbers is always a whole number. Similarly, we could consider the set of whole numbers with the associated operation of multiplication; again, the set (whole numbers) exhibits closure under the chosen operation (multiplication). On the other hand, the set of whole numbers under the operation of division is not closed; for instance, one divided by four is 1/4 or 0.25, which is decidedly *not* a member of the set of whole numbers.

The second axiom, known as the 'associative law', is the most obscure of the axioms. It won't play much of a role in our discussion, but we need to include it for completeness' sake. Let's say that you have not two but *three* elements, a, b, and c, that you wish to combine according to the rules of the associated operation, yielding some final element z within the set. The obvious thing to do is to combine two of them first, say b and c, and then combine the result with a: $z = a * (b * c)$, where the parentheses indicate which of the two indicated '$*$' operations should be performed first. On the other hand, one could also imagine combining a and b first, and then combining the result of that operation with c: $z = (a * b) * c$. The requirement of the second axiom, that the 'associative law' hold, is nothing more than the requirement that both of these double-operations yield the same result z, for any possible a, b, and c within the set. One reason why the associative law is

relatively uninteresting is that it's hard to think of an operation that *doesn't* obey it – you can easily convince yourself that ordinary arithmetic operations such as addition and multiplication do. However, many of the rich and powerful results which form the mathematical theory of groups do require that the associative law hold true, and so it must be included as an axiom.

The thoughtful reader might recognize that there are even more possible ways to combine the three elements a, b, and c: what about $z = (b * c) * a$ or even $z = (c * a) * b$? These are all different combinations of the three elements a, b, and c – shouldn't they give the same result z as above? The answer is, emphatically, not necessarily. In these two operations, the *order* of the elements themselves (and not just the order of the operations) has been changed relative to those of the previous paragraph. The operation associated with the group is *not* required to be the same when the order of the elements being operated on is the switched. In other words, it is not necessarily true that $x * y = y * x$ – that x and y 'commute' – for all elements x and y in the set. Groups that exhibit this additional property, not necessarily required of groups, are known as 'commutative' or 'Abelian' groups. Most of the Lie groups that we will introduce towards the end of this chapter are *not* commutative, or 'non-Abelian' – a fact that will be easily demonstrated. This failure of the elements of the Lie groups to commute with each other will have substantial consequences when we discuss the physical application of Lie groups in the context of gauge theory in Chapter 8.

The third axiom requires that the set must possess an object which is an 'identity element' of the chosen operation. By this it is simply meant that the set possesses an element, call it 'I', which has the property that $I * x = x$ for *any* object x in the set. Note that for the whole numbers under addition the identity element is just '0', which when added to anything just gives the original thing back again. The whole numbers under the operation of multiplication also possess an identity element, which of course is just the number '1'.

The fourth and final axiom associated with the definition of a mathematical group, on the other hand, excludes the whole numbers, under either addition or multiplication, from forming a group. For the fourth axiom states that for each and every element x in the set, there must be one and only one element x', also in the set, for which $x * x' = I$, where again I is the identity element. For example, if the operation '$*$' represents multiplication, for which the identity element I is just

the number 1, this axiom requires that for each element x there be a corresponding element x' such that $x \times x' = 1$, or, equivalently, $x = 1/x'$. The element x' is known as the 'inverse' of x under the chosen operation. In these terms, the fourth axiom states that each element in the set must have a unique inverse under the chosen operation.

Thus, to be explicit, the whole numbers 0, 1 ,2 ,3 ,..., do *not* form a group when associated with the operation of multiplication. For example, what number x' has the property that $3 \times x' = 1$? Obviously, the answer is 1/3, and the number 1/3 is not a whole number.

To briefly recap, in case you've gotten lost in the fray: a group consists of a set with an associated operation, or rule of combination. The set and operation must satisfy four axioms in order that the whole system can properly be called a group: closure (the combination of two elements must always yield a third element within the set), associativeness ($(a*b)*c = a*(b*c)$), the possession of an identity element (which, when combined with any other element, gives the other element back again), and finally, for each element in the group, the existence within the set of a unique corresponding inverse (which, when combined with the element, yields the identity element). Additionally, if the group exhibits the property that $a*b = b*a$ for any elements a, b in the set, the group carries the special designation of being 'Abelian'.

It should be noted that there is nothing that can be 'proven' or 'disproved' about these four axioms. Concerning themselves only with the ethereal world of abstraction, mathematicians are free to arbitrarily introduce a structure known as a group, and are equally free to arbitrarily require that anything that is a group satisfy these four criteria. Nobody can say these axioms are right or wrong – they are simply the rules that mathematicians have chosen to require of something which they have decided for some reason or other to call a 'group'.

With this definition of the term 'group' out of the way, though, one can then develop a *theory* of groups, based on postulates which can be proven to follow directly from the four defining axioms. This theory is no more or less arbitrary than the axioms underlying the definition of the group – all that one can say is that since the arbitrary mathematical entities known as 'groups' satisfy (by fiat) all four of these criteria, then they will possess a number of other properties, and exhibit a number of other characteristics, which can be mathematically *proven* to follow

directly from the four axioms. Just for the sake of being definite, two questions (of many possible questions) that you might hope your theory of groups might answer are 'How many distinct Abelian groups (if any) contain exactly 675 elements?' or 'For the group with 3,698 elements, how many distinct subsets of elements in the group's set form groups in their own right under the associated operation?'. Things like that.

To the mathematically inclined, there is a deep beauty in the creation of the web of interlocking results that lead eventually to the solutions of basic questions one might ask about mathematical structures such as groups. It's sort of like doing a crossword puzzle – when you're finished, you haven't produced anything that will launch the next Fortune 500 company, but you have met and triumphed over an intellectual challenge.

There is a difference though – in mathematics, you have not triumphed over the arbitrary machinations of another human being (the designer of the puzzle), but rather over the absolute fabric of logical relations. The body of knowledge you have developed has the enviable characteristic of being demonstrably and absolutely true, given the set of assumptions (axioms) underlying your contemplations, irrespective of the foibles of your own human limitations – indeed, irrespective of the existence of humans themselves. And – as an added bonus – if it should so happen that the set of axioms on which your intellectual fortress is built is somehow relevant to the physical world, then you can sometimes even walk away with a deeper understanding of your natural surroundings.

The beauty of group theory, of course, is that its relevance to the worlds of both mathematics and natural science far exceeds the self-contained boundaries within which it was first developed.

II. A SPECIFIC EXAMPLE: THE GROUP MOD(4)

There's nothing that illustrates an abstraction better than a concrete example, so here's a description of a particular group – in fact one of the smallest possible groups, having only four elements in its set. This group is sometimes referred to by mathematicians as 'Mod(4)'.

The division of the terrestrial day into 24 hours is pretty much arbitrary – let's consider a culture that instead chooses to divide the day up into only four hours.

Also, let's assume that this culture is sufficiently relaxed that they don't care about the minutes within the hour. Thus, clocks in this culture will consist of dials with four markings – one through four – and a single hand which moves four times a day from one hour to the next. Figure 6.1 shows this clock.

The four markings (the numbers one through four) will form the elements of our set. The associated operation will be that of adding elapsed time (in hours) to the current time of day in order to get the time of day after the elapsed interval. Thus, if it's now one o'clock, and we want to know what time of day it will be after two more hours have elapsed, we consult the rules of our operation to deduce that $1 * 2 = 3$. The operation represented by the '$*$' in this case looks an awful lot like regular old addition – but it's not!

For example, what if the current time is 2 o'clock, and we want to know what time it will be after 3 more hours have elapsed. We consult our clock face to find that, after passing through four o'clock, the time cycles back to one o'clock. The day has changed, but that doesn't concern us, because the elements of the set are just the four hours representing the time of day; the date itself is of no relevance. So, after 3 more hours have elapsed, the time of day is one o'clock: $2 * 3 = 1$.

It's easy to see that the set of numbers is closed under this operation, which is sometimes referred to as 'clock arithmetic'. No matter how many hours you add to the current time, all you're going to do is spin the dial around to one of the four hours between one and four o'clock – the result of the operation of combining any two numbers between one and four in this way is just a third number between one and four. In addition, no matter what time of day you start with, if you add four elapsed hours to that time, you go through exactly one full day and get back to the original time of day: $x * 4 = x$, for any $x = 1, .., 4$ in the set. Comparing this expression to that of the definition of the identity element above, we see that the number 4 is thus just the set's identity element. Also, once you get the feel of clock arithmetic, it's easy to convince yourself that the associative law (the second axiom) holds.

That the fourth and final criterion (that every element in the set have a unique inverse within the set) is satisfied is demonstrated by considering the following three clock arithmetic operations: $1 * 3 = 4$, $2 * 2 = 4$, and $4 * 4 = 4$ (can you verify that these operations are correct according to the rules of clock arithmetic with four elements?). Recall that the inverse of an element is the element which, when

combined with the original element according to the rules of the operation, yields the identity element. So, we see that the elements 1,2,3,4 have the inverses 3,2,1,4, respectively. And that's it – the numbers one through four, under the operation of clock arithmetic, form a group!(***2)

It can't be emphasized enough that, to a mathematician, this particular group Mod(4), or any other group for that matter, is purely an abstraction. The symbols 1,2,3,4 are nothing more than labels for the elements of the group's set. In fact, the elements bearing these convenient labels could be, as far as the mathematician is concerned, absolutely anything – a pencil, a hammer, a bowl of spaghetti, and a glass eye, say, would work just fine. Then, in order to be the group Mod(4), the rule of combination would say that when you operate on a pencil and another pencil you get a hammer ($1 * 1 = 2$), on the pencil and the pasta you get an eyeball ($1 * 3 = 4$), and so forth.

As absurd as this seems, as long as you have the same number of elements as Mod(4), and the operation gives the exact same relations between all the elements as that of Mod(4), then the mathematician is happy to deem your odd assortment of objects and silly rule for combining them the group Mod(4). She might wonder why you went to all the trouble, but she would indeed be compelled to grant you the designation that you so desire. Even more than that, she would be exhibiting tremendous insight do so, for the ability to *abstract* – to see the deep connections between things which on the surface seem to have nothing in common – is perhaps the single most defining characteristic of the unique human intellect.(***3)

Thus, rather than the nature of the objects themselves, it's the interrelations between the objects in the group – the *pattern* of relationships between the elements of the group that is given by the associated operation – that establishes the unique mathematical identity of the group. It is through such patterns that the connection between the tangible world of particle physics and the ethereal world of abstract mathematics was first established in the decade of the 1960's.

III. INTO THE CONTINUUM: THE LIE GROUPS R(2) AND U(1)

Groups need not have finite numbers of elements. Consider the whole numbers from negative infinity to positive infinity (including zero) under the operation of addition. With a little thought, you can readily convince yourself that this forms a group with an infinite number of elements.

Now, between each consecutive pair of whole numbers, add in all the fractions which fall between them – there's an infinite number of fractions between each consecutive pair of whole numbers, and an infinite number of consecutive pairs of whole numbers. This infinitely infinite set of numbers, known as the 'rational numbers', also forms a group under the operation of addition (the sum of two fractions is always just another fraction).

If we add in 'transcendental' numbers – numbers whose decimal expansions (like that of $\pi = 3.14159...$) do not terminate or exhibit repeating patterns, we get the *real* numbers. Mathematicians have been able to show that, believe it or not, that there are an infinite number of transcendental numbers between each consecutive pair of *fractions* – which is a little hard to imagine, since two consecutive fractions would themselves seem to have to be infinitely close together (since there are an infinite number of them between each two whole numbers, such as zero and one). But it's demonstrably true.

Of course, this infinitely infinite set of 'real' numbers also forms a group under addition. In fact, real numbers have a special berth within natural science, in that the result of the measurement of any physical quantity is expected to be a member of the set of real numbers. The set of real numbers forms a 'continuum' of possible numbers – there are no 'gaps' between successive real numbers, no holes between which you could squeeze another number that might be the result of a physical measurement. The set of real numbers under the operation of addition is a *continuous* group.

Similarly, all Lie groups are continuous. In fact, the set of real numbers under addition is one of the most straightforward examples of a Lie group, although an example that is not particularly useful for illuminating the special properties of Lie groups. For instance, recall the claim from above that most Lie groups are *non-Abelian*. Obviously, the real numbers under addition *are* Abelian, since for any numbers x and y, $x + y = y + x$ (for this particular example of a group – the real numbers under the operation of everyday addition – we can substitute the '+' sign for the more general symbol '$*$').

Without yet saying exactly what a Lie group is, let's discuss another relatively simple example of a Lie group. This Lie group will also be Abelian – we'll get to an example of a non-Abelian Lie group in due course. Nevertheless, this particular Lie group will be of direct relevance, as it forms the basis of the 'gauge theory'

of electromagnetism – the most modern and up-to-date recasting of the quantum theory of electricity and magnetism – and is one of the two Lie groups underlying the unified Standard Model of electroweak interactions. In addition, it will be a good example of the true depth of the principle of mathematical abstraction.

Place a dot on a piece of paper with a pen (either in your imagination, or in actuality; which ever you prefer). Draw a closed shape of your liking on the paper somewhere within the vicinity of the dot (to avoid a possible confusion which we won't belabor, let's require that this shape not contain the dot). Now, hold the pen vertically and place the tip of the pen on the dot, so that you can rotate the paper around the dot. The dot will be the only point on the paper that doesn't move when you rotate the paper; the vertical line represented by the pen represents the 'axis' about which the rotation takes place, just as the wheels of a car rotate about the axis defined by the axle.

If you then go ahead and rotate the paper, the position of the shape you drew will change. Just how much the position changes depends on just how much you rotate the paper – if you rotate the paper minutely, then the position of the shape changes some, but very little. If you give the paper a real twist, the position of the shape will change substantially (unless, of course, the twist is just about 360 degrees, or a multiple of 360 degrees). In fact, the possible positions of the shape after the rotation form an infinite continuum – for every possible angle between 0 and 360 degrees (i.e., for every real number between 0 and 360) there is a unique position at which the shape comes to rest after the rotation. Note also that if you rotate the paper through some angle, and then again through some other angle, the result is just as if you rotated the paper once through some third angle, which is just the sum of the first and second angles.

This is indeed starting to smell a lot like a continuous group. However, to see that this system does indeed obey the postulates required of a mathematical group requires a somewhat sophisticated application of mathematical abstraction.

Recall, one more time, that a group consists of both a set *and* an a rule of combination, or *operation*, on that set. It's natural to think of an 'operation' as some sort of an action – in fact, the word 'operation' carries the direct denotation of action. So, if you're not extremely thoughtful, you'd be inclined to think of the act of rotation as the 'operation' in this system. However, this is *not* the case – if we want to discover how this system satisfies the requirements of a mathematical

group, we need to recognize that the act of rotation represents the *elements* of the group's set, and *not* the operation. The elements of the set are all the possible rotations of the paper about the pen's axis, and there's one for each possible angle by which you could rotate the paper. So, just as the possible angles between 0 and 360 degrees form a continuous set of numbers, the possible rotations of the paper form a continuous set of elements of this group.

The *operation* – the rule of combination – is just the successive application of two rotations. As mentioned above, a rotation by 32.4 degrees followed by a rotation of 19.1 degrees is the same thing as a single rotation of 51.5 degrees. So, if we represent with the symbol R_θ the element of the group's set corresponding to the counterclockwise rotation of the paper through an angle θ, then we can write

$$R_{19.1} * R_{32.4} = R_{51.5}$$

The result of the combination of the group elements corresponding to 32.4 and 19.1 degrees of rotation is a third element, also within the group, corresponding to a rotation of 51.5 degrees. Note that if the two rotations are large, so that the total rotation is somewhat greater than 360 degrees, you will go all the way around and wind up with a total rotation which is identical to that of one of the rotations less than 360 degrees – like going all the way around the face of the clock in the clock arithmetic of the group Mod(4).

So, once we make the leap of abstraction that recognizes the different possible rotations as the *elements* of the group's set, and recognizes the accumulation of the effect of two successive rotations as the *operation* associated with the group, we see that the axiom of closure (that the result of the operation must lie within the set) is satisfied. There's an easily identified identity element: just a rotation by 0 degrees ($R_0 * R_\theta = R_\theta$ for any angle of rotation θ since rotating first through θ degrees and then through 0 degrees leaves you with a total rotation of θ degrees). That each element has an inverse is also easy to see: a rotation through an angle θ, followed by a rotation through an angle of $360^\circ - \theta$, yields a combined rotation of 360°, which as we know is the same thing as a rotation of 0° – the identity element. Thus, the inverse of a rotation through the angle θ is just a rotation through an angle of $360^\circ - \theta$. Finally, I'll leave it to you (if interested) to show that the associative property (axiom two above) holds.

Since the surface of the paper you have been rotating (perhaps in your mind) is two-dimensional, this system, which we now know to be a formal mathematical group, is known as 'the group of rotations in two dimensions', or 'R(2)' for short. As we've seen, R(2) is a continuous group, which is one of the characteristics of Lie groups (what the other characteristics are we haven't yet discussed). But also, R(2) is *Abelian*, i.e., the order in which you perform the rotations (when combining them according to the rules of the operation) makes no difference whatsoever. A rotation through 32.4 degrees followed by a rotation through 19.1 degrees is the same as a rotation through 19.1 degrees followed by a rotation through 32.4 degrees – in both cases, what you wind up with is a rotation through 51.5 degrees. In fact, you may wonder, how can a group operation *not* be Abelian – how could the order in which you combine the elements possibly matter? The answer is coming soon, so hold that thought.

Before moving on to non-Abelian Lie groups, however, it is convenient here to first introduce another Abelian Lie group – the group U(1). The Lie Group U(1) is a group involving *complex*, or *imaginary* numbers, so let's pause for a minute to remember (or learn) what a complex number is.

What follows is a bit more technical than most of the rest of the book, and if you get lost, remember the bottom line: the group U(1) of rotations in one *complex* dimension is, mathematically, just the same thing as the group R(2) of rotations in two *real* (ordinary) dimensions. In both cases, elements of the group are specified by a single angle θ. In the context of R(2), this angle θ can be represented physically as the amount by which you rotate a piece of paper to achieve a given rotation R_θ. In the case of U(1), on the other hand, the angle can be represented physically as an amount by which you change the *phase* of a quantum-mechanical wave function. Recall our discussion of phase in Chapter 3: it describes where you are relative to the crest of a uniformly undulating wave. With every change of phase of 360°, you move exactly one wavelength forward. But in a uniformly undulating wave, all undulations are the same, so it's as if the 360° phase change brought you right back to where you started from, just like a 360° rotation of a piece of paper.

Given our eventual interest in the phase of the wavefunction, foreshadowed in Chapter 3, this particular physical conception of the group U(1) will be of particular relevance. Anyway, it's time for the details.

You probably recall that whenever you square a number, no matter what number

you start with, you end up with a positive result. This all boils down to the fact that two minuses, when multiplied together, make a positive, so if you square a negative number, the two minus signs combine to give a plus sign, and the result is positive. There's no way around this.

Well, no way unless you're a mathematician, and thus schooled in the discipline of making up whatever wacky thing you need to in order to satisfy your current whim.

In this spirit, let's *define* a number i, which is exactly that number, which we all agree shouldn't really exist, that when multiplied by itself (squared) gives the result -1: $i \cdot i = i^2 = -1$. Thus, if b is any *real* number at all, then the quantity $b \cdot i$ (the little dot is a more compact way of representing multiplication) has the property that its square is negative: $(b \cdot i) \times (b \cdot i) = (i \cdot i) \times (b \cdot b) = -1 \times b^2 = -b^2$. Numbers such as $b \cdot i$, which have *negative* squares, are known as 'imaginary' numbers – for obvious reasons. A 'complex' number is just a combination of a real number and an imaginary number. In other words, if z is a complex number, then we can always write $z = a + b \cdot i$, for some real numbers a and b. Note that if a is zero, the $z = b \cdot i$ is an imaginary number; if, on the other hand, b is zero, then $z = a$ is just a real number.

We all know how to order real numbers. Given any two real numbers, virtually anyone could say which one is bigger – which one they would prefer to receive, say, in dollars as a gift from a benevolent uncle. On the other hand, how do you discern the size of a *complex* number $z = a + b \cdot i$? Which is 'bigger': $100.2 + 22.6 \cdot i$, or $1.3 + 67 \cdot i$? By convention (of course, it turns out to be a very useful convention), we define the size, or 'modulus' $|z|$ of a complex number $z = a + b \cdot i$ to be just the square root of the sums of the squares of its real and imaginary pieces: $|z| = \sqrt{a^2 + b^2}$.

Thus, in particular, there is more than one complex number z that has a size of $|z| = 1$. You can easily verify on a calculator that the numbers $1 + 0 \cdot i$, $0.64 + 0.36 \cdot i$, and $\frac{12}{13} + \frac{5}{13} \cdot i$ all have a size of 1 according to the rule for calculating sizes. In fact, there is a *continuum* of complex numbers of size one, since for any real number a between -1 and $+1$, you can find a real number b (also between -1 and $+1$) such that $\sqrt{a^2 + b^2} = 1$, so that the number $z = a + b \cdot i$ has size one.

In fact, whenever you multiply together two complex numbers of size one, the resulting complex number also has size one. Thus, it turns out that the set of

complex numbers of size one, together with the operation of multiplication(***4), forms a continuous group. This group is known is U(1).

To get a better feel for the group U(1), let's recall for a moment the demonstration involving the rotating sheet of paper from the text just above in which the two-dimensional rotation group R(2) was introduced. If we were to place a one inch long (or one centimeter long, if you prefer) arrow so that the base of the arrow rests right at the dot representing the axis of rotation, the tip of the arrow could lie at one of an infinitude – of a *continuum* – of points, depending on which direction the arrow happens to point. In fact, the set of possible points shown by the inch-long arrow to be one inch from the dot on the paper are determined by nothing more than the set of possible angles – between 0 and 360 degrees – which pin down the direction the arrow points in its trajectory away from the dot around which we rotate the paper. Since the elements of the group U(1) are just complex numbers of size one, they are thus similarly represented by the points at the tip of the arrows with angles from 0 to 360 degrees, as exhibited in Figure 6.2(***5). For those who remember some high-school level geometry, it's just like a circle on a two-dimensional plot, except instead of having an x and a y axis, we have a 'real' and an 'imaginary' axis.

Thus, the elements of the group U(1), the group of complex numbers of size one, are equivalent in every way to the possible angles θ which define the rotations that are the elements of the two-dimensional rotation group R(2). Furthermore, although we haven't demonstrated it, it turns out that the multiplication of two complex numbers of size one is completely equivalent to the operation of combining rotations for R(2). If a size-one complex number z_1 is represented in Figure 6.2 by an arrow at an angle of 32.4 degrees, and z_2 by an arrow at an angle of 19.1 degrees, then the product $z_2 \cdot z_1$ is represented by an arrow at an angle of $32.4 + 19.1 = 51.5$ degrees – just as following a 32.4 degree rotation with a 19.1 degree rotation is just the same as a single 51.5 degree rotation. The bottom line: R(2), the group of rotations in two dimensions, and U(1), group of complex numbers of size one, are one and the same group.

At first blush, R(2) and U(1) look very different, but after some deep thought, we see that they are one and the same thing. The fact that the two groups are based on very different number systems (two real vs. one complex dimension), and employ entirely different operations (successive rotation vs. complex-number

multiplication) is immaterial. Even less material is the fact that they happen to have physical manifestations (the description of rotation in two dimensions vs. the possible changes in phase of a quantum-mechanical wavefunction) which are vastly different in kind.

Instead, the mathematician sees that both R(2) and U(1) contain the same number of elements (the infinitude of real numbers between 0 and 360 degrees) and that the *pattern* of interrelations between the elements of each group, as specified by their associated operations, is identical. That's all it takes – to the mathematician, and even thus the physicist, the two groups are one and the same. This is an archetypical example of the workings of mathematical abstraction – and a very necessary step in the process of unveiling the deep connection between these mathematical entities and the underlying structure of the physical universe.

And final, a word about the notation. The rotations which comprise the group R(2) were performed with a sheet of paper – a plane – for which it takes two numbers (usually denoted x and y) to specify the location of any point on the paper relative to the fixed dot about which we rotate. In the case of the complex group U(1), however, we saw that we admit the possibility of something equivalent to rotation with just a *single* complex number – a *single* complex dimension. Thus, we can interpret the nomenclature which led mathematicians to the designation U(1) for this complex group. The 'U' refers to the word 'unit' – the elements are all the complex numbers of 'unit' size, or as we have been saying, 'size one'. The specifier '(1)' refers to the fact that the 'rotations' which carry size-one complex numbers from one to another are rotations in a *single* complex dimension.

IV. ADDING THE NEXT DIMENSION: THE LIE GROUP R(3)

By making the transition, at this point, from two to three (real) dimensions, not only do we add another spatial direction to our mental constructions, we also incorporate another dimension of richness in the properties of the associated Lie groups.

The three-dimensional version of the group R(2) of rotations of a plane (represented, say, by a piece of paper) is of course just the group of rotations of a three-dimensional space (represented, say, by a cardboard box). As you might have already guessed, the three-dimensional group of rotations is known as R(3).

The extra possibilities opened up by the expansion from two to three dimensions allows the group R(3) to exhibit two critical properties which are not exhibited by its two-dimensional counterpart R(2).

The first of these is that the group R(3), as heavily foreshadowed, is non-abelian: the order in which you combine the elements of the group matters. The second is that we can think of the infinite continuum of elements which comprise R(3) as being 'generated' by a finite number of prototypical elements of the group. This latter characteristic is the essence of what makes any continuous group specifically a Lie group. When we associate Lie groups with the underlying nature of the physical forces in Chapter 8, the specific number of such 'generators' (R(3), for example, has 3), and the fact that the order of combination matters, will have direct and profound consequences on the nature of the associated force, and thus the behavior of the physical universe at its most fundamental level.

To explore the properties of the group R(3), the set of possible rotations in three-dimensional space, find a cardboard box or a thick book or some other similarly shaped object, and mark one of the eight corners with a pen (even if you got by just fine without the prop in the discussion of R(2), it might be helpful to actually go through the physical exercise in this case). The marked corner will form our 'origin' – the fixed point about which we will perform our rotations, similar to the dot on the paper of the previous discussion. The three edges which connect at the chosen origin define three lines in three dimensional space. Label them 'x', 'y', and 'z'. Many of you will recognize these as the x, y, and z axes of a three-dimensional 'cartesian' coordinate system.

Tape a piece of paper on a table, and set the box squarely on top of it, in a way such that the marked corner (the 'origin') is about in the middle of the piece of paper. Mark the place on the paper at which the origin rests.

Now, practice rotating the box exclusively about the x axis – i.e., so that the edge you've labeled 'x' remains in the same place as you turn the book back and forth. Rotate the box back to its original position, and then do the same exercise about your y and then your z axes (edges), again making sure you return the box to its original position before going on from the y axis to the z axis rotation. This is all just warm-up; the only thing you might want to notice is that, by the time you've gotten through all three of these exercises (the x, y, and z axis rotations), there will be one and only one point on the box that hasn't moved the entire time

– the corner you marked as your 'origin'. Now, however, we're ready to begin our exploration of the group R(3). In what follows, we'll refer to these three types of rotations – by some unspecified angle, but with one of the three axes (x, y, or z) fixed, as one of the three 'exercise' rotations.

Place the box squarely in front of you, as before, with the origin of the box resting on the dot on the paper. Make a mental note of the position of the box. Now, pick up the box, and in the air above the table, spin the book around in some arbitrary way so that it's orientation is space is random. Now, preserving this orientation, lower the box back to the table so that the origin again touches the dot on the piece of paper. (Because the surface of the table is rigid and won't let any of the box pass through it, not all possible orientations of the box will work, but there are enough orientations that do work that this doesn't really present a problem. Just pick one of the random orientations that does work).

Now, compare this position of the box to the original position, of which you made a mental note. The two positions are related by an arbitrary rotation of the three dimensional box about its marked corner – the fixed 'origin' of the coordinate system. The set of all such possible rotations, including the ones you couldn't do because the table was in the way, form the elements of the group R(3). We can label these different elements of the rotation group R(3) by the angles at which the x, y, and z axes end up after the rotation.

Again, it must be emphasized that these actions – the rotations of the axes – are the *elements* of R(3), and *not* the associated operation. The operation, of course, is just the *combination* of two rotations by their successive application.

Note that the 'exercise' rotations exclusively – the rotations exclusively around the x, y, or z axis, by any angle between 0 and 360 degrees, lie within this set of elements. Note also that amongst these *elements* resides the 'trivial' possibility of no rotation whatsoever (just leaving the book sitting there unmoved); we still want to think of this as a rotation, albeit one that leaves the position of the box unchanged. This 'rotation' of the box by nothing is just the identity element of the group R(3), just as it was for the rotating paper of R(2).

It's fairly straightforward to convince yourself that the set of rotations R(3), in concert with the operation given by the successive application of rotations, forms a group. If you rotate the box, and then rotate again, what you always wind up with is just some other rotation. So, R(3) is 'closed' under this operation. The

identity element was introduced above; the inverse of an arbitrary rotation is just the rotation that 'undoes' the arbitrary rotation by returning the box to its original position. The associative law holds, although again we won't bother to show it.

There's an interesting way to specify the different rotations which form the elements of the group R(3), which gets to the heart of its designation as a Lie group. If the arbitrary rotation that you just did was general enough, you will see that it is *not* possible to move the book from its original (resting) position to its final (held) position with just a single 'exercise' rotation about either the x, y, or z axis. However, any arbitrary rotation *can* be achieved by a succession of three 'exercise' rotations of the appropriate angle – one about the x axis, another about the new y axis (note that after the rotation about the x axis, the orientation of the y axis has changed, giving a 'new' orientation to the y axis), and a third about the even newer z axis. If you play around with the box a bit, you can probably convince yourself that this is in fact the case. Thus, instead of distinguishing between the different possible rotations by designating the angles at which the former x, y, and z axes end up, we can instead simply designate the angles $\theta_x, \theta_y, \theta_z$ associated with the three exercise rotations necessary to produce the arbitrary rotation in question.

This is it – the essence of what makes a group a Lie group. The group contains a continuum of elements – an infinite, 'dense' set of elements, yet its structure is delineated by a finite number of elements, known as 'generators', from which the continuous infinite of elements are easily obtained. In the case of R(3), there are three such generators, which are just the small 'exercise' rotations about the x, y, and z axes. Once you know what the generators of the Lie group are, all you need to do is just figure out how much each generator contributes in order to form any given element in the group (any given rotation in three dimensions). In the case of the Lie group R(3), this amounts to just picking the values (between 0 and 360 degrees) of the three angles $\theta_x, \theta_y, \theta_z$ that produce the rotation that you desire. Absolutely any element of R(3) – any rotation in a three dimensional space – can be produced in this way.

Note that in our discussion of R(2), the group of possible rotations of a flat piece of paper, or a 'plane' (a two dimensional space), we made no mention of generators. This is not because the group R(2) (which *is* a Lie group) has no generators. Recall that, for the the group R(2), the elements (rotations) are specified by the *single* angle θ through which we rotated the piece of paper. Thus, if you think about

it, R(2) is a Lie group generated by a single basic 'exercise' rotation: just a small rotation about single available axis of rotation in the system (this axis was just the line which went perpendicularly through the dot on the paper). Thus, R(2) and U(1), which we argued are the same abstract Lie group, are Lie groups characterized by a single generator.

Thus, to recapitulate, Lie groups are 'continuous' groups (composed of an infinitely dense succession of elements) whose elements are derivable from a finite number of 'generators'. The properties of the *generators* alone – their number (typically relatively small; only three for R(3)) and their relations under the group's operation – completely establish the characteristics of the infinite continuum of elements in the full group.

Our discussion of the general properties of Lie groups is not complete, of course, because as of yet we have neglected to discuss the issue of the behavior of the generators under the group's operation, i.e., what happens when you start trying to combine the generators according to the rules of the group's operation. The critical property to consider, it turns out, is whether the generators 'commute', which you may recall is just a fancy way of asking whether or not the order of the elements in the operation makes a difference or not. If the order *does* make a difference (we'll see that this is the case for R(3) in just a moment), then the elements don't 'commute', and the group, you may recall, is given the designation 'non-abelian'. For non-abelian Lie groups, the difference between the result of combining pairs of *generators* in different orders is very well defined. In fact, this order-difference in the combination of the Lie group's generators is another very important factor (in addition to the *number* of generators) determining the characteristics of the given Lie group.

To see that the generators of R(3) *don't* commute – that the order of the generators (different 'exercise' rotations) matters when you combine them according to the group's operation (which is just the successive application of the two exercise rotations in question), once again place your labeled box squarely in front of you. Let's let the symbol 'X_{90}' denote an 'exercise' rotation of 90 degrees about the x axis (which should still be labeled as such on your box). Similarly, the symbols 'Y_{90}' and 'Z_{90}' denote rotations of 90 degrees about the y and z axes, respectively.

Now, let's consider the behavior of two of these generators under the group's operation(***6). The combination, or successive application, of the 90-degree exercise

rotations about the x and y axis can be performed in two different orders:

$$R_{xy} = X_{90} * Y_{90}$$

$$R_{yx} = Y_{90} * X_{90}$$

We'll adhere to the rather confusing mathematical convention that the rotations are performed *from right to left*, so for example, the first of these expressions directs you to rotate by 90 degrees about the y axis first, and then by 90 degrees about the x axis (which will have changed position, but will still be labeled with the 'x' on your box).

Try this, and see if you can confirm what is depicted in Figures 6.3a and 6.3b – the two different orders of combination result in combined rotations R_{xy} and R_{yx} which are completely different! The order in which the elements are combined together by the group's operation (successive rotation) *does* matter – the generators of R(3), and thus more generally the elements of R(3) which you construct by taking different amounts of these fundamental 'generating' rotations, *do not commute*. The group R(3) is *non-abelian*!

To make this point more clear, it's helpful to consider a counterexample: a Lie group with three generators in which the generators *do* commute. Let's go back to the sheet of paper rotating about a point – the prop we used to motivate the group R(2) – and consider now the case where we have three independent sheets of paper stacked one on top of another, all lined up and ready to rotate about the same axis of rotation, as shown in Figure 6.4.

Rotations in this system are again determined by three angles – in this case $\theta_L, \theta_M, \theta_U$, the angles by which the lower, middle, and upper sheets of paper are rotated. The three 'generators' of these rotations, from which all possible rotations of the system can be made, are just small rotations of each of the three sheets of paper (lower, middle, and upper) individually about the common axis of rotation. It's easy to see that in this case the generators *do* commute – it obviously makes no difference which order you apply the separate rotations of the three sheets of paper to compose any given rotation in the group.

As we'll discuss in Chapter 8, a physical theory based on this *abelian* Lie group (which is known as $R(2) \otimes R(2) \otimes R(2)$ since it's just a pasting together of three

copies of the two-dimensional rotation group R(2)) would have significantly different characteristics than one based on the *non-abelian* group R(3), even though both of these groups have the same number (three) of generators.

V. THE LIE GROUP SU(2)

Somewhat above, we dwelt for some time on the Lie group U(1) – the (abelian) Lie group of complex numbers of size one. Recall that a complex number z is any number of the form $z = a + b \cdot i$, where i is the 'imaginary' square root of -1, i.e., i is the number defined by the relation $i = \sqrt{-1}$, or equivalently, $i \cdot i = -1$. The quantities a and b are just any ordinary, everyday ('real') numbers. The 'size' $|z|$ of the complex number z is just given by the expression $|z| = \sqrt{a^2 + b^2}$. We stated without showing it(***7) that whenever you multiply two size-one complex numbers together, you end up with another size-one complex numbers. So, size-one complex numbers, under the operation of multiplication, form a group – the group U(1). We also argued that, considered as an abstraction, the group U(1) of size-one complex numbers under multiplication, and the group R(2) of rotations in two dimensions, are one and the same. In other words, considering only the number of elements, and the elements' pattern of interrelations as given by the group's operation, and *not* the identity of the elements themselves, R(2) and U(1) are identical.

Now, without offering any motivation (ample motivation will be provided in the next chapter, in which we begin to discuss the physical application of Lie groups), let's consider the following Lie group: that of rotations in two *complex* dimensions.

To understand (or really, to define) what it means to 'rotate' in two complex dimensions, let's first recall some aspects of rotation in two normal, everyday dimensions. Let's go back to the prop we used to introduce R(2), the group of rotations in two normal (real) dimensions. This prop was simply a sheet of paper with a dot on it. We placed the point of a pencil on the dot; the pencil then became the axis about which we rotated the paper. We saw that rotations in two (real) dimensions were described by a single generator, and so the value of a single angle (representing the amount by which the paper was to be rotated) was sufficient to specify which of the infinite number of possible rotations we were interested in performing.

On this sheet of paper, draw (or imagine drawing, if you prefer) a one-inch long arrow whose base is at the axis of rotation (the dot about which the paper is rotated), and whose head is anywhere on the page. Now, rotate the page, and make note of the fact that, while the position of the arrow changes (i.e., it would take different coordinates x and y to describe the location of the head of the arrow), its *size* does not change – no matter how you rotate the page, the arrow is still one inch long. It would be different if the page were made of some elastic material that you could stretch or squeeze – in this case, the arrow would change its size. But deformation (stretching and squeezing) is different than rotation. Rotation – pure rotation, without deformation – does not change the size of an object – it's 'size-preserving'.

This, then, is the key to the notion of 'rotation' in two *complex* dimensions – we simply ask that the size of two-dimensional complex objects (whatever that means!) be preserved by the complex rotation(***8). The set of rotations in two complex dimensions is just the set (mathematical) motions which preserve the size of objects which have their form in two complex dimensions. Of course, we can't picture such objects, but the mathematical rules for representing such objects (and for figuring out what happens to these representations when they are rotated through two complex dimensions) are fairly easily developed.

Just as the set of rotations in two normal dimensions forms a Lie group (R(2)), the set of rotations in two complex dimensions also forms a Lie group, known as 'SU(2)' (read 'ess-you-two'). Again, the numeral '2' refers to the fact that two dimensions (in this case, complex dimensions) are at play. The letter 'U' stands for the word 'unitary', which just means 'size-preserving', and which we just argued is the bellwether property of a rotation. The letter 'S' stands for the word 'special', and represents the fact that not all two-dimensional complex rotations are interesting – we only want the interesting, or 'special', rotations(***9). This latter point is a technicality which is not a concern of ours.

Recall that the operation under which the set R(2) became a group was the successive application of two rotations, the result of which is a third, combined rotation which is always a member of R(2). Likewise, the result of the successive application of two complex two-dimensional rotations is always just some other complex two-dimensional rotation. This operation satisfies all the other requirements listed at the beginning of this chapter, and so SU(2), the set of 'rotations'

(size-preserving operations) in two complex dimensions is in fact a group – a Lie group.

If SU(2) is in fact a Lie group, then its infinitude of elements must all derive from a small number of basic elements, or generators. Recall that for R(2), there was but a single generator – a small rotation about the axis through the dot and perpendicular to the sheet of paper. Any rotation in the group R(2) – any amount of turning of the sheet of paper about that single axis – can be 'generated' by the application of this basic rotation about the single axis; one simply has to chose an appropriately sized angle (from the continuous infinitude of possible angles) for the rotation.

Similarly, we saw that the group R(3) of rotations in three dimensions has three generators – small rotations about each of the x, y, and z axes, or, in terms of the prop we used, the three small 'exercise' rotations about each of the three different edges of a box which emanate from one of its corners. Any rotation in R(3) – any possible final orientation of the three box edges after the execution of any set of turns about the fixed box corner – can be achieved by three rotations of the appropriate size about the three box edges, or axes. These three different types of rotations – one for each axis or box edge – generate the group R(3).

Furthermore, these rotations do not 'commute' – the order in which you perform any two successive rotations does makes a difference, leading in general to a different final orientation of the box (or coordinate system) after both rotations have been performed in succession. The precise way in which the rotations fail to commute – the difference between the successive application of two of the generating rotations in first one order and then the other – is a characteristic property of the Lie group. The list of these properties – the order-differences in the successive application of each pair of generating rotations – is known as the 'algebra' of the Lie group(***10). It is this algebra which differentiates between different Lie groups with the same number of generators, such as the two three-generator Lie groups R(3) and R(2)$\otimes$R(2)$\otimes$R(2).

So then, what about SU(2) – how many generating rotations do we need in order to be able to achieve any two-dimensional *complex* rotation with the appropriate choice of the angle of each of the generating rotations?

If you have a good picture of normal two-dimensional rotations (R(2)) in mind, you might be inclined to say that just one generator is required. After all, as we've

mentioned a number of times, there's only one axis of rotation in two dimensions. However, these are *complex* dimensions, and so our intuition may not be reliable – we are completely at the mercy of our mathematical powers of abstraction and induction. In fact, recall from the description of the group U(1) above that even in *one* complex dimension, there's a 'rotation' we can do – the reapportionment of the real and imaginary parts of the complex number $a + b \cdot i$ so that the size $\sqrt{a^2 + b^2}$ of the new complex number is unchanged from the old.

So, we need one generator to specify the desired amount of rotation between the two separate complex dimensions of SU(2), equivalent to the single generator we needed in order to specify the desired amount of rotation between the two real dimensions of R(2). But, in addition, we need a generator to specify exactly how to position our object relative to the real and imaginary axes *within* each of the two separate complex dimensions. This adds two more generators, and so, just like R(3), SU(2) has *three* generators.

But we now know enough not to stop here, for if we really want to compare SU(2) and R(3), we need to know more than just the number of required generators. We also need to compare the behavior of the combination (under the group's operation) of pairs of generators when we change the order of their combination – we need to compare the *algebra* of the two groups.

Once you know how to represent the elements of SU(2) and R(3) mathematically(***11), this is quickly done, since there are only three unique ways to pair two out of three generators. When you do perform these three calculations, first for pairs of generators of SU(2), and then for pairs of generators for R(3), you find something rather surprising: the algebra of the two groups is precisely the same! In other words, the commutation, or ordering, difference between the first and second generators of SU(2) is precisely the same as that of the first and second generators of R(3), and so forth, for each of the three possible pairs of generators for each group.

So, at this point a warm feeling washes over you as you conclude, just as you did in the case of the single-generator abelian groups U(1) and R(2), that the three-generator non-abelian groups SU(2) and R(3) are nothing more than different manifestations of the same Lie group. Again, you exclaim, the powers of abstract reasoning have triumphed over the pedestrian world of the merely apparent, revealing SU(2) and R(3) for what they really are: the same individual, simply

dressed in slightly different clothing. In essence – mathematically – they are one and the same.

This epiphany, as it turns out, is almost – but not quite – correct.

Let's go back to our intuitively accessible R(3), and the box prop that we used to evince its properties. Consider any one of the three generating 'exercise' rotations of the box, and rotate the box through 360 degrees of that rotation. You get back to where you started. A rotation of 360 degrees, or a multiple thereof, about any axis, is the same as doing nothing – in other words, is just the group's identity element. Such is not the case for the group SU(2). For SU(2), it takes *720 degrees* – twice 360 degrees – of (complex) rotation in order to get back to where you started! To foreshadow the next chapter just a bit, this will turn out to lead to a natural association of the Lie group SU(2) with the 'spin', or intrinsic angular motion, of a particle rotating on its own axis. The Lie group R(3) will have a natural association with 'orbital' angular momentum – the energy of the angular motion of a particle locked in orbit about a fixed center.

Requiring a rotation of 720 rather than 360 degrees to restore a system to its original orientation may sound a bit strange, but there's actually a relatively simple physical system which has such a property. And here it is...

Place the palm of one hand near the cheek on the same side of your body, and face the palm straight upward (you'll need to be standing for this to work). Place a small box or book on your palm so that it rests comfortably with no other support than the palm itself. Making sure the top of the box always faces upwards, rotate your forearm about your elbow so that your palm (and the box) passes through 360 degrees of rotation about your elbow. If you've done it right, you'll be in a fairly comfortable position, with the top of the box still facing up, but now *below* rather than *above* your elbow. The system is in a different configuration than it was before the 360 degree rotation! A continuation of the rotation by another 360 degrees, for a total of 720 degrees, restores the system to its original configuration, with the book resting on your palm right near your cheek, above the elbow. So, even though this system has little to do with SU(2), we see that it's actually not that out of the ordinary for a system to require 720 rather than 360 degrees of rotation in order to get back to its original orientation.

And so, the mathematician hedges. She would say that, 'locally', SU(2) and

R(3) are the same group – the number of generators, and the 'algebra', or ordering-differences between the small 'generating' rotations, are the same for both groups. But she would say that 'globally', there are some differences – it takes 720, and not 360, degrees of rotation to get back to where you started with SU(2), while for R(3), of course, it only takes 360.

Physically, when we begin to apply Lie groups to the structure of the natural world (next chapter!), we'll see that the local ('Lie algebra') properties of the group establish the properties of the physical quantity associated with the Lie group, and so, for example, SU(2) and R(3) are both generally able to describe the behavior of systems with angular (rotational) motion. However, the global properties of the group establish what specific types of systems can be described. Thus R(3), with its mere 360 degrees of rotation, is more limited, and can only describe systems with orbital motion. SU(2), with its 720 degrees of possible rotation, is thus more general, and can describe systems with either (or both) orbital or spinning motion.

VI. LIE GROUP WRAP-UP

Mathematicians make it their business, of course, to study and delineate properties of Lie groups. Naturally, one of the central questions in the theory of Lie groups is simply the determination of what the possible Lie groups are – what abstract sets and associated operations, or equivalently, what collections of generators and their associated 'Lie algebras' – lead to systems which satisfy the requirements of mathematical group-hood which were introduced above. Mathematicians hardly stop there, though, for in their fertile and abstract imagination, they envision all sorts of interesting properties of Lie groups.

A particularly fascinating aspect of the study of Lie groups is the exploration of their 'topological' features – topology being the field of modern mathematics in which the very notions of shape and connectedness (e.g., how many 'holes' or 'handles' an object has) are distilled and applied to abstract mathematical systems. The issue of the topology of Lie groups, a topic which saw considerable advances in the middle part of the twentieth century, is now playing a central role in theoretical exploration at the very cutting edge of formal mathematical physics. The practitioners of this most advanced brand of mathematical science hold considerable hope that their efforts may lead to the what is no less than the holy grail of particle physics – a self-consistent 'theory of everything', in which all the known

workings (interactions) of nature are at last understood to be different manifestations of a single, overarching proto-force of great physical (if not mathematical) simplicity(***12). This whole subject, of course, lies well beyond the scope of our more earth-bound narrative.

For our purposes, though, we can extricate ourselves from the abstractions and complexities of this chapter after a final, yet brief, discussion of a few more specific examples of Lie groups. We've discussed the two- and three-dimensional rotation groups R(2) and R(3) in terms of their (relatively) easily grasped actions on the orientation of physical objects in real, everyday space. We tenuously extended these notions to one- and two-dimensional complex spaces with the introduction of U(1) and SU(2), and discussed their mathematical relation to R(2) and R(3), respectively.

With the properties of these groups in hand, the mathematician has enough intellectual ammunition available to induce the properties of higher-dimensional rotation groups, both real and complex. Although not that central of a point in our discussion, it turns out that the connection between the complex and real (everyday) rotation groups – the fact that R(2) and U(1) are one and the same mathematically, and that R(3) and SU(2) are very closely related – is fortuitous, and is not a general property connecting the worlds of real and complex rotation groups. The next-higher-dimensional real rotation group R(4) (the group of rotations in four real dimensions) has nine generators (can you see why)? On the other hand, the next-higher-dimensional complex rotation group SU(3) (the group of rotations in three complex dimensions) has only eight generators (don't bother trying to see why – the only way it can be done is by working it out mathematically).

But who cares, really. The point is that we know, using the tools of mathematics, exactly what the properties of these higher dimensional rotation groups (a small subset of the set of possible Lie groups) are. We can deduce their set of generators, work out the 'algebra' of these generators, and even derive, if interested, the groups' global 'topological' properties. And, armed with this, we can fully explore the implications of physical theories based on each of these groups (what we mean by the statement 'physical theory based on these groups' will be the subject of the next two chapters). And again, to foreshadow a bit, what we've found so far is that the groups which provide successful physical theories are U(1) and SU(2), together providing the basis for the theory of the 'electro-weak' force, and SU(3), which

provides the basis for the theory of the strong nuclear force – the force, you may recall, that is responsible for binding quarks together into protons and neutrons. So, we'll need to make substantial use of U(1), SU(2), and SU(3) in the chapters that lie ahead.

That's it for our introduction of Lie groups. Take a break, get yourself a doughnut and coffee (or organic rice cake and sparkling water), stretch your legs a bit, and we'll meet again in Chapter 7.

****

(***1) Here, 'Lie' is pronounced the same as 'Lee'.

(***2) Here's an interesting challenge: can you find another distinct operation (rule for combining the numbers 1 through 4) that still satisfies all the axioms required of a group? If so, then you've discovered a *second*, distinct group with four elements.

(***3) Whether or not the superior intellectual capabilities of the human species renders it somehow intrinsically more deserving or worthwhile than all the other animals is a debate into which this author would prefer not to enter, being a cat 'owner' and not wishing to offend the true masters of the household.

(***4) The rule for multiplying together two complex numbers $z_1 = a + bcdoti$ and $z_2 = x + y \cdot i$ is what you might expect if you know a little algebra: $z_1 \times z_2 = a \cdot x + a \cdot y \cdot i + b \cdot x \cdot i + b \cdot y \cdot i \cdot i = (a \cdot x - b \cdot y) + (a \cdot y + b \cdot x) \cdot i$, since $i \cdot i = i^2 = -1$.

(***5) Strictly speaking, in this 'complex plane', the condition $\sqrt{a^2 + b^2} = 1$ is just the equation of a circle of radius one. A segment forming the radius of this circle reaches every point on the circle as it sweeps through an angle of 360 degrees, and so each point on the circle can be labeled by an angle between 0 and 360 degrees.

(***6) Here, we're being a bit fast and loose about what specific elements of R(3) are actually the generators. The three generators are, strictly speaking, 'exercise' rotations (about the three axes x, y, and z) by *infinitesimal* angles – angles *theta* which are vanishingly small but still not quite zero. Such a concept will really only make sense to those familiar with calculus, which most of us are not. Feel free here to think of the generators as the three 90° exercise rotations.

(***7) It's actually relatively easy to show that this is true if you have a good innate sense of mathematical reasoning. Just take two size-one complex numbers $a + b \cdot i$ and $c + d \cdot i$ and multiply them together according to the rules of algebra, remembering that $i \cdot i = -1$. See if you can show that what you wind up with is size-one! (Remember that the complex number $a + b \cdot i$ is size-one provided that $a^2 + b^2 = 1$.)

(***8) In two ordinary dimensions, the size, or length, of an arrow whose base is at the origin and whose head is at the coordinates x and y is given by $s = \sqrt{x^2 + y^2}$. Now, consider two *complex* numbers $x = a_x + b_x \cdot i$ and $y = a_y + b_y \cdot i$. Similarly, we define the 'size' of a 'complex arrow' with its base at the origin and its head at the (complex) coordinates x and y to be $s = \sqrt{|x|^2 + |y|^2}$ (recall that the size $|x|$ of the complex number $x = a_x + b_x \cdot i$ is itself given by $|x| = \sqrt{a_x^2 + b_x^2}$). Rotations in two complex dimensions change the complex coordinates x and y of the head of the arrow in a way such that the complex arrow's 'size' s is unchanged.

(***9) Recall from our discussion in Chapter 3 that the overall *phase* of a wave is not physically relevant. The 'special' complex rotations are the rotations which leave this overall phase unchanged. A 'non-special' rotation which does the same thing as a given 'special' rotation (except that it also changes the overall 'complex phase') has the same quantum mechanical content as the special rotation, and so need not be considered – we need only consider the 'special' rotations to take into account all of the physically relevant possibilities.

(***10) A bit of further explanation may make this point more clear. Let x and y be two generating rotations of the Lie group R(3). Or, more generally, let x and y be two generating elements of a general Lie group. Let $\alpha = y * x$ represent the combination of y with x under the operation of the group – in the case of R(3), the successive application of rotation x followed by rotation y. Then, let $\beta = x * y$ represent the combination in the other order – the combination of x with y, or the successive application of rotation y followed by rotation x. Both of these combinations α and β must be members of the group; for example, in the case of the Lie group R(3), α and β are just some other three-dimensional rotations. *If* the group is non-abelian, i,e, if its elements do not 'commute', then α and β will be *different* elements – different three-dimensional rotations. Furthermore, the *difference* between α and β will itself just be another element, say 'γ' in the group (the difference between any two rotations α and β is just some other rotation

γ). Thus, we can write, for any two generators a and b of the Lie group, that $y * x - x * y = \gamma$. For each possible pair of generators x and y, we will have a different result γ for the ordering difference $y * x - x * y$. The list of γ's, for all possible pairs of generators x and y, specifies the 'algebra' of the Lie group. So, as discussed, the 'algebra' of the Lie group is just a precise specification of the way in which the generators of the Lie group *fail* to commute. Note that is the Lie group is abelian, then all the generators commute, and $y*x = x*y$, or $y*x-x*y = 0$ for all pairs of generators x and y. For an abelian group, then, all the ordering-difference γ's are exactly 0. On the other hand, then, for a non-abelian group at least one of the γ's is not 0.

(***11) For the mathematically inclined: this is most easily done with two-by-two complex matrices for SU(2) and three-by-three real matrices for R(3); the group operation is then simply matrix multiplication.

(***12) Such efforts are not completely divorced from the world of demonstrable experimental fact. For example, for a number of years, candidate theories growing out of this effort seemed to require that the very strength of the electromagnetic force vary substantially over the 10 billion year or so evolution of the universe from the big bang to modern times. What experimental constraints, you might ask, might address the issue of the strength of the electromagnetic force many, many years ago? Well, it turns out that the relative abundances of various nuclear species (elements and isotopes) after a nuclear explosion is extremely sensitive to the strength of the electromagnetic force. It also just so happens that about a billion years ago, in what is now Africa, tectonic forces conspired to push some Uranium 235 into high enough concentration that a chain reaction – a spontaneous atomic bomb explosion, if you will – ensued. Geological study of the elemental and isotopic abundances in the remnants of this spontaneous explosion show unequivocally that the strength of the electromagnetic force has remained essentially unchanged over the last billion years, in direct contradiction to the predictions of these advanced theories. In most recent years, however, this problem has been overcome, and some promising approaches have emerged towards constructing a unified theory for which force strengths are fixed over time.