CHAPTER 5

The Greatest Forms of the Beautiful

[Particle Physics and Electroweak Unification]

The mathematical sciences particularly exhibit order,
symmetry and limitation; and these are the greatest
forms of the beautiful. —Martin H. Fischer

June 1981
Trieste, Italy

My darlings, Fatima and Hassan,

This letter is only for the two of you and, as promised, it is a "proper" letter, addressed not to the children who sat on my knee but to the young adults you are now becoming. I think back to those days (which doubtless seem long ago to you, but are only yesterday to me) when you would leaf through my papers, fascinated by the fact that I could "read" the funny, squiggly alphabet of equations. I remember how eager you were to learn to draw those symbols, how you traced their foreign shapes on page after page . . . almost like some rite of initiation. "What does this mean, Abba?" you would ask, and I would have to fend you off with incomplete answers. I have waited a long time to share with you the thoughts that occupy my days, so this letter marks an occasion in my life as much as in yours.

It is the lunch hour, but I have skipped the long queues in the cafeteria and crossed through the wall from the ICTP* into Miramare Park. A deep calm pervades the wooded groves, and in the

---

*International Centre for Theoretical Physics.

clearing a picture-perfect fairy-tale castle hovers on a cliff over the shimmering Adriatic Sea. On my walks here, I have stumbled on picturesque little fountains, burst on spectacular views, and in quiet places where the curtain of leaves parts, I have even glimpsed the elusive deer who call this park home. I imagine the two of you running around, uncovering these half-hidden pleasures, and the thought makes me smile. My first few days in Trieste, each time I saw a graceful statue, a beautiful building, or a lovely vista, I would turn around instinctively, looking for you. This city is replete with sights I want to share with you and your mother. I miss you all so much, it makes my heart ache. If God wills, one day the four of us will visit Trieste together.

There is one particular bench that I have come to think of as my own. From here, I can see the immensity of soothing water that stretches across to the horizon and also the hills that cup around Trieste like the palm of a gentle hand. Sitting here, far from the clamor of mindless chores and institutional responsibilities, I find my perspective on physics broadening; perhaps perspectives are elastic and stretch to fill the space provided. In this wide expanse I spread my thoughts: I lay out exciting new ideas, poured into my head during the seminars, and pull out the crumpled thoughts I had pushed to the back of my mind. Then I let all these ideas drift like clouds across my mind's sky while I sit back and watch, marveling at the pictures they form. Some of these shapes I will attempt to point out to you now. But, as I told you when you were little children, fluffy cotton-wool clouds vanish into wisps and vapor when you get close enough, and so, even while you strain to see the pictures I show you, keep in mind that with time and distance they may well change.

To begin: I told you years ago that I am a particle physicist. The particles I study are the building blocks of nature, much like the pieces in your Lego set. With these fundamental particles, as with Lego bricks, we must know their possible combinations and various attributes before we can use them to construct anything.

In the early part of this century, physicists thought our list of essential pieces and attributes was complete. Every phenomenon that had yet been observed could be ascribed to one of two interactions—gravity or electromagnetism. Three subatomic particles—negatively charged electrons, neutral neutrons, and positively charged protons—were thought to be the indivisible constituents of matter. The atom was built by placing electrons in orbit around a nucleus, consisting of neutrons and protons packed together.

This is how the atom is pictured in your schoolbooks, but if you approach it with a questioning mind—as you should get in the habit of doing—you will instantly realize that this simple model is in flagrant violation of the laws of electromagnetism. The recognition, and resolution, of these contradictions led to the inclusion of a few more particles and two additional interactions on our list. Since these new forces made their presence felt only when we started poking around at atomic scales, they were called the nuclear forces.

One obvious problem with the simple atomic model had already been resolved by quantum mechanics. Despite their attraction to the nucleus, electrons were kept at bay because they were constrained to move in orbits, and there was a smallest orbit beyond which they could never go. This stopped the atom from imploding, but what kept it from exploding? The very formation of the nucleus, in fact, went against the dictates of electromagnetism. By law, a group of protons should experience intense electrostatic repulsion, and yet somehow they were held tightly together. There was obviously a very strong binding agent at work, overpowering the protons' urge to fly apart; it was named the strong (nuclear) force.

Even with this addition, the acrobatics of the subatomic particles could not be fully explained. One particularly troubling move, beta decay, involved the transformation of a neutron into a proton and electron. The most obvious explanation was put forth: perhaps the neutron was a composite object made up of a proton and electron; beta decay would then simply be the dissolution of the bond that held these two constituents together. But, after careful consideration, this possibility was ruled out, and we were left confronting the perplexing display of a neutron spontaneously morphing into a proton and an electron. The three forces we knew of pushed and pulled and otherwise affected the motion of an object, but they did not change its identity. It became obvious that an entirely new kind of interaction was responsible for beta decay—the so-called weak force.

The number of fundamental forces had now risen to four, but the problems were far from over. In complete disregard for norms and tradition, beta decay appeared to flout the law of energy conservation. Experiments revealed that the combined energies of the electron and proton after decay were less than that of the original neutron. This was nothing short of a catastrophe. In an increasingly uncertain world, one of the few tenets physicists held was the law of energy conservation. Regardless of the nature of the interactions involved, the total energy of a system before and after any physical process was always the same. Energy could not be created or destroyed; it could only change form.

It seemed as if the axioms that had provided warmth and shelter were falling to pieces, and we were cast into cold, unbounded ignorance. At times like this, when gaping holes appear in a system of thought, there is only one thing to do. We must take the pieces apart and reassemble them to create a structure that encompasses the successes of old but leaves room for explanations yet to come.

In a way, it is similar to assembling a jigsaw puzzle, except that you have no image of what the completed picture should look like. Our dining table has been strewn with puzzle pieces enough times for you to know that the first thing to do with a new puzzle is to assemble the frame. Corner pieces are the most valuable finds, because they provide clues about two different sides and the way in which they meet. Once the frame is in place, it becomes easier to work in from the edges, or even to link pieces in random disconnected vignettes, because you know that somehow they will fit inside a circumscribed boundary; it may seem paradoxical, but the containment is liberating.

Laws are to a theory as the frame is to a puzzle. Once we determine the laws by which a theory abides, we have established the borders within which all our explanations must be contained. We move freely within this frame while remaining aware of the lines that must not be crossed. Energy conservation had always formed a crucial part of every frame, but faced with the conundrum of beta decay, some desperate physicists considered crossing this boundary. Wolfgang Pauli was unwilling to take this sacrilegious step, so he chose the only other option. Based on blind faith in the law, he claimed that energy *was* in fact conserved; the missing energy had merely adopted a form we could not detect. Pauli claimed that the neutron decayed into an electron, a proton, and an invisible something extra. He claimed that if the energy carried off by this phantom was accounted for, the total energy before and after the decay would be the same—just as the law of energy conservation decreed.

There were many who refused to believe in the existence of a smooth criminal who left no fingerprints behind; they tried, instead, to come to terms with the fact that energy had actually been destroyed. But Pauli held his ground. In true Sherlock Holmes fashion, he started listing the attributes the thief must have in order to execute this perfect crime. He found that the deed could be done by a particle that had no mass,* no charge, and carried a spin of ½. He called this the neutrino. Since most particle detectors work on the principle that charged particles leave visible tracks, it was easy to argue that an electrically neutral, massless particle would escape detection.

---

*Subsequent research has revealed that the neutrino may perhaps possess a small, but nevertheless finite, mass.

Two decades after Pauli placed this ‘Wanted’ ad, the neutrino was finally captured, and experiment vindicated the bold trust Pauli had placed in theoretical reasoning. This changed forever the way physics was done. The neutrino was the first particle whose existence was inferred in this manner, but it was far from being the last.

Even after it had been found, the neutrino continued to shock us with its maverick behavior; it turned out to be chiral. Before you can feel the horror that chilled the physics community at this discovery, you will need to know what chirality is.

Think of the electron as a little ball that rotates about an axis through its center. This axis can be oriented either up or down—corresponding to “spin up” and “spin down”—so you can picture it as a little arrow that points in the appropriate direction. (Keep in mind that these are just visual metaphors, the electron is not in fact a little ball, the quantum mechanical property called spin is not a literal manifestation of the English word, and these arrows are purely imaginary!)

In order to specify the motion of an electron, we have to state the orientation of the arrow (spin) as well as the direction in which the electron rotates. This combination of the two attributes is referred to as helicity or handedness.

To understand how helicities are assigned, try this: Point the thumb of your right hand up and curl your fingers around. You will see they turn in a counterclockwise direction. By analogy, a right-handed electron is one that rotates counterclockwise around the vertical axis when the arrow points up. By flipping your thumb, you can see that a right-handed electron will rotate clockwise when the arrow points downward. In similar vein, a left-handed electron rotates clockwise when the arrow points up and counterclockwise when down.

Notice that the two are mirror images of each other. If we place a right-handed electron (clockwise rotation, spin down) in front of a mirror, its reflection will rotate counterclockwise while spin remains unchanged: the mirror electron is thus left-handed. This interchange of left and right is in accord with expectations and familiar from everyday experience: if you raise your right hand in front of a mirror, your reflection raises its left. But other than this left-right exchange, we expect the mirror world to be identical to our own. In other words, we expect the laws of physics to be unchanged upon reflection; we expect nonchirality.* The other alternative is a chiral theory, in which the world and its mirror image have the freedom to act independently, and no one took that option seriously. Almost implicitly, we assumed that every sensible theory must be nonchiral. Imagine the horror physicists felt when they found that the neutrino was shamelessly chiral—it simply had no reflection! All neutrinos are left-handed. They lack a right-handed counterpart† and, hence, a mirror image.

I have tried to proceed systematically, but I have introduced so many new ideas that it would be quite natural if, in your zeal to understand each line, you have by now lost sight of the larger picture. Before I go any further, let me reiterate: Our study of atomic structure revealed discrepancies in existing theories; gravity and electromagnetism had to supplemented by the two nuclear forces—the strong force, which kept the atom stable; and the weak force, which caused particles to transmute and was discovered to be the mechanism behind the radioactive decay of nuclei. Our investigation of beta decay resulted in the inclusion of a new particle in the subatomic roster: electrons, protons, and neutrons were joined by the notoriously hard-to-pin-down neutrino, which also turned out to be chiral. This was such a re-

---

**Chiral* comes from the Greek word for "hand." A nonchiral theory is "neutral"; it does not prefer any one chirality to the other. A chiral theory is biased in favor of a particular choice.

†In recent years, there has been some evidence that neutrinos may not in fact be completely massless, and consequently right-handed neutrinos may exist. Even if this turns out to be true, the left- and right-handed neutrinos will still not be perfectly symmetric, so the theory will continue to be chiral.

markable departure from expectation that it took a while, and much evidence, for the physics community to digest the information and accept it as fact.

But there was still more to come. The discovery of the neutrino was very welcome because it exonerated our theoretical structure, but it was considered to be a special case. We did not anticipate, nor did we want, the hailstorm of subatomic particles that soon pelted down on us.

You might be wondering where all these "new" particles were coming from, and why we hadn't seen them before. The answer is that most of these particles are unstable. They exist only for brief flashes of time, before nature redistributes their attributes into more efficient packages—and these stable parcels are what we see in the world around us. With the advent of particle accelerators and colliders, we were able to spy on the subatomic world for the first time. From Einstein, we know that mass is just a store of energy. If that seems too abstract a notion, think of particles as vessels. Each particle has its own characteristic vessel, of a particular size and shape; the amount of energy the vessel can hold determines the mass of the particle.

In these large, complicated machines, particles are accelerated to very high speeds, so that they acquire even more energy, and then they are made to collide. The vessels break open, and molten energy rushes out, like a genie escaping its bottle. This unleashed energy does not stay free for long; some of it adopts a less constricted form, like heat or light or motion, but most flows into a waiting set of empty containers—thereby taking on a new particle incarnation. Initially, energy tends to flow into fewer, larger vessels; but these are unwieldy, and it is instantly poured out again, distributed among the small, stable vessels of familiar particles. The intermediate step happens so quickly that we were unaware of it until advances in technology made it possible to see those transitory particles whose presence we had never suspected.

For a while it was exciting to find new particles, but the novelty quickly faded. Over the next couple of decades, the tale of discovery was told so many times that it lost its charm. Jokes were made* about awarding the Nobel Prize to the physicist who actually managed to avoid the discovery of a new particle. Indifferent to their reception, particles continued to rain down from the skies, and physicists scrambled to accommodate them somehow.

But even scientists have their limits, and when the muon was discovered in 1936, it was almost an affront. There was no reason for the muon to exist. For all intents and purposes, it was identical to the electron, just two hundred times heavier. Isidor Rabi (who later won the Nobel Prize) expressed his exasperation: "Who ordered that?" he asked.

One can easily sympathize with the plight of the physicists at the time. If particles decided to replicate themselves in ever heavier versions, there would be no end to the madness. There seemed to be no logic anymore behind the appearance of the new particles. People began to wonder which of these were truly fundamental. Were all of them entirely new, or were some of them combinations of the old and familiar?

While strolling through town a few days ago, I noticed ropes strung at various places along the roads. I remembered having seeing similar ropes lining the paths in Miramare Park, and the repetition piqued my curiosity. Upon enquiry, I found that they were there as protection against the Bora, a wild, wolfish wind that howls across Trieste en route to the gulf. As it squeezes through the gaps in the surrounding mountains, the bitterly cold Bora becomes icicle sharp. On its prowl across the city, this wind rips the tiles off roofs, sends parked motorcycles flying in the air, and blows people off their feet—hence the ropes. Some-

---

*Oppenheimer is credited as the author of this joke, but I have not found the original statement anywhere, only indirect reports.

thing similar was happening in particle physics a few decades ago. A razor-sharp wind tore through the edifices of our theories, causing entire floors to cave in. New particles were blown in, and they flew every which way with reckless abandon, knocking people off kilter.

The Bora is rather mercurial; it blows in fits and starts. Even Triestinos accustomed to this erratic behavior are sometimes tricked into thinking the wind has run its course, and they move away from the ropes, only to find the Bora lying in wait. It was so in particle physics as well. Every time there was a brief lull from the storm and people moved in to survey the damage, the temperamental wind threw another tantrum. When it finally blew over, all that remained was a broken, unrecognizable mess. As physicists rummaged around in this windswept chaos, organizing and classifying particles, they noticed certain patterns emerging. No one understood where these correlations came from, but that was immaterial. Patterns in themselves are powerful and persuasive. Wherever there is repetition, there is the hope that reality can be peeled back one layer further, that an underlying structure exists on which the visible form is built.

In the end, it was the American physicist Murray Gell-Mann whose tabulation made this structure transparent. Gell-Mann found that a set of six elementary particles, which he whimsically named "quarks," could be combined in various ways to form the proton, the neutron, and most of the other, newer subatomic particles. This discovery also exposed the true nature of the nuclear forces: the weak force changed one type of quark into another, whereas the strong force was the glue binding quarks together. The strong force shackled triplets of quarks to form neutrons and protons, and it caused the quarks in neighboring neutrons and protons to attract each other with such power that the stability of the nucleus was guaranteed.

Not all particles could be built from quarks: the electron was found to be an elementary particle in itself, as were its heavier cousin the muon, a yet heavier version called the tau, and the neutrinos associated with all three. Together, the electron, muon, tau, and their respective neutrinos were referred to as leptons. Being free of quarks, the leptons were indifferent to the strong force, though they were still subject to the weak.

At the end of all this reorganization, the tally for "matter particles" had risen from three (electron, proton, neutron) to twelve (six leptons and six quarks), and the fundamental forces had gone from two (gravity and electromagnetism) to four (the strong and weak nuclear forces included). Not a huge improvement as far as numbers go, but not bad either—particularly in the wake of the recent particle explosion. With these amendments to our list of ingredients for the universe, we could once again find recipes for everything we had ever encountered, whether directly through our senses or indirectly through instruments like telescopes and particle detectors.

A lucky glance at my watch reminds me that the lunch hour is over and the conference is about to resume. I must get back to ICTP now. I will finish this conversation with you tonight.

Long after I put my pen down, I was composing this letter in my mind. Throughout the day, I kept searching for the right words, an expressive turn of phrase, an image, a metaphor—any vehicle that would convey my ideas to you. When I finally stepped out of the lecture hall this evening, the mental toil of the day had left me exhausted. Like a homing pigeon, I headed to the bus station at Piazza Oberdan.

Already, in the short time I have been here, Bus 42 has become my favorite place to relax. For an unhurried hour and a half, it traces a path out from the heart of Trieste through ever quieter streets to Opicina, a picturesque little town next to the border. As we curve along the gently winding roads, breathtaking vistas roll out steadily. We ride past the huge stone monuments of Monrupino, past the Grotta Gigante cave whose depths glimmer with crystal, past the lighthouse Faro della Vittoria where Winged Victory stands, keeping watch over Trieste.

Most of the passengers are locals who chat away happily on their ride home, the lilting cadences of their speech forming the perfect accompaniment to the lush scenery. Picturesque houses cling to tree-lined hills, bathed in rainbow hues by the setting sun as the boundaries between sea and sky melt in an otherworldly glow. In this setting, the rapturous notes and decadent melodies of opera seem inevitable; the lavish beauty of Italy calls these sounds into being.

Passengers alight every so often, and still I stay on, until we reach the top of the hill where the entire view lies before our eyes, like a luxurious silken tapestry unfurled. This being the end of the route, the driver parks the bus. He looks at me questioningly, but when I nod no, he smiles and heads out for a quick break. About ten minutes later, we begin our descent back to town.

It is dusk now, and the hills are dappled in pinpricks of light; houses, shops, and villages are represented by constellations of tiny dots. Instinctively I start looking for a structure, a way to group them together and make sense of them. How dense would the lights have to be to constitute a village? What might cause those outlying patterns? Are there any specific shapes I can identify that would help me recognize a specific grouping again? While my mind idly plays these games, the particulars fall away and the larger theme I have been unconsciously grappling with all day shines into view.

The mathematical meaning of symmetry is not far removed from the meaning ascribed to this word in common parlance. If I ask you whether or not a given object is symmetric, you will most likely have an immediate answer. Even without conscious consideration, you respond instinctively to geometric similarities between two aspects of the object—if such similarities exist. You might need to focus a little harder to identify and describe the symmetry, but a first glance suffices to recognize that a repetition of some kind is at play.

More often than not, symmetry is pleasing to the eye, and as a result it is heavily employed in the ornamentation and design of traditional objects. The aesthetic value ascribed to symmetry has been a constant through centuries of changes in taste and fashion. Despite all that divides the two of you from the great Mughal emperors, you have this in common with them: the intricate motifs that tessellate the arches and ceilings of the Badshahi Mosque take your breath away; you are fascinated by the geometric fretwork of the marble lattices in the Lahore Fort; and you are delighted when you come across a familiar pattern, repeated on a different scale, in an alien context or on an unrelated structure.

Being an ardent admirer of Mughal architecture, I have often thought about why I find these designs so appealing, and I think it is this: patterns are reassuring. They promise the return of the familiar, and so they hold us safe. Yet patterns have no natural end. Iterations can go on forever, and in that way the known carries us into the Unknown, offering a hint of the divine. Perhaps this is why a devout believer like Abdus Salam* was drawn to symmetry like a moth to a flame.

Inherent in the word "symmetry" is the implication that some operation can be performed on a system without affecting the outcome. A mirror-symmetric shape, for instance, can be flipped over and both the object and its mirror reflection will appear unchanged. An equilateral triangle can be rotated about the center, and as long as its three vertices fall on the same three points, even the most careful observer (as long as he has not seen you perform the rotation) will never know that anything happened. Similarly, a square can be rotated by a right angle—or two, or three, or four—without anyone ever knowing the difference.† The ultimate symmetry, of course, is that of a circle: it

---

* Abdus Salam was a Nobel laureate and the founding director of the ICTP.

† In general, a regular *n*-sided polygon can be rotated through any of the angles that result when a circle is cut into *n* equal slices, and none of these manipulations will leave behind a trace.

can secretly be rotated through any angle at all without appearances giving it away.

When we say an object is symmetric, we specify—or imply—the action that leaves it unchanged. The collection of all such actions comprises the symmetry group of the object. The symmetry group of an equilateral triangle thus has three members, that of a square has four, and more generally speaking, the symmetry group of the *n*-sided polygon has *n* members. These groups all have a finite number of members, reflecting the discrete nature of the corresponding symmetries. Such groups are interesting and useful, and they generate very pretty patterns, but for our present purposes it is *continuous* symmetries that are far more interesting, for these underlie gauge theories, such as quantum electromagnetism and the theories of the strong and weak nuclear forces.

The gauge theory formalism is immensely powerful because it not only predicts the action of a force, but goes deeper to explain why a force has the strength and range it does, and also why it mediates between particular particles; it does all this by drawing on arguments of symmetry. A continuous symmetry, as the name implies, is one in which the transformation (rotation, in the above examples) can be smoothly varied and is not restricted to discrete values. A simple example is the group of rotations of a circle. The angle can be varied uninterruptedly, and stopped at any point; throughout the process, the circle is left unchanged.

Spatial symmetries are familiar to us and hence are easy to recognize, but they are far from being the only kind; there also exist "internal"* symmetries—so called because they become manifest only in an abstract space, not in the visible three dimensions. These symmetries show up in a theory in the guise

---

*Such is the case for the gauge theories of electromagnetism, as well as the strong and weak nuclear forces.

of mathematical transformations that leave the equations unchanged.

As a basic example, consider the case where a quantity x figures in the equations strictly through its square x^2—there is no expression that contains x alone. Since $(-x)^2 = x^2$, we conclude that the theory is blind to the difference between x and $-x$. In other words, it is symmetric under the transformation of x into $-x$, and the theory is said to have a (discrete) symmetry. No part of this argument depends on the physical quantity represented by x; it could be anything at all. If, instead, the equations have such a form that they remain unchanged as x cycles through a whole range of connected values, the theory is said to have a continuous symmetry.

In order to give rise to a gauge theory, the underlying symmetries must be not only continuous but also local rather than global. What are global and local transformations, you ask? If the same transformation is performed throughout a system, it is said to be global; if the transformation can be varied across different points, it is said to be local.

As an example, think of daylight saving time. One fine day, an entire country wakes up and decides to set its clocks an hour forward. What was, until yesterday, 7 a.m. instantly becomes 8 a.m., but life goes on undisturbed and people can make and keep appointments the same as always, because the decision is a global one. Regardless of whether we call it 7 a.m. or 8 a.m., as long as the entire country has set its clocks forward by the required hour, everyone agrees on the time, which is all that is needed for the system of life to maintain an unchanged appearance.

A local transformation is far more interesting, because it has the potential to be far more chaotic. Suppose the people rise in protest, arguing for the right to determine their own individual times. They see no reason why the government, or indeed any agency, should tell them how to set their personal clocks. Eventually the political powers tire of arguing and give in. Since all citizens are free to set their clocks as they choose,* some precautionary measures clearly need to be taken in order to avoid widespread confusion.

Before we can decide what these measures should be, think for a moment about all that would remain unchanged. Irrespective of how we set our clocks, the duration of events would not be affected. School would still run for six hours, regardless of the fact that one child might say it lasted from 9 p.m. till 3 a.m., while his classmate would claim it started at 4:34 a.m. and ran until 10:34 a.m. People would be required to put in as many hours at the office as they did before; it would take the same amount of time to commute from one place to another or to cook a meal, and the same number of days would elapse between successive birthdays. In short, everyone would agree on how long something took, just not on the time at which it started or ended.

In the absence of a universal time, people would need to share their personal times with each other whenever they made appointments. If my dentist asked me to come in at 5:32 a.m., I would have no idea what she meant unless I asked what her current personal time was, and then tallied that against my own. This exchange of information would allow me to convert her time into mine, so that I could arrive at her office when she expected me. Such determinations of time are continuous (assuming the townspeople's watches can be varied in the infinity of subdivisions between seconds) and local (in that they change freely from person to person or place to place); thus they constitute gauge transformations.

It now becomes obvious that if a system (like a city) is to be symmetric (remain apparently unchanged) with respect to a

---

*Time, as it is viewed on a conventional clock, isn't quite a continuous symmetry, since traditional clocks show time in discrete seconds; think of this as an advanced civilization where time can be measured to infinite precision. People are no longer limited to saying "it is 9:45"; they can as easily declare it to be "8:35 and 49.234234984938 seconds."

gauge transformation—that is, while a particular parameter undergoes continuous changes that vary from place to place (like the determination of a personal time)—a transfer of information must take place (there must be a way to compare the time displayed on two distinct clocks). The mechanism for the exchange of information is what we call a force. The presence of this force is what makes the symmetry possible.

Physicists put it the other way around. Symmetry under gauge transformations gives rise to a force, they say. The very fact that system is invariant under a particular set of gauge transformations *implies* that an interaction exists to make this symmetry possible; gauge theories, therefore, come with interactions built-in. So, it is only natural to expect that the fundamental forces of nature should be described by gauge theories. And, indeed, they are.*

We have spoken of symmetries and transformations, discrete and continuous, local and global, but I have not yet said what the word *gauge* refers to. It is a tricky matter, knowing when to probe into the genealogy of words. Often names are just names, like quarks for instance—that word makes no sense. At other times, it is downright treacherous to read too much into names; the Up quark isn't any more vertically upright than the Down quark, and the Strange and Charm quarks aren't stranger or more charming than the Top or Bottom quarks. However, it just so happens that the use of the word *gauge* is somewhat revealing, and since gauge theories are an abstract concept, quite difficult to wrap one's head around, I thought I would tell you where the name comes from.

---

*It is somewhat of a moot point, almost an issue of semantics, whether or not general relativity is a gauge theory. The other three forces (electromagnetism, and the strong and weak force) can all be formulated as Yang-Mills theories, which means the gauge group satisfies a particular set of criteria (is a compact, semi-simple Lie group). Gravity stands apart because, even though it is a gauge theory in a broader sense of the word, it is definitely not of the Yang-Mills type.

In English, the word *gauge* refers to a measuring device, and this meaning carries over into physics. A gauge theory is one in which the means of measurement does not need to be universally agreed upon, but rather can be independently determined at each point in space. In other words, there is, everywhere, the freedom to "pick a gauge." The key point, of course, is that the physical world is blind to this choice; no matter how arbitrarily the measuring system is chosen from point to point, all appearances remain unchanged. A gauge theory contains concealed freedoms; there are a host of invisible choices we can make, without altering reality. But this liberty comes at the cost of introducing forces into the theory.

This intimate intertwining of forces and symmetries has spectacular implications; some of them are too technical for me to even attempt to share with you, but there is one which, given my current setting, is too tempting to resist. I turn to the tale of electroweak unification—the seminal work for which Abdus Salam, together with the American physicists Sheldon Glashow and Steven Weinberg, was awarded the Nobel Prize just two years ago.

The Nobel Prize is more than an award for excellence; it is a nod of acceptance by the physics community and signals the inclusion of a theory in the standard core of knowledge. The electroweak theory was subjected to some very rigorous testing before this prize was awarded. To my mind, at least, there is another argument in its favor: the fact that Glashow, Salam, and Weinberg, with their markedly different personalities, beliefs, and motivations, assembled the theory while working, for the most part, independently of each other.

Salam's motivations were almost religious. He credits his culture and heritage with having inculcated in him an appreciation for symmetry, a belief in unity, and the faith that a single underlying cause ties together disparate phenomena. Weinberg, on the other hand, considers religion "an insult to human dignity." In my mind, this dichotomy only renders their theory that much more beautiful. If two people with diverging beliefs can arrive at the same structure, surely it cannot be just a subjective construction of someone's mind; it must be a reflection of some objective reality that exists "out there."

Glashow, Salam, and Weinberg showed that the two apparently unrelated symmetries underlying electromagnetism and the weak nuclear force could in fact be traced back to a single, larger symmetry that had been "broken."

A quick comparison between electromagnetism and the weak force reveals the magnitude of this discovery. The contrast between the two forces couldn't be any starker! Electromagnetism is felt across large distances, but the weak force dies out at nuclear scales; electromagnetism is not chiral, but the weak force is; when subjected to electromagnetic interactions, particles retain their identity, but the effects of the weak force are alchemical and can transform one particle into another. Yet Glashow, Salam, and Weinberg managed to stitch these overtly different entities into one. Not since Maxwell amalgamated electricity and magnetism had physics experienced such a triumphant confluence.

Of course, there were many subtleties involved. The beauty and strength of gauge theories come from their heritage of symmetry, and this underlying structure had to be preserved. The key was to realize that symmetry was demanded only from the equations, and not from their solutions.

The idea that electromagnetism and the weak force might be related dates back to Julian Schwinger. Sheldon Glashow, his student, tried to make the connection, and a continent away, so did Abdus Salam and John Ward. Their theories were uncannily parallel, and both came up against the same problem: the weak force had to be mediated by massive bosons, but there seemed to be no way to endow force carriers with mass without blowing up the theory, so the work was put aside. A few years later, Abdus Salam and Steven Weinberg heard something that fired their imaginations. Drawing on recent developments in the theory of superconductivity, scientists had begun to question one of their most basic assumptions—the emptiness of the vacuum.

The vacuum isn't necessarily a void, they said, but merely the ground state, the lowest energy configuration—the baseline on which everything else is built, and from which nothing further can be extracted. It is possible, then, for the vacuum to be permeated by a field. Using this insight, Salam and Weinberg each arrived at the same solution: that bosons acquire mass through interactions with a new field that could be included in the particle physics repertoire without upsetting the symmetry of the equations. This Higgs field, as it came to be called, represented neither matter nor force carriers, but instead an entirely different—and previously unobserved—kind of particle known as a scalar. When the modified equations were solved, it was found that all the fields vanished in the vacuum, except the Higgs. The usual example given is that of a marble and a Mexican hat. Assume that the marble represents the state of the universe, the hat represents energy, and the value of the Higgs field corresponds to the distance of the marble from the center of the hat.

When the marble is balanced perfectly, right in the center of the hat's hump, the Higgs field vanishes, but the system has a finite energy, given by the height of the hump. The energy is lowered if the marble falls, and the vacuum corresponds to the lowest part of the brim (before the edges turn up again). When the marble falls to this position, the value of the Higgs field (the distance from the center of the tip) becomes equal to the radius of the hump. So, at one time, either the Higgs field or the energy can be minimized, but not both simultaneously. Poised on the hump of the sombrero, the marble is surrounded with infinite possibilities, each as good as the next, but its position is precarious and practically impossible to maintain. The energy of the system isn't minimized until the marble rolls down the hump, breaking this impasse. The direction in which the marble falls is completely random; the point on the rim where it lands is not distinguished in any way until, by virtue of the marble's arrival, it becomes the chosen vacuum, the point of reference for everything that happens from then on.

The details of electroweak theory are subtle, intricate, and almost breathtaking, and it can only really be done justice using mathematics. But a flavor of the argument is conveyed by the analogy Abdus Salam used in later years: "Look at ice and water," he writes. "They can co-exist at zero degrees centigrade, although they are very distinct with different properties. However, if you increase the temperature you find that they represent the same fundamental reality, the same fluid. Similarly, we thought that if you could conceive of a Universe, which was very, very hot . . . the weak nuclear force would exhibit the same long-range character as the electromagnetic force. You would then see the unification of these two forces perfectly clearly." Since we experience these forces only as they are now and not as they used to be in millennia past, we see electromagnetism as being inherently different from the weak force, whereas in fact the two can be traced back to the same root.

Once electromagnetism and the weak force had been subsumed into a single formalism, it was only natural to ask whether the strong nuclear force could also be incorporated. Abdus Salam says this is like asking if, in our analogy of ice and water, we can include steam. Several scientists began to wonder if the three quantum field theories could be encompassed in what came to be called a grand unified theory.* According to grand unification, the early universe was like an ideal and endless water bath where the temperature was exactly the same everywhere. Consequently, there was no means of distinguishing any one point from another, and every direction was exactly the same, from here on out to infinity. While these intensely high energies reigned and absolute symmetry prevailed, no struc-

---

*One such scheme was proposed by Salam and Jogesh Pati, another by Howard Georgi in collaboration with Glashow, and later, with Steven Weinberg and Helen Quinn.

ture was possible. As the universe expanded and cooled down, asymmetries developed naturally; the figurative water turned to ice. With its crystalline structure, ice distinguishes among different directions, whereas for water (at least under the idealized hypothetical circumstances we are considering) all directions are completely equivalent and interchangeable. Water is the more symmetric state, but ice has more structure: this appears to be a common theme.

It is by breaking the perfect symmetry of the early universe that the fundamental forces became manifest, and matter coalesced to form the elementary particles that congealed into atoms which dance around as the forces dictate, giving rise to our vast and varied world. Perfect symmetry might be aesthetically appealing, but in practice it is sterile. Figuratively, I suppose, one could say that the choice is between remaining frozen in stagnant perfection and descending into the frantic whirl of being; the complexity of life is possible only when things and places are different from each other and change is possible.

Craftsmen of old knew this truth instinctively. It is said that when a particularly enchanting piece of art was made, in which pattern, repetition, and symmetry played a role in enhancing the beauty and quality of the design, artisans would introduce a small but deliberate flaw on the grounds that the realm of perfect symmetry is reserved for the divine.

But, insipid as unbroken symmetry might be, a complete lack of symmetry would be far worse. Luckily, a benevolent providence has ordained for us the best possible combination. Sandwiched between the possibilities of stagnation and chaos, we find ourselves in this “best of all possible worlds,” where there is simplicity of design but not implementation. The fundamental laws of our world may still be found by appealing to symmetry, but the applications are not thus limited. Even though the structure of our world is immeasurably rich, it is still economical in essence.

I know that this letter spans many new ideas, from the specific properties of a particular particle to general organizing principles in physics. Do not be frustrated if you can't grasp them all right now. Draw comfort from the fact that I have struggled with the expression of these concepts as much as, if not more than, you will struggle with their comprehension.

I was wrestling with words last night, striving for a clarity on paper that was perhaps not even there in my mind, when my eye fell on Rilke's *Letters to a Young Poet*. I bought this book in town last weekend, after learning that Rilke had spent some time in this area. He lived at Duino Castle, which, much like Miramare, is perched on a cliff over the cerulean waters of the Adriatic. There is a walking trail, just over a mile long, that leads up from the fishing village of Sistiana to Duino. It is abandoned now, but since it is named after Rilke, I like to imagine that it is a path he trod often, perhaps while pondering life's great mysteries.

I had started reading the *Letters* but had not yet gone very far. Every paragraph demanded to be savored, and I was reading slowly to extend my pleasure. When I picked up the book last night, I came upon a passage that spoke directly to me. A gentle admonition floated up across the years, and it was exactly what I needed to hear. I cannot think of any more fitting way to end this letter than to share it with you.

Rilke wrote, "Have patience toward all that is unsolved in your heart and try to love the questions themselves, like locked rooms and like books that are now written in a very foreign tongue. Do not now seek the answers, which cannot be given you because you would not be able to live them. And the point is to live everything. Live the questions now. Perhaps you will then gradually, without noticing it, live along some distant day into the answer."

I hope and pray that both of you learn to delight in walking with questions, and that you are granted the exultation of arriving at answers. The very wise Henri Poincaré wrote, "The scientist does not study nature because it is useful to do so. He studies it because he takes pleasure in it, and he takes pleasure in it because it is beautiful. If nature were not beautiful, it would not be worth knowing and life would not be worth living."

I wish for you to truly experience—and revel in—this "intimate beauty which comes from the harmonious order of [nature's] parts." Of earthly joys, not many are greater than this, except of course love—and that you will always have, for to you belongs my whole heart.

Your adoring Abba

P.S. Hassan, I purposely added more postage than was needed, on both this letter and the one I wrote your mother. As you must have noticed, each stamp is different. You may keep all of these for your collection, but please do not drown your sister's postcards in an attempt to pry the stamps off—I have used only duplicates on those.