II

HOW SCIENCE WORKS

CHAPTER 4

THE IRON RULE OF EXPLANATION

Enter the rule that defines modern science and gives it unprecedented knowledge-making power.

TO ACT OUT THE iron rule of explanation, allow me to introduce two scientists, both alike in dignity, yet separated by character and culture, education and personal style.

Professor Juliet Capulet, angular and elegant, was raised in schoolrooms, opera houses, and contemporary art museums. She is an aficionado of city life, with its acute organization, its grids and layers of structure upon structure. An ex-champion in the Multinational Mathematical Gymkhana, she prefers a hypothesis that explains methodically and exactly, even if it accounts for only a few isolated chapters of the volume of evidence. For a theory that meets her standards while comprehending everything, she is happy to wait.

The energetic and unkempt Professor Romeo Montague grew up surrounded by grassy expanses and thickets, by birds and beetles. He likes to think with his hands: having disassembled every household appliance from the vacuum cleaner to the sump pump, he set his heart on a life in science to better grasp nature in all its diversity and unruliness. He sees the virtues of mathematical exactitude, but ever restless and eager to push ahead, he will accept a back-of-the-envelope explanatory sketch with delight if that sketch makes sense of a broad array of phenomena.

Montague cares more about the quantity of evidence explained; Capulet cares more about the higher qualities of the explanation—simplicity, beauty, exactitude. Intensity against extent. Rigor versus reach.

Now let me send these characters back in time to stage some philosophico-scientific theater for your edification and enjoyment. It is based on a true story. The action takes place somewhere in the first half of the nineteenth century, a time during which the nature of heat was the subject of spirited debate. On one side of the debate is caloric theory, created and developed by French scientists from the 1780s through the 1820s, according to which heat is a kind of stuff. This "caloric fluid" flows steadily from warmer to cooler substances, causing them to become hotter and—newly engorged with fluid—to expand. By 1830, caloric theory has achieved a series of spectacular successes: the accurate calculation of the speed of sound through air, the delineation of a mathematical formula capturing precisely the rate with which heat flows through a material such as a metal rod, and an illuminating theory of the efficiency of steam engines and other heat-driven motors.

On the other side is the kinetic theory of heat (sometimes called the mechanical theory of heat), according to which heat is a kind of motion: it is the frenetic, disordered movements of the small particles of which things are made. The theory has a distinguished pedigree—it has been around since the seventeenth century, advocated as we will see by Sir Francis Bacon—but in the light of the French caloric exploits, it is looking rather outworn. Yet there are signs of a revival. The American-born scientist Benjamin Thompson, later to become the Bavarian noble Reichsgraf von Rumford, has just shown that unlimited quantities of heat can be generated by friction, by using a blunt drill to bore a cannon barrel immersed in water. Eventually the water boils—and keeps on boiling as long as the drill is grinding, well after all the system's caloric fluid ought to have drained away. The English chemist Humphrey Davy has conducted similar demonstrations and made similar arguments.

Professors Capulet and Montague find themselves at the center of this standoff, each seeking to explain the behavior of heat as best they can. Capulet's theoretical weapon of choice is the mathematical machinery of caloric theory: systematic, quantitative, precise. Montague, by contrast, is an enthusiast for the kinetic theory: although it cannot yet match the numerical accuracy of the equations of caloric theory, it provides intuitively appealing explanatory mechanisms for a wide range of phenomena, not least the generation of heat by friction.

Who is right: Capulet or Montague? Caloric or kinetic? In 1830, both choices are sound. Twenty years later, the kinetic conception will have overcome the caloric conception. But that has not happened yet. The choice between the two is dictated not so much by evidence as by temperament and taste. And so Montague and Capulet eye one another uneasily from opposite sides of the ideological stage.

In a few moments, the curtain will rise. That leaves just enough time for a brief authorial disclaimer: what follows is fiction, not historical fact. The background is historically real, as is the important problem with which the play begins. After that it is for the most part philosophical embellishment, intended not to tell the story of the science of heat but to direct the spotlight past the players and onto the backdrop, exposing a critically important, fixed plane of scientific agreement that lies behind the essentially subjective scientific dispute concerning which theory to believe.

Very well; let the drama begin. Capulet and Montague are deep in argument. The question is heat's ability to radiate across empty space, as when the sun pours its life-giving warmth across the interplanetary vacuum onto the earth. Such a thing seems impossible if the kinetic theory is correct, as there is nothing in a vacuum to vibrate, so nothing that can transfer the tiny motions that make up heat from source to destination. Perhaps Professor Capulet is right to reject kinetic theory, then—for surely here it finds a definitive refutation?

An imaginative Professor Montague wonders, however, whether heat might not travel through empty space in a different form than it travels through air, earth, and water. Perhaps, he suggests, there is a process—call it radiation—by which the vibratory energy of a hot body is transformed into "heat rays" that, like light rays, fly through empty space at high speed and, colliding with solid matter on the other side, induce the sort of vibrations that amount to heat proper? Professor Capulet considers that to be a transparently desperate attempt to save the kinetic theory from what is to any neutral observer an obvious falsification.

Montague's saving maneuver is, however, not only logically permissible, but factually quite correct. The sun's heat arrives at the beach thanks to electromagnetic radiation of certain frequencies, much of it in the infrared frequency band. These heat rays are not themselves a form of heat; rather, heat is transformed into electromagnetic energy, which travels as a ray and is then, upon striking the sand and sundry bathers' bodies, changed back to heat, exactly as Professor Montague's convoluted explanatory narrative maintains.

That said, Capulet's skepticism has a sound rationale. She needs only one kind of thing, caloric fluid, to explain the movement and behavior of heat. Montague has posited two kinds of things, heat itself and heat radiation, each popping up to do the job it is good for and then conveniently metamorphosing into the other when that's what's needed instead. At the very least, Montague seems to be violating the old methodological dictum "not to posit two entities when one will do"—usually, in honor of the medieval philosopher William of Occam, called "Occam's razor." It is hardly unreasonable for Capulet to threaten Montague with a close shave.

This is one more example, then, of the essential subjectivity of scientific reasoning: the wisdom of Montague's willingness to postulate radiation—a whole new way that heat might be shipped around the universe—depends on your perspective. To Montague, it is an inspired discovery; to Capulet, poetic indulgence. One scientist senses the first glimmer of the sun's rays and the lark's joyful song; the other only darkness and the plaintive anthem of the nightingale.

Montague and Capulet disagree as to which theory best explains the evidence. And they disagree on how particular pieces of evidence—most notably, heat's ability to travel through a vacuum—bear on each other's theories. Is this the version of *Romeo and Juliet* in which the lovers survive, only to have family squabbles lead slowly but inexorably to irreconcilable differences?

We expect Elizabethan drama to end stirringly in true love or violence, and modern drama (perhaps) in misery or self-recognition. The drama of science, however, never need end at all. Montague and Capulet can continue their dialogue indefinitely without running out of lines. That is because there is always a purely scientific way to perpetuate a scientific argument: make more measurements or conduct new experiments.

The unending script, the code of conduct for scientific argument according to which Montague and Capulet may continue their debate indefinitely, is provided by the methodological precept that I call the iron rule of explanation. What the rule says is simple enough: it directs scientists to resolve their differences of opinion by conducting empirical tests rather than by shouting or fighting or philosophizing or moralizing or marrying or calling on a higher power. That is all; it makes no attempt to interpret the evidence, to decide winners and losers. Indeed, its function is not so much to resolve the dispute as to prolong it. This perpetuation of the dramatic conflict for its own sake is the essence of the scientific method, and as the perpetuator-in-chief, the iron rule establishes itself as the heart, the soul, the life force of scientific inquiry.

The first act has ended in impasse. Montague said that heat traveling across empty space takes the form of radiation; Capulet said that it is a spume of caloric fluid jetting through the void. The iron rule shows them how to go on arguing, how to build a second scientific act around their disagreement, by improvising a series of experimental scenes—that is, by conducting a succession of empirical tests.

To decide the question of how heat moves through the void, Montague and Capulet might, for example, agree to measure the difference that a pane of glass makes to the velocity with which heat travels across a laboratory-made vacuum. If Montague is right and heat is transmitted by radiation, then heat rays should speed through the glass like light rays, barely slowing at all. If Capulet is right and heat is an invisible, flowing liquid—caloric fluid—then the glass should stop the heat, or at least slow it down significantly, as it works its way by diffusion from one side of the pane to the other.

Suppose they perform the experiment. The glass pane makes no difference to the time taken for heat to traverse the vacuum chamber, so the test looks to count in favor of Montague's theory and against Capulet's. Capulet might surrender, but then again she might fight on. After all, her caloric hypothesis predicts a significant slowdown only in conjunction with certain auxiliary assumptions—in particular, the assumption that the glass pane acts as an effective barrier to caloric fluid. Perhaps that's not the way it works. Perhaps firing heat at a pane of glass is like directing a fire hose at a fishing net: the stuff courses through holes invisible to us but huge by comparison with the particles that make up the fluid (if it is made of particles at all). To test Capulet's "fishnet" response, Montague concedes, more tests will be needed. Perhaps stack up many panes of glass at different angles? If the fishnet analogy holds, sufficiently many layers should slow caloric fluid right down.

Or suppose that the glass pane experiment comes out the other way: the glass slows heat flying down the chamber to a crawl. Exit Montague?

Figure 4.1. Montague proposes an experiment.

Not necessarily. Glass is transparent to the radiation that we call light, but perhaps it is opaque to heat radiation, acting like a thick black curtain. To test this "blackout" response, Capulet concedes, more tests will be needed. Perhaps try various other potential barriers? If heat rays are like light rays, then in the same way that some materials are transparent to light, some ought to be transparent to heat.

Whatever the outcome of the test, then, two things are certain. First, the "loser" will have the chance to save their theory by rejecting one auxiliary assumption in favor of another. Second, both scientists will agree on the kinds of further observation that will put the new auxiliary assumptions to the empirical test. Whichever way the glass pane experiment goes, for example, Capulet and Montague might agree to pursue their argument by trying different materials as barriers in the vacuum chamber to find out what lets heat through and what does not.

Because they follow the iron rule, Capulet and Montague are in a certain kind of agreement. It is not agreement as to the best theory of heat. It is not agreement as to what the evidence says about various theories of heat. It is rather a kind of procedural agreement, an accord as to how to go on arguing: by observation and experiment, and not in any other way.

Simply saying, "Resolve your disputes through empirical testing" is not enough to establish an accord, however; what's also needed is a shared sense of what counts as an empirical test. The iron rule provides the necessary definition, an objective criterion for empirical testing to which all scientists subscribe.

You have two hypotheses that you want to decide between. Is heat a special kind of substance, different from ordinary matter, or is it the disordered motion of small particles of matter? Is the earth 20 million years old or more than 100 million years old? Do its continents move, or do they stay in place? To answer these questions, the iron rule says, proceed as follows: find an experiment or observation that might have one of two possible outcomes, where the first outcome would be explained by the first hypothesis (or rather, a cohort containing the first hypothesis) but not the second, and vice versa. Perform the experiment or make the observation. See what happens.

Here, then, in short, is the iron rule:

1. Strive to settle all arguments by empirical testing.
2. To conduct an empirical test to decide between a pair of hypotheses, perform an experiment or measurement, one of whose possible outcomes can be explained by one hypothesis (and accompanying cohort) but not the other.

There lies the nub of the scientific method and so, once its subtleties have been spelled out in the chapters to come, the denouement of the Great Method Debate.

THE IRON RULE GIVES all scientists the same advice as to what counts as a relevant experiment or observation, regardless of their intellectual predilections, cultural biases, or narrow ambitions. It does not, however, pretend to prescribe what to believe on the basis of such tests. It is a rule for doing rather than thinking.

This uniformity of doing, this procedural consensus, may not sound like much. It is an etiquette for argument, an agreement on how to disagree. What can it bring to science, beyond a certain civility and orderliness? In spite of its apparently modest ambitions, the procedural consensus is precisely what secures the triumph of modern science.

The first benefit of the procedural consensus prescribed by the iron rule is simple continuity. Throughout history, religious traditions have been vulnerable to schism—to irreparable ideological separations yielding daughter traditions that lose the ability to reason with one another. In their mutual future lies suspicion, political jockeying, sometimes appalling violence. Islam split into its Sunni and Shiite branches; early Christianity into Roman Catholic and Eastern Orthodox camps; Roman Catholicism into numerous Protestant factions that denounced and battled the papal loyalists.

Philosophical and political traditions have the same susceptibility. Even natural philosophy, the study of nature that is the precursor of modern science, has at times fractured into factions that, though they coexisted peacefully, found little to talk about. Physics after Aristotle bifurcated into two schools, the Epicureans and the Stoics, the one affirming an atomistic view in which the universe is composed of nothing more than particles careening blindly through the emptiness of space, the other a view in which matter fills the universe according to the dictates of rationality. An individual seeking enlightenment while Stoicism and Epicureanism held sway, from about 300 BCE to 300 CE, could choose one school or the other or attempt to learn from both, but the schools themselves remained detached intellectual traditions, separated by both their physics and their philosophy of life, until economic decline, barbarian invasions, and Christianity finished them off.

Since the creation of modern science in the seventeenth century, empirical inquiry has seen nothing like this intellectual fission. There was no irreconcilable split between the caloric and kinetic schools of heat, nor did Kelvin's work cause biologists to cut off all conversation with physicists. They continued to work together because in the iron rule they had a shared rule of engagement and no choice but to use it.

Peaceful dialogue is a start, certainly better than the alternative, but it is hardly sufficient to guarantee progress, let alone science's fabled power to zero in on the truth. There must be more. The clues lie in the pioneering work of Popper and Kuhn on the importance of motivation and morale in science.

Scientific argument, unlike most other forms of disputation, has a valuable by-product, and that by-product is data. The iron rule encourages, instructs, obliges, or forces contending scientists to engage each other with observable fact alone. Theological and philosophical contentions are ruled out of court; as long as the protagonists are practicing science, their arguments must be conducted by experimental means. All of their need to win, their determination to come out on top—all of that raw human ambition that, on the modern sociological view of science, would subvert any objective code of inquiry—is diverted into the performance of empirical tests. The rule thereby harnesses the oldest emotions to drive the extraordinary attention to process and detail that makes science the supreme discriminator and destroyer of false ideas.

I wrote earlier about the long, tedious years of brain crushing and distillation required for the rival scientists Roger Guillemin and Andrew Schally to analyze the structure of the hormone TRH. (Schally estimated that in the course of his efforts to find the structure of another hormone, called LRF at the time, he had to process the hypothalami of 160,000 pigs to obtain less than a thousandth of a gram of the sought-after substance.) What kept them going? In part, as Schally pointed out, the Kuhnian belief that substances such as TRH and LRF existed and that current techniques were capable of revealing their structure. Everyone had that belief, however. The thing that made the difference between the Nobel Prize winners Guillemin and Schally, on the one hand, and their rivals who failed to complete the project, on the other, was the lust for victory. As Schally, who was born in 1926 in Vilnius, then part of Poland, said of one of those lesser rivals:

> He is the Establishment . . . he never had to do anything . . . everything was given to him . . . of course, he missed the boat, he never dared putting in what was required: brute force. Guillemin and I, we are immigrants, obscure little doctors, we fought our way to the top.

That extraordinary combativeness could have manifested itself in fierce metaphysical argument or in fine rhetoric. But because Schally's enterprise was bound by the iron rule, his ardent spirit was directed exclusively toward the production and analysis of TRH and so toward the generation of empirical data.

There is still more to the motivational power of the iron rule's procedural consensus. Think about chess. The rules of the game are simple and known to all. No player wastes time musing on or debating the rules; they simply follow them. Consequently, enormous energy is liberated to pursue the problem how best to play within the established regulations.

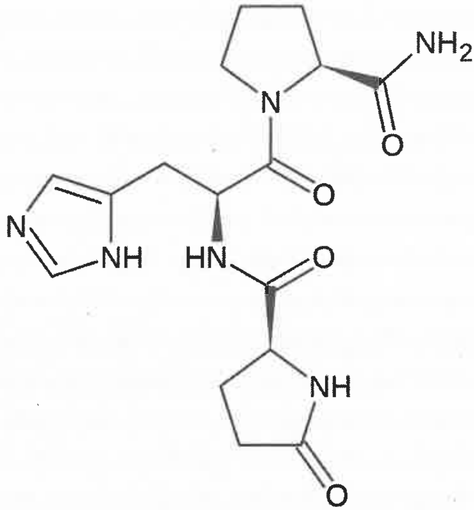

Figure 4.2. The fruits of procedural consensus: the molecular structure of TRH.

There are numerous compilations of openings, analyses of midgame strategies, commentaries on individual contests. No one doubts that this effort is important (to those who care about chess), because no one questions the framework within which it is carried out. Following the rules is just what chess is.

The iron rule establishes certain "rules of the game" in a similar way and with similar effect. The game is scientific argument, and the iron rule specifies what counts as a "legitimate move"—namely, performing an experiment or making an observation that generates relevant empirical evidence. Because the rule imposes a standard for what counts as empirical testing that never changes, questions of legitimacy are settled once and for all. Scientists' minds and money are thereby freed to focus on what can be done within the prescribed framework and thus on the question not of what makes for a legal move but of what makes for a good or excellent or brilliant move.

The consensus on testing enables, for example, an across-the-board agreement that certain experiments ought to be performed, even experiments that consume disconcertingly large quantities of labor and capital. In some cases, scientists who may have quite opposing theoretical perspectives and ambitions come together to form a temporary community sharing an experimental purpose and capable, because of its size and unity, of doing things beyond the means of a single research group. Of the scientists who searched for the top quark at Fermilab near Chicago in the 1990s or the Higgs boson at the Large Hadron Collider near Geneva in the 2010s, some were hoping to confirm the predictions of physics' Standard Model—which implies that these particles exist—while others were hoping to overturn the Standard Model ("I *hate* the Standard Model," said one to me), and others still were agnostic. They agreed, nevertheless, that the experiments were worth all their time and a lot of money.

More typically in science, research groups undertake their experimental projects independently, as in the case of Eddington's expedition to see the eclipse or Guillemin and Schally's independent attempts to isolate TRH. They set off alone, but thanks to the consensus brokered by the iron rule, they do so with the approval of their competitors, who agree that the experiment or observation will make a valid empirical contribution to the argument. This approval constitutes a kind of moral support, and in the modern era, where science is conducted mostly with other people's money, it also often enough secures financial support, as boards consisting of scientists with opposing views agree to direct funding to what seem to all members to be worthwhile projects—even if they hope for contrary outcomes.

The game of science is not always played against direct rivals. Sometimes it is more like a roomful of solitaire players, each striving to outdo the others in a competition to solve the multifarious problems posed by nature. The contest makes sense, though, only if everyone strives under the same unchanging regulatory regime.

And so, spurred on but at the same time boxed in by the iron rule, scientists conduct new experiments or make new observations that no one would otherwise have had the motivation or the resources to bring about—using satellite antennae to pick up patterns in the faint radio echoes of the Big Bang, gravimeters to detect minute differences in the thickness of the earth's crust, calorimeters to measure impalpable flows of energy in chemical reactions, gentle yet comprehensive techniques of excavation to reveal fossilized feathers, shattered pots, fallen temples—each carefully calibrated to favor, on empirical grounds only, one theoretical cohort over another.

My inspiration for emphasizing the importance of regulatory unity, I hope you see, is Kuhn's conception of paradigm-driven science. I have my differences with Kuhn. His games are individual scientific research programs—Newtonian physics, say, or evolutionary biology—each of which has its own distinctive set of rules, implicit in the paradigm, determining the permissible moves within the program. My game spans all of science: its iron rule governs Newtonian physics, microeconomics, and molecular genetics alike. For Kuhn, then, the rules change with the times—revolutionary times, that is—whereas like Popper's rule of falsification, my procedural consensus is forever, or at least for as long as science endures.

Further, a Kuhnian rule-making paradigm is a psychological idée fixe, alternatives to which cannot even be conceived by the scientists within its grip. I hold, by contrast, that science's iron rule is more like a sporting convention than a mental prison. Chess players can easily imagine different rules (and occasionally entertain themselves by doing so), but for the purposes of the competition, they agree to stick with the official set. Likewise, scientists abide by the iron rule not because it is impossible for them to conceive of an alternative, but because they know that the rule characterizes what it is to do science, and just as chess players want to play chess, they want to do science.

Although my disagreements with Kuhn are deep and important, I emphasize the extent to which I have taken over Kuhn's fundamental insight: that establishing beyond question the legitimacy of evidential maneuvers, whether by paradigm or by the unchanging iron rule, provides an intellectual security that supercharges the scientific enterprise's power to generate expensive but telling empirical tests.

The ideas in the last few pages are some of the most important that I have to share. Together they capture the way in which the procedural consensus orchestrated by the iron rule powers the scientific knowledge machine. I'll state them one more time.

First, the consensus ensures continuity. Provided that there are resources and the will to continue the investigation, there is no prospect that Montague and Capulet will arrive at a position where they can think of nothing to do or say that might bring them any closer together. There are no duels or divorces in science, no schisms to parallel the historical schisms in religion, politics, and philosophy, where opposing camps stop talking or worse. Always there is something that even the most bitter adversaries can agree to do next: another test.

Second (and consequently), the iron rule channels hope, anger, envy, ambition, resentment—all the fires fuming in the human heart—to one end: the production of empirical evidence. Capulet and Montague's complex feelings about one another must, as long as they follow a scientific script, work themselves out in the lab, the field, or the observatory.

Third, the fundamental conventions of the game, the legitimate evidential moves, are fixed for all time. This gives the players, the scientists, a quiet but firm certainty in the foundation of their enterprise, a stable intellectual and moral platform on which to construct great experimental enterprises, with all their financial, emotional, and physical demands.

In these ways, a protocol as weak as the iron rule—an agreement to argue by empirical testing alone conjoined with a fixed standard to determine what counts as an empirical test—gives science something that no other form of inquiry before it has had, a pool of empirical observations that dwarfs in its size, scope, subtlety, and precision anything the ancient or medieval natural philosophers could bring themselves to produce.

THE GREAT RESERVOIR of evidence tapped by the iron rule is essential fuel for science, but can it drive scientific progress unaided? Does a great mass of empirical data, left to stew, convert itself spontaneously into proportionally great quantities of theoretical truth?

It is not as easy as that. Empirical knowledge cannot be transformed directly into theoretical knowledge. As Eddington's telescopes, Kelvin's estimates of the earth's antiquity, and Montague and Capulet's measurements of heat radiation show, observations must be carefully interpreted if they are to tell us what theories to believe. There lies a daunting problem. I have given up on the objectivity of scientific reasoning and proposed in its place a purely procedural agreement. The agreement may bring forth gratifyingly large quantities of data. But without an objective scheme of interpretation, data in any quantity is useless. It says a different thing to each scientist.

How, then, does science go about finding the truth? A promising idea lies at the juncture of Shakespeare's death and the dawn of the Scientific Revolution. Let us go there.