CHAPTER TWO

Ducks, Rabbits, and Normal Science: Recasting the Kuhn's-Eye View of Popper

SHORTLY AFTER the publication of his enormously influential book *The Structure of Scientific Revolutions,* Thomas Kuhn offered "a disciplined comparison" of his and Popper's views of science in the paper "Logic of Discovery or Psychology of Research?" It begins with these lines:

> My object in these pages is to juxtapose the view of scientific development outlined in my book [*Structure*], with the better known views of our chairman, Sir Karl Popper. Ordinarily I should decline such an undertaking, for I am not so sanguine as Sir Karl about the utility of confrontations. . . . Even before my book was published two and a half years ago, I had begun to discover special and often puzzling characteristics of the relation between my views and his. That relation and the divergent reactions I have encountered to it suggest that a disciplined comparison of the two may produce peculiar enlightenment. (Kuhn 1970, 1)

"Peculiar enlightenment" is an apt description of what may be found in going back to Kuhn's early comparison with Popper and the responses it engendered. What makes my recasting of Kuhn peculiar is that while it justifies the very theses by which Kuhn effects the contrast with Popper, the picture that results is decidedly *un*-Kuhnian. As such I do not doubt that my recasting differs from the "peculiar enlightenment" Kuhn intended, but my task is not a faithful explication of what Kuhn saw himself as doing. Rather it is an attempt, at times deliberately un-Kuhnian, to see what philosophical mileage can be gotten from exploring the Kuhnian contrast with Popper. This exercise will serve as a springboard for the picture of experimental knowledge that I want to develop in this book.

Kuhn begins by listing the similarities between himself and Popper that place them "in the same minority" among philosophers of science of the day (Kuhn 1970, 2). Both accept theory-ladenness of observation, hold some version of realism (at least as a proper aim of science), and reject the view of progress "by accretion," emphasizing instead "the revolutionary process by which an older theory is rejected and replaced by an incompatible new one" (ibid., 2).

Despite these agreements Kuhn finds that he and Popper are separated by a "gestalt switch." Popper views the overthrowing and replacement of scientific theories as the main engine of scientific growth. Scientific knowledge, Popper declares, "grows by a more revolutionary method than accumulation—by a method which destroys, changes, and alters the whole thing" (Popper 1962, 129). Kuhn views such revolutionary changes as extraordinary events radically different from the "normal" scientific tasks of "puzzle solving"—extending, applying, and articulating theories. Moreover, the growth of scientific knowledge, for Kuhn, is to be found in nonrevolutionary or normal science. Although, in Kuhn's view, "normal science" constitutes the bulk of science, what has intrigued most philosophers of science is Kuhnian revolutionary science—with its big changes, gestalt switches, conversion experiences, incommensurabilities, and the challenges thereby posed to the rationality of theory change. Although there are several chapters on normal science in *Structure,* the excitement engendered by these revolutionary challenges has seemed to drown out the quieter insights that those chapters provide. Kuhn's description of normal science, when discussed at all, is generally dismissed as relegating day-to-day science to an unadventurous working out of "solvable puzzles" and "mopping-up" activities. In the view of Kuhn's critics, as Alan Musgrave (1980) puts it, normal science is either "a bad thing which fortunately does not exist, or a bad thing which unfortunately does exist" (pp. 41–42).

In this vein, Popper (1970) responds to Kuhn in "Normal Science and Its Dangers." Popper was aghast at Kuhnian normal science with its apparent call to "abandon critical discourse" and embrace unquestioning allegiance to a single accepted paradigm, encompassing theories as well as standards and values for their appraisal. Kuhnian normal science, were it actually to exist, Popper declares, would be pathetic or downright dangerous:

> In my view the "normal" scientist, as Kuhn describes him, is a person one ought to be sorry for. . . . The "normal" scientist, as described by Kuhn, has been badly taught. He has been taught in a dogmatic spirit: he is a victim of indoctrination. (Popper 1970, 52–53)

Most troubling for Popper is the alleged inability of normal scientists to break out of the prison of their paradigm or framework to subject it to (revolutionary) tests, and, correspondingly, Kuhn's likening theory change to a religious conversion rather than to a rational empirical appraisal.

Finding much to dislike about normal science, philosophers find little use for it when it comes to solving these challenges to the rationality of theory change. The big problems enter in revolutionary, not normal, science, and that is where those who would solve these problems focus their attention. I find Kuhnian normal science to be far more fruitful. It is here that one may discover the elements of Kuhn's story that are most "bang on," most true to scientific practice; and the enlightenment offered by the Kuhn-Popper comparison suggests a new way of developing those elements. Far from being the uncritical affair Popper fears, normal science, thus developed, turns out to offer an effective basis for severe testing. This, in turn, provides the key to getting around the big problems alleged to arise in revolutionary science, or large-scale theory change.

Let us pursue a bit further the contrasts Kuhn draws between his and Popper's philosophies of science. Except where noted, all references are to Kuhn 1970.

2.1 TURNING POPPER ON HIS HEAD

Kuhn asks, "How am I to persuade Sir Karl, who knows everything I know about scientific development and who has somewhere or other said it, that *what he calls a duck can be seen as a rabbit?* How am I to show him what it would be like to wear my spectacles?" (p. 3; emphasis added).

Kuhn's tactic is to take the linchpins of Popper's philosophy and show how, wearing Kuhnian glasses, they appear topsy turvy. While in Popper's view, what sets science apart from other practices is its willingness to continually subject its theories to severe and crucial tests, to the Kuhnian eye "it is normal science, in which Sir Karl's sort of testing does not occur, rather than extraordinary science which most nearly distinguishes science from other enterprises. If a demarcation criterion exists (we must not, I think, seek a sharp or decisive one), it may lie just in that part of science which Sir Karl ignores" (p. 6). But because normal science, for Kuhn, does not involve Popperian-style testing, Kuhn provocatively declares, "In a sense, to turn Sir Karl's view on its head, *it is precisely the abandonment of critical discourse that marks the transition to a science*" (p. 6; emphasis added).

If only we would view the highlights of the Popperian landscape through his spectacles, Kuhn proposes, we would see how Popper's view gets turned on its head. Specifically, we would see why, where Popper sees a fundamental theory failing a severe test, Kuhn sees a paradigm failing in its "puzzle solving ability" (crisis), and why, where

Popper sees a lack of testability, Kuhn sees a lack of puzzle solving. In so doing, Kuhn assures us, we would begin to see the sense in which "severity of test-criteria is simply one side of the coin whose other face is a puzzle-solving tradition," and with this, Kuhn proclaims, "Sir Karl's duck may at last become my rabbit" (p. 7).

I propose that we look at the high points of the Popperian landscape that the Kuhn's-eye view brings into focus. Why? First, I think that in identifying these points of contrast Kuhn makes a number of well-founded descriptive claims about scientific practice. Second, I think these claims have important normative epistemological underpinnings that have gone unnoticed. The highlights that interest me underlie the following portions of the above passages:

- "It is normal science, in which Sir Karl's sort of testing does not occur, rather than extraordinary science which most nearly distinguishes science from other enterprises." (P. 6)
- "It is precisely the abandonment of critical discourse that marks the transition to a science." (P. 6)
- "Severity of test-criteria is simply one side of the coin whose other face is a puzzle-solving tradition." (P. 7)

To extract the epistemological lessons I am after, however, it is not enough to turn our gaze toward the Kuhn's-eye view of Popper; spectacles capable of seeing the normative dimension are required.

Except for some tantalizing clues, Kuhn fails to bring out this normative dimension. This is not surprising given his view on what doing so would require. What more is there to say, Kuhn seems to ask, once one is done describing the behavior of practitioners trained in a certain way? Explaining the success or progress of science for Kuhn just *is* a matter of "examining the nature of the scientific group, discovering what it values, what it tolerates, and what it disdains. That position is intrinsically sociological" (p. 238).

Stopping with this kind of descriptive account has several shortcomings. First, it is not known which of the many features scientific communities share are actually responsible for scientific success. And even where we can correctly identify these features, there is an epistemological question that needs answering: why does following these practices yield reliable knowledge? Correlatively, why do enterprises not characterized by these practices turn out to be less successful sciences or not sciences at all?

Of Kuhn's key observations of normal scientific practice, I will ask: What is so valuable about normal science (as Kuhn describes it)? Why does the lack of a normal "puzzle-solving tradition" seem to go hand in hand with ineffective sciences or nonsciences? The bulk of my inquiry will be directed toward identifying the constraints and methods that enable what is learned in normal science to do what Kuhn rightly says it does: it extends, articulates, and revises theories, identifies genuine anomalies, functions creatively in developing alternative theories, and allows communication between paradigms (however partial).

2.2 Popper in Light of Kuhn: J. O. Wisdom

J. O. Wisdom (1974), responding to Kuhn's "Logic or Psychology," offers an insightful way of assimilating Popper and Kuhn. He goes so far as to suggest that "Kuhn's theory and (when correctly interpreted) Popper's (when developed) are identical" (p. 839). In Wisdom's reading, the difference between Kuhn's picture of revolutionary science as something that occurs only after periods of normal science followed by crisis and Popper's emphasis on testing and falsifying theories is only apparent. (He is far less sanguine about accommodating Kuhn's "sociology of acceptance.")

How does Wisdom manage to convert Kuhn's rabbit into a Popperian duck? As we saw in chapter 1, Popper himself is well aware that for a theory genuinely to fail a test it is not enough that one of its consequences turns out wrong. Even for Popper, Wisdom rightly notes, before a failed prediction is taken as falsifying a theory, "a whole crop of loopholes have to be investigated: checking mathematics, accidental disturbances, instruments in working order, bungling, unsuspected misconceptions about the nature of the instruments, adequate articulation of the fundamental theory" (ibid., 838), and such investigations are the sorts of things done in normal science. Normal science is called upon, in effect, to pinpoint blame. First it is needed to determine if a genuine anomaly is at hand, to obtain a Popperian "falsifying hypothesis." Even then the anomaly may be due to other errors: faulty underlying assumptions or faulty initial conditions. With enough anomalies, however, a (Kuhnian) crisis arises, and as it becomes more and more unlikely that initial conditions are to blame for each and every anomaly, there are finally grounds for pinning the blame on the fundamental theory. Out of Kuhnian crises spring forth Popperian severe tests!

Wisdom's insightful analysis is in my view completely correct as far as it goes. But it goes only so far as to show how Kuhnian normal science can serve as a handmaiden to Popperian testing. By starting from the Popperian point of view of what the task is and utilizing Kuhnian ideas to fill out that task, he gets the order wrong and misses most of the uses of normal science. Normal science is needed to get extraordinary science off the ground all right, but starting with the latter as primary fails to do justice to Kuhn's main point of contrast with Popper.

Kuhn is quite clear about the main point of contrast (e.g., in his "Reflections"). He says, "my single genuine disagreement with Sir Karl about normal science" is in holding that when a full-bodied theory is at hand "the time for steady criticism and theory proliferation has passed" (Kuhn 1970, 246). Scientists can instead apply their talents to the puzzles of normal science. "[Sir Karl] and his group argue that the scientist should try at all times to be a critic and a proliferator of alternate theories. I urge the desirability of an alternate strategy which reserves such behaviour for special occasions" (p. 243).

Evaluating Kuhn's alternative research strategy requires us to reanalyze the aims of normal science. The result will neither be to turn Kuhn's rabbit into Popper's duck (as Wisdom has), nor to turn Popper's duck into Kuhn's rabbit (as in Lakatos's rational reconstruction of Kuhn), but to convert Kuhn's sociological description of normal science to a normative epistemology of testing.

2.3 Normal Science as Normal Testing

Let us begin by asking what is involved when practitioners turn their attention and apply their talents to the tasks of normal science. Kuhn identifies three classes of problems. "These three classes of problems—determination of significant fact, matching of facts with theory, and articulation of theory—exhaust, I think, the literature of normal science, both empirical and theoretical" (Kuhn 1962, 33). Kuhn elaborates on each:

1. *Determination of significant fact.* This concerns "that class of facts that the paradigm has shown to be particularly revealing of the nature of things" (Kuhn 1962, 25). Examples include stellar position and magnitude, the specific gravities of materials, wave lengths, electrical conductivities, boiling points, acidity of solutions. A good deal of work goes into the goal of increasing the accuracy and scope with which such facts are known.

2. *Matching of facts with theory.* This concerns arriving at data that "can be compared directly with predictions from the paradigm theory" (ibid., 26). Typically, theoretical and instrumental approximations limit the expected agreement, and it is of interest to improve upon it. Find-

ing new methods and instruments to demonstrate agreement constantly challenges experimentalists.

3. *Articulating the paradigm theory.* This has three parts:

• Determining physical constants (e.g., gravitational constant, Avogadro's number, Joule's coefficient, the electronic charge).

• Determining quantitative laws. Kuhn's examples include determining Boyle's law relating gas pressure to volume, Coulomb's law of electrical attraction, and Joule's formula relating heat to electrical resistance and current.

• Conducting experiments to "choose among the alternative ways of applying the paradigm" (ibid., 29) to some closely related area. Each of these three classes of experiment has an analogous theoretical task.

In each class, Kuhn stresses, some current background theory or paradigm is needed to define the normal problem and provide means for determining if it is solved. Conjectured solutions to each problem may be viewed as hypotheses within a larger-scale background theory. Clearly, finding and evaluating hypotheses of these sorts is not a matter of uncreative hack science, nor does Kuhn suggest otherwise. What normal science does require, for Kuhn, are shared criteria for determining if problems are solved:

> No puzzle-solving enterprise can exist unless its practitioners share criteria which, for that group and for that time, determine when a particular puzzle has been solved. The same criteria necessarily determine failure to achieve a solution, and anyone who chooses may view that failure as the failure of a theory to pass a test. (Kuhn 1970, 7)

I do. But to emphasize that these are tests of the hypotheses of Kuhnian normal science, I will refer (in this chapter) to such testing as *normal testing.* Kuhn observes that "Tests of this sort are a standard component of what I have elsewhere labeled 'normal science' . . . an enterprise which accounts for the overwhelming majority of the work done in basic science" (ibid., 4). For reasons that will later become clear, I think that all experimental testing, if construed broadly, is based on normal testing. From experimental tests we acquire *experimental knowledge* (as characterized in chapter 1).

To digress about my terminology: The appellation "normal testing" of course stems from Kuhn's normal science, with which it correlates. Later, when it is more developed and the deviations from Kuhn are clarified, I will call it *"standard testing."* But Kuhn sees normal science in contrast to nonnormal science, whereas I shall reject the idea that there are two kinds of science. If all science is normal (or even standard), is it not peculiar to use such an adjective altogether? No. In logic we talk about normal systems and corresponding standard models, and, while various nonnormal systems and nonstandard models exist, one may debate whether the latter are needed for some purpose, for example, the logic of arithmetic. Likewise, various nonstandard accounts of testing exist. My position is that for scientific reasoning they are either not needed or offer inadequate models.

Normal Hypothesis Testing: The Test of Experiment

Perhaps to achieve a stark contrast with Popperian tests, Kuhn calls normal problems "puzzles," where the loser when a conjectured solution fails is not the fundamental theory but the practitioner who was not brilliant enough. But since "blaming the practitioner," even for Kuhn, means only that the practitioner's conjectured solution fails to hold up to testing, it is less misleading to talk in terms of testing conjectured solutions to normal problems.[1] Kuhn supports this reading:

> There is one sort of "statement" or "hypothesis" that scientists do repeatedly subject to systematic test. I have in mind statements of an individual's best guesses about the proper way to connect his own research problem with the corpus of accepted scientific knowledge. He may, for example, conjecture that a given chemical unknown contains the salt of a rare earth, that the obesity of his experimental rats is due to a specified component in their diet, or that a newly discovered spectral pattern is to be understood as an effect of nuclear spin. (P. 4)

These are certainly the kinds of substantive hypotheses that philosophers of science want a theory of testing to address, something that tends to be lost in calling them "puzzles." The next steps, the test of experiment, have a hypothetico-deductive flavor:

> The next steps . . . are intended to try out or test the conjecture or hypothesis. If it passes enough or stringent enough tests, the scientist has made a discovery or has at least resolved the puzzle he had been set. If not, he must either abandon the puzzle entirely or attempt to solve it with the aid of some other hypothesis. (P. 4)

1. Hilary Putnam (1981) proposes that Kuhn's puzzles follow the form of an explanatory scheme that he calls Schema II in contrast to the pattern of a test based on a prediction (Schema I). In Schema II, both a theory and a fact are taken as given, and the problem is to find data to explain the fact on the basis of the theory. I agree with Putnam that Schema II captures the form of many normal problems, and concur that this normal task has been too little noticed by philosophers of science. Still, the task of checking or arguing for a conjectured explanation falls under what I would term a normal test.

Further, Kuhn finds it "unproblematic" to call successful conjectured solutions "true":

> Members of a given scientific community will generally agree which consequences of a shared theory sustain the test of experiment and are therefore true, which are false as theory is currently applied, and which are as yet untested. (P. 264)

So, as I read Kuhn, to infer that a hypothesis sustains the test of experiment it must have passed "enough or stringent enough tests," and, correspondingly, to regard a normal hypothesis as true is to accept it as correctly solving the associated normal problem.

Several characteristics of normal tests emerge from these passages. The tests are directed at a specific normal hypothesis, say, *H*. The test criteria determine whether *H* passes or fails (further divisions could be added). If *H* passes "enough or stringent enough tests," then *H* is taken to solve the puzzle, that is, to be correct or true. If *H* fails, it is concluded that *H* does not solve the puzzle (that *H* is incorrect or false). Blaming the background theory is tantamount to changing the puzzle and is disallowed. Indeed, in my reading, the main purpose of calling a normal problem a "puzzle" is to call attention to the fundamental restriction on what counts as an admissible solution: if a conjectured solution fails the test, only the conjecture and "not the corpus of current science is impugned" by the failure (p. 5).

Kuhn cites, as an example, how some eighteenth-century scientists, finding anomalies between the observed motions of the moon and Newton's laws, "suggested replacing the inverse square law with a law that deviated from it at small distances. To do that, however, would have been to change the paradigm, to define a new puzzle, and not to solve the old one" (Kuhn 1962, 39). This solution was not admissible. The normal scientist must face the music.

These are useful clues, but not quite enough. After all, a community could agree on any number of rules or criteria to pass or fail hypotheses, to call a test stringent, and so on. Frustratingly, Kuhn does not spell out what more is required for normal test criteria to pass muster, leading some to suppose that community consensus is enough for him. Yet a careful look at the demands Kuhn sets for normal tests suggests how to flesh out the earlier clues.

Underlying the stringency demand, I propose, is the implied requirement that before a hypothesis is taken to solve a problem it must have stood up to scrutiny: it must be very unlikely that the hypothesis really does not solve the problem. This *reliability* (*or severity*) *requirement,* discussed in chapter 1, will be developed as we proceed. For now, to avoid prematurely saddling Kuhn with my notion, the "reliability" of a test will be interchangeable with its "severity." That a test is reliable, for me, describes a characteristic of a *procedure* of testing, or of test criteria.[2] It does not say that the test never errs in declaring a problem solved or not, but that it does so infrequently. In other words, the normal scientist declares a problem solved only if the conjectured solution has withstood a scrutiny that it would very likely have failed, were it not correct.

It is only by some such reliability or severity requirement, I maintain, that Kuhn is right to locate the growth of knowledge in normal science. We can simultaneously unpack this requirement and motivate it on Kuhnian grounds by continuing our normative questioning.

Justifying the Pursuit of Normal Science

Kuhn observes that the bulk of scientific practice is directed at normal testing, rather than at Popperian testing, where Kuhn construes Popperian testing as criticizing fundamental theories. What interests me is whether there is an epistemological rationale for this focus on one sort of activity rather than another. The answer, I propose, is that one learns much more this way. Focusing on normal testing is a better research strategy.[3]

To bring the answer into focus we first must ask, Why, in the face of a rich enough theory to "support a puzzle-solving tradition," is it fruitful to concentrate on normal problems? (No answer precludes there also being something to be gained by seeking alternative theories.) When I look at Kuhn while wearing my spectacles, I discern this reply: if one has an interesting theory, one with predictions, suggestions for improvement, challenging puzzles, and so on, then taking up its challenges *will teach us a great deal,* and *a portion of what is learned will remain despite changes in theory.* With respect to the solved problems in normal research, Kuhn says that "at least part of that achievement always proves to be permanent" (Kuhn 1962, 25). To ignore its challenges is to forfeit this knowledge.

One can go further. Pursuing normal problems is a good strategy because, if there are anomalies that call for changes in theory, doing so will reveal them as well as help point to the adjustments indicated.

> In the developed sciences . . . it is technical puzzles that provide the usual occasion and often the concrete methods for revolution. . . . Be-

2. The reliability of a test procedure must be distinguished from the reliability of any particular hypothesis that passes the test, as that is generally understood.

3. Kuhn (1970, p. 243) identifies the key difference between himself and Popper as one of research strategy.

> cause they can ordinarily take current theory for granted, exploiting rather than criticizing it, the practitioners of mature sciences are freed to explore nature to an esoteric depth and detail otherwise unimaginable. Because that exploration will ultimately isolate severe trouble spots, *they can be confident that the pursuit of normal science will inform them when and where they can most usefully become Popperian critics.* (Kuhn 1970, 247; emphasis added)

The rationale for pursuing normal problems is that (if done right) some positive payoff is assured. If normal science yields problem solutions, then new knowledge is brought forth. If normal testing determines that an anomalous result is real—that "it will not go away"—then there is knowledge of a real experimental effect (i.e., experimental knowledge). Further normal testing will indicate whether adjustments and revisions are called for. If, alternatively, the effect is a genuine anomaly for the underlying theory, normal science will let this be found out by means of gross or repeated failures (Kuhnian crisis). Even such crises, Kuhn notices, serve a creative function in developing alternative theories. Finally, normal science is the source of the most effective and severe tests of fundamental or basic theory:

> Though testing of basic commitments occurs only in extraordinary science, it is normal science that discloses both the points to test and the manner of testing. (P. 6)

and

> Because the [theory] test arose from a puzzle and thus carried settled criteria of solution, *it proves both more severe and harder to evade than the tests available within a tradition whose normal mode is critical discourse rather than puzzle solving.* (P. 7; emphasis added)

This last sentence gets to the heart of why, from the Kuhn's-eye point of view, "severity is the flip side of puzzle solving," and why one finds the most severe tests of theories, just what Popper seeks, in practices that have been engaged in the puzzle solving of normal science (normal testing). There is a clear epistemological ground for this, and a bit more spadework will uncover it. The last part of the quotation points the way. We have to ask what is wrong with "a tradition whose normal mode is critical discourse rather than puzzle solving"?

2.4 The Kuhn's-Eye View of Demarcation

It is important to keep in mind that the critical discourse Kuhn is disparaging is the special kind of criticism that he imagines Popper to be championing: a relentless attack on fundamentals. It helps, in grasping

Kuhn here, if each time we read "critical discourse," we tack on the phrase "rather than puzzle solving." For Kuhn, finding a practice whose normal mode is critical discourse rather than puzzle solving is the surest tip-off that its scientific status is questionable. Hence Kuhn's provocative claim that a demarcation criterion may be found in the portion of science badly obscured by Popperian spectacles (normal science). Nevertheless, as Kuhn remarks (p. 7), Popper's demarcation line and his often coincide, despite the fact that they are identifying very different processes.

On the face of it, their two demarcation criteria are nearly opposite. For Popper, the hallmark of science is criticism and testability, whereas Kuhn, in deliberate contrast, declares that "it is precisely the abandonment of critical discourse that marks the transition to a science" (p. 6). To call what goes on in good normal testing an abandonment of critical discourse is highly misleading, because normal science itself is based on severe and critical normal tests. However, on my reading, what Kuhn takes good sciences to abandon is not normal testing—where all the fruitful learning really takes place—but rather "critical-discourse-rather-than-puzzle-solving." Good sciences do not and should not do what Kuhn takes Popper to be championing: relentlessly attacking fundamental theories, looking always for rival theories, and doing so to the exclusion of the positive learning of normal science. Although I do not endorse this provocative, idiosyncratic usage of "critical discourse," for the purposes of this chapter it helps us to construe Kuhn's demarcation criterion plausibly.

Astrology

To illustrate his contrast with Popper, Kuhn chooses astrology, out of a wish to avoid controversial areas like psychoanalysis (p. 7). His focus, he says, is on the centuries during which astrology was intellectually respectable. The example functions not only to make out his demarcation but also to show "that of the two criteria, testing and puzzle solving, the latter is at once the less equivocal and the more fundamental" (p. 7). Astrology was unscientific, says Kuhn, not because it failed to be falsifiable, nor even because of how practitioners of astrology explained failure. The problem is that astrologers had no puzzles, they could not or did not engage in normal science.

Engaging in normal science requires a series of puzzles and strict criteria that virtually all practitioners agree to use to tell whether the puzzles are solved. But a practice does not automatically become scientific by erecting such a series of puzzles and rules to pronounce them solved or not. Becoming a genuine science is not something that can occur by community decree, nor does Kuhn think it is.[4] Kuhn balks at those who would find in him recipes for becoming scientific, apparently unaware of how he invites this reading by failing to articulate the kinds of tests needed to carry out normal science legitimately and why only these tests qualify. Still, in several places Kuhn hints at the criteria normal testing requires (reliability or stringency). The most telling of all, I find, is his critique of astrology.

With astrology, Kuhn observes, not only are the predictions statistical, but there is a tremendous amount of "noise" from background uncertainties.

> Astrologers pointed out, for example, that . . . the forecast of an individual's future was an immensely complex task, demanding the utmost skill, and extremely sensitive to minor errors in relevant data. The configuration of the stars and eight planets was constantly changing; the astronomical tables used to compute the configuration at an individual's birth were notoriously imperfect; few men knew the instant of their birth with the requisite precision. No wonder, then, that forecasts often failed. (P. 8)

Kuhn's point seems to be this: astrology, during the centuries when it was reputable, did not fail to be scientific because it was not testable nor because practitioners did not take failures as grounds to overthrow astrology. Plenty of perfectly good sciences act similarly. The reason the practice of astrology was unscientific is that *practitioners did not or could not learn from failed predictions.*[5] And they could not learn from them because too many justifiable ways of explaining failure lay at hand. They could not use failures or anomalies constructively.

> The occurrence of failures could be explained, but particular failures did not give rise to research puzzles, for no man, however skilled, could make use of them in a constructive attempt to revise the astrological tradition. There were too many possible sources of difficulty, most of them beyond the astrologer's knowledge, control, or responsibility. Individual failures were correspondingly uninformative. (P. 9)

The above passage is the most revealing of all. For failed predictions to "give rise to research puzzles," a failure must give rise to a fairly well defined problem; specifically, the problem of how to explain it. It must be possible, in other words, to set up a reliable inquiry to

4. This is stressed by Hoyningen-Huene (1993) in his analysis of Kuhn's *Structure of Scientific Revolutions.* (See, for example, p. 193.)

5. My reading is not affected by the fact that Kuhn thinks it wrongheaded to call failed solutions "mistakes", and that he limits mistakes to errors in applying some rule, e.g., mistakes in addition.

determine its cause and/or the modifications called for. This is the kind of information normal tests can provide.

> Compare the situations of the astronomer and the astrologer. If an astronomer's prediction failed and his calculations checked, he could hope to set the situation right. Perhaps the data were at fault. . . . Or perhaps theory needed adjustment. . . . The astrologer, by contrast, had no such puzzles. (P. 9)

To "set the situation right" one needs to be able to discriminate between proposed explanations of the failure. Unless one can set up a stringent enough test of a hypothesized explanation, so that its passing can reliably be attributed to its being correct, that failed prediction will be unconstructive or uninformative.

By the same token, so long as there is no way to cut down these alternative explanations of failure, there is no ground for arguing that the failures *should* have been attributed to the falsity of astrology as a whole. In other words, if failed predictions do not give rise to research puzzles (reliable inquiries into their cause), then one cannot come to learn whether and, if so, how they can be explained within the global background theory. Thus, they cannot warrant discrediting the whole theory; they cannot warrant (Popperian) critical discourse.

It becomes clear, then, that mere critical discourse is not enough for genuine science. In fact, Kuhn must see the case of astrology as one in which the normal day-to-day practice is critical discourse (i.e., critical discourse rather than puzzle solving). Constructive criticism, one might say, requires at least being able to embark on an inquiry toward solving the Duhemian problems that will arise.

Unwarranted Critical Discourse

The practitioners of astrology, Kuhn notes, "like practitioners of philosophy and some social sciences . . . belonged to a variety of different schools, and the inter-school strife was sometimes bitter. But these debates ordinarily revolved about the *implausibility* of the particular theory employed by one or another school. Failures of individual predictions played very little role" (p. 9, n. 2). Practitioners were happy to criticize the basic commitments of competing astrological schools, Kuhn tells us; rival schools were constantly having their basic presuppositions challenged. What they lacked was that very special kind of criticism that allows genuine learning—the kind where a failed prediction can be pinned on a specific hypothesis. Their criticism was not constructive: a failure did not *genuinely indicate* a specific improvement, adjustment, or falsification.

Thus, I propose to construe the real force of Kuhn's disparaging practices "whose normal mode is critical discourse" as disparaging those practices that engage in criticism even where the criticism fails to be driven by the constrained tests that exemplify good normal science.[6] What is being disparaged, and rightly so, is unwarranted and unconstructive criticism. When the day-to-day practice is criticism that is not the result of the stringent constraints of normal testing, then that criticism is of the unwarranted or unconstructive variety. It is mere critical discourse. Nonsciences engage in mere critical discourse, not genuine criticism that allows learning from empirical tests.

Learning from tests requires being able to learn not only from failed predictions but also from successful ones. For practitioners of astrology, both failed and successful predictions were uninformative. They could not learn from successful predictions because they would not provide a warrant for crediting any astrological theory. Credit does not go back to any astrological theory because there were no grounds for attributing a successful prediction to some astrological cause, for example, to the stars and planets being in particular positions. Successful astrological predictions are likely, even if astrology is false: the tests are not severe. We will later see how to make this notion of severity concrete.

Astrology exemplifies an extreme situation in which severe tests are precluded. The situation might be described in modern statistical terms as having too much uncontrolled variability, or as lacking a way to distinguish the "signal" from the noise. The situation, Kuhn notes, is typical of practices that one might call "crafts," some of which eventually make the transition to sciences (e.g., medicine). (Current theories of the stock market might be said to be crafts.)

The transition from craft to science, Kuhn observes, correlates with supporting normal science or normal testing. However, Kuhn's demarcation slogan makes him appear to be saying that the transition comes about by "abandoning critical discourse." Kuhn fails to identify the kind of abandonment of criticism that is actually conducive to making a practice more scientific. It is unwarranted and unconstructive criticism that should be abandoned and replaced by the warranted criticism of normal testing.

Let us go back to a practice that, unlike astrology, is sufficiently developed to support normal testing (puzzle solving). If a hypothesized

6. This should be qualified to refer only to enterprises for predicting, controlling, or understanding the physical world, in short, to intended sciences. It would not be a disparagement, say, of art.

solution to a normal problem fails a test, it could, theoretically, be accounted for by alleging a fundamental flaw in the underlying theory—but such a criticism would very likely be unwarranted (at least if just from this one failure). Thus, to regularly proceed this way would very often be in error, thereby violating the reliability requirement of normal testing. On these grounds normal science calls for abandoning this type of criticism. For the same reason it admonishes the practice of dealing with a failed solution (failed hypothesis) by changing the problem it was supposed to solve. An enterprise that regularly allowed such a cavalier attitude toward failure would often be misled. Likewise, in the case where a hypothesis passes a test: if there is too much leeway allowed in explaining away failures, passing results teach little if anything.

Changing the problem, blaming one's testing tools or the background theory *where these are unwarranted,* is the kind of criticism that should be disallowed. Only then can the practice of hypothesis appraisal be sufficiently constrained to identify correctly genuine effects, gain experimental knowledge—or more generally, to accomplish the tasks of normal science *reliably.*[7] Thus recast, Kuhn's demarcation criterion intends to pick out those practices that afford experimental learning. I suggest we view such a demarcation criterion as indicating when *particular inquiries,* rather than whole practices, are scientific. It becomes, roughly:

> *Demarcating scientific inquiry:* What makes an empirical inquiry scientific is that it can and does allow learning from normal tests, that it accomplishes one or more tasks of normal testing *reliably.*[8]

This criterion becomes more specific when particular types of normal testing results are substituted. For example, an important type of normal test result is a failed prediction. The difference between a scientific and an unscientific treatment of a failed prediction is the extent to

7. Popper makes it clear that he allows the critical method to refer to the minor tests that Kuhn counts as puzzles. Popper (1974) addresses this point with an example. The heating engineer needing to figure out how to install a central heating system under unusual conditions may have to throw away the rule book of normal practice to find the solution. "When he works by trial and the elimination of error, and when he eliminates the error by a *critical* survey of tentative solutions, then he does not work in this routine manner; which for me makes him a scientist. But Kuhn . . . should either say that he was not a scientist, or an extraordinary one" (Popper 1974, 1147).

8. I am assuming that the empirical inquiry is aiming to find out about the world.

which it is used to learn (about its cause, about needed modifications, and so on).

So far our analysis has brought us to the following recasting of the Kuhnian observations with which we began: To understand the nature of the growth of scientific knowledge, one should look to tests of local hypotheses (normal experimental testing). An adequate account of normal testing should serve each of the functions Kuhn accords it, with the additional proviso that it do so reliably and with warrant. From our vantage point, what distinguishes Kuhn's demarcation from Popper's is that for Kuhn the aim is not mere criticism but constructive criticism.[9]

2.5 Paradigmatic Models of Error

The Kuhnian picture of the role of background paradigms in normal science provides a useful framework for pursuing these ideas about experimental testing. While I would deny that a practitioner needs to work within a single paradigm, it will do no harm to see a given inquiry as within a single paradigm and it will make it easier to gain access to Kuhn's story. By the time the story is completed, delineating paradigms will not matter anyhow.

Kuhn's notion of paradigm is notoriously equivocal.[10] We may allow that a Kuhnian paradigm includes theories, specific hypotheses, and an ontology, as well as research aims and methods both for directing normal research and testing hypotheses. (In "Reflections on My Critics," p. 271, Kuhn says he would prefer to use the term "disciplinary matrix.") For Kuhn, sharing a paradigm is what accounts for "relative unanimity in problem-choice and in the evaluation of problem-solutions" (p. 271). However, we must carefully distinguish what Kuhn runs together here. What goes into choosing a problem is quite different from what goes into criticizing proposed solutions. We need, in short, to distinguish the paradigm's role in providing *research guides*—a source of problems and guides for solving problems or puzzles—from *experimental testing models*—tools for testing hypothesized solutions or for normal hypothesis testing. The second category includes tools for criticizing such tests. These two categories of tasks may overlap, but no matter. The point of the distinction is to see why

9. I hope it is clear that my use of the term "constructive criticism" differs radically from Lakatos in his attempt to revise Popper in light of Kuhn. For Lakatos, constructive criticism means replacing large-scale theories (progressively). For me, it means obtaining experimental knowledge by local arguments from error.

10. See, for example, M. Masterman 1970 and D. Shapere 1984, chaps. 3 and 4.

changing one's research program is not the same as changing one's experimental testing tools.

Paradigm as Research Guide

Let us begin with the role of a paradigm in providing a set of research guides. Its role is to supply the questions that need answering and suggest the kinds of answers afforded by given instruments, experiments, and tests. For Kuhn, a paradigm also creates a situation that fosters exploring subjects in esoteric depth and satisfies the psychological and practical requirements for sustaining such exploration.

This activity requires researchers to "accept" the program, but only in the sense that they choose to work on its problems and utilize the tools it offers for doing so. In Larry Laudan's terminology, they choose "to pursue it"; and, as he rightly urges, it is important to distinguish acceptance in this sense from taking an epistemological stance toward the theory.

Kuhn, in contrast, often suggests that the paradigm must have a grip on the minds of those scientists working within (i.e., accepting) it, allowing them to perceive the world through the paradigm. Is it true that working on a research program demands total immersion in the paradigm? Kuhn is right to insist that solving the problems of normal science requires creative brilliance (it is not hack science by any means), and perhaps total immersion is the most effective way to attain solutions. This could probably be investigated. Nevertheless, the results of such an investigation would be irrelevant to the questions about the epistemological warrant a theory might be required or entitled to have.

The factors that enter into choosing to take up a research program, pledging allegiance to it, living in its world—all the things Kuhn associates with accepting a global theory or paradigm—include psychological, sociological, pragmatic, and aesthetic values. Only by assuming these to be inextricably entwined with theory testing does Kuhn perceive the latter to turn on these values as well. Before considering global theory testing we need to focus on the second role of the global background theory or paradigm: its role in normal testing.

Experimental Testing Models: Error Paradigms

Significantly, Kuhn remarks that he was originally led to the concept of a "paradigm" in thinking of the concrete problem solutions or exemplars that practitioners share and that enable them to agree whether a problem is solved (p. 272). This is the role I propose to give to certain *experimental testing models* or *testing exemplars*. At one level, I am saying precisely what Kuhn says here about the role of shared exemplars. Through them one gains the ability to see a number of experimental problems or contexts in the same way, permitting the application of similar techniques. One grasps standard ways to ask questions to arrive at experimentally determinable answers, and one learns what does and does not count as a satisfactory solution, an adequate fit, and a "good approximation." Nevertheless, I depart from Kuhn in several important ways. While this results in a view of normal science very different from Kuhn's official position, it is in keeping with my normative recasting of Kuhn. The main differences (to be discussed more fully in later chapters) are these:

1. *Normal tests are not algorithms or routines.* If one unearths what actually goes on in normal testing, one sees that even sharing experimental test examples does not secure the relatively unproblematic means of testing normal hypotheses that Kuhn imagines. Uncertainties about experimental assumptions, significance levels, the appropriateness of analogies, and so on have to be confronted. As a result, there is often much disagreement about the results of normal testing, and there is plenty of opportunity for biases, conscious and unconscious, to enter. Where consensus is reached, it is not because of anything like a shared algorithm. It is because, in good scientific practices, the very problems of interpreting tests, critiquing experiments and such, themselves "give rise to research puzzles" in the sense we have discussed. Nor do these mechanisms for reliable tests in the face of threats from biases become inoperative in large-scale theory appraisal.

2. *Normal testing exemplars correspond to canonical experimental models.* In my view, standard examples or normal testing exemplars are not a set of tools available only to those working within a given global theory or paradigm. Instead they consist of any models and methods relevant for testing solutions of normal problems, and these come from various background theories, from mathematics and statistics, and from theories of instruments and experiments.

It is no objection to this idea of a pool of shared models that in a particular paradigm the model comes dressed in the special clothing of that paradigm. What is shared is the corresponding bare-bones or *canonical version* of the model. For example, a particular instrument used to test a predicted quantity may be characterized as having a specific Normal distribution of errors, and the rule for declaring the prediction "successful" might be that the result be within two standard error units. The corresponding bare-bones or canonical model would be the mathematical family of Normal distributions and the corresponding statistical rules for declaring a difference "statistically significant." That canonical model, once articulated sufficiently, is available for use for a host of experimental inquiries—it is not paradigm specific.[11] It is like a standard instrument; indeed, a physical instrument is often at the heart of a canonical model.

3. *The use of exemplars in normal testing is open to objective scrutiny.* Paradigmatic experimental examples are exemplary because they exemplify cases where the kinds of errors known to be possible or problematic in the given type of investigation are handled well—that is, ruled out. Include also examples of infamous mistaken cases, especially those thought to have surmounted key problems. Kuhn's use of astrology is an excellent example of what I have in mind; it is a classic case of a nonscience. As we proceed, other examples will arise.

The use of paradigmatic exemplars is open to a paradigm-independent scrutiny, that is, a scrutiny that is not relative to the background theory within which their use takes place. Although they may be applied in a routine manner, their appropriateness typically assumed, the exemplars are used in *arguments* appraising hypotheses, and these arguments have or fail to have certain properties (e.g., reliability, severity). For example, the rule for determining whether agreement with a certain kind of parameter (with a certain distribution of error) is "good enough" may be the two-standard-deviation rule. Normal practitioners can and do criticize such rules as appropriate or not for a given purpose.[12]

Interestingly, Kuhn's attitude toward the exemplars of normal tests is analogous to Popper's treating the decisions required for testing as mere conventions. They simply report the standards the discipline decides to use to declare a problem solved or not.[13] By Kuhn's own lights, however, before normal practitioners may take a puzzle as solved the hypothesized solution must have passed stringent enough tests. The arsenal needed for normal testing, then, is a host of tools for detecting

11. It does not follow that scientists will recognize that the canonical model is appropriate to a given problem. As we will see (e.g., in chapter 7), important mistakes could have been avoided in Brownian motion tests if the applicability of certain statistical models had been noticed.

12. Kuhn's use of the examples of astrology and astronomy is a good case in point. Suppose he had used the latter to exemplify a practice *lacking* a puzzle-solving tradition. Such a use would be criticized as allowing erroneous characterizations of good scientific practices.

13. Laudan (1996) makes this point in "The Sins of the Fathers." Kuhn's relativism about standards, Laudan says, is the exact counterpart of Popper's methodological conventionalism or Carnap's principle of tolerance.

whether and how conjectured hypotheses (of a given type) can fail. They call for methods capable not only of determining whether a hypothesis correctly solves a problem, but also of doing so reliably. Appraising the use of exemplars in normal testing turns on how well they promote these aims—the very aims we extricated in our normative recasting of Kuhn.

In this connection one might note how Kuhn, while claiming that a gestalt switch separates him from Popper, is able to criticize Popper's research strategy on the grounds that one does not learn much through critical discourse rather than puzzle solving.

2.6 GETTING SMALLER—A CONSEQUENCE OF SEVERITY AND INFORMATIVENESS

The most valuable idea that comes out of the testing-within-a-paradigm concept at the same time gets to the heart of Kuhn's contrast with Popper. The idea is that testing a hypothesis, if the test is to be informative, is not to test everything all at once but to test piecemeal. The necessity of proceeding piecemeal follows from the desiderata of normal tests: that they be reliable (severe) and informative (constructive). This requires two things of tests: a hypothesis H is taken to solve a problem only if it passes sufficiently stringent or severe tests, and some hypothesis (or other) is likely to be taken to solve the problem if it actually does so (at least approximately).

A major flaw in Popper's account (recall chapter 1) arises because he supplies no grounds for thinking that a hypothesis H very probably would not have been corroborated if it were false. ("Not-H" included all other hypotheses that were not yet considered.) Satisfying the aims of good normal testing, in contrast, directs one to select for testing the hypothesis where "H is false" (not-H) does not refer to all other hypotheses in the domain in question. Rather it refers to a specific way in which a conjectured solution, H, could be wrong—could be erroneously taken as actually solving the problem. (For example, H might assert that an effect is systematic, not an artifact, and not-H that it is an artifact.) I will elaborate on this in chapter 6.

The twin desiderata of normal tests, which we can express as "be stringent, but learn something," compel a localization of inquiry in another sense, one that David Hull (1988) precisely puts his finger on. In the section "Summa Contra Kuhn," Hull explains that "scientists are willing to accept certain problems as solved and proceed to new problem areas" even on the basis of an apparently small number of tests "because they are confident that *error ramifies*. If the hypotheses that they are accepting in order to attack new problems are mistaken, the results of related, though partially independent, research are likely to signal that something is wrong" (p. 496; emphasis added).

Later I will have much more to say about the points of the last two paragraphs. Here my main concern is how my reading of Kuhn substantiates his contrast with Popper. The aims of normal testing give an entirely new impetus to a slew of microinquiries that from behind Popperian spectacles[14] might appear unexciting, without risk.

Far from desiring boldness in Popper's sense, Kuhn observes how normal practitioners often set out to achieve an outcome that is already anticipated or seek to redetermine a known result—bolstering the impression of normal science as lacking novelty. A very different impression arises if it is seen that a central aim of normal science is to improve on its own tools. It is not so much the new information about the scientific domain that is wanted but new ways of minimizing or getting around errors, and techniques for ensuring Hull's point, that important errors "ramify rapidly."

As Kuhn puts it, "Though its outcome can be anticipated . . . the way to achieve that outcome remains very much in doubt" (Kuhn 1962, 36). A good deal about method is likely to be learned by finding out how to achieve the expected outcome. Consider, for example, the continued interest in using eclipse results to estimate the deflection of light. As will be seen in chapter 8, a main problem was finding improved instrumental and analytical techniques. Seeking alternate ways to elicit a known solution is often an excellent route to discovering a new and clever mode of interacting with nature. One of the most effective ways to test and learn from such interaction is through quantitatively determined effects.

Quantitative Anomalies

If we continue to look at Kuhn with normative glasses, we see that he has put his finger on the rationale of normal tests. They provide experimental constraints that allow learning from tests—both to ground the three types of normal hypotheses needed to extend and flesh out theories (section 2.3), as well as to substantiate a crisis.

A crisis emerges when anomalies repeatedly fail to disappear. Crises, while heralding troubling times, as Kuhn stresses, also have an important positive role to play: "[T]hough a crisis or an 'abnormal situation' is only one of the routes to *discovery* in the natural sciences, it is prerequisite to *fundamental inventions of theory*" (Kuhn 1977, 208).

14. Perhaps I should say wearing what Kuhn takes to be Popper's spectacles.

While "prerequisite" is probably too strong, crises have an especially creative function because of the knowledge they embody. Examining Kuhn's views on crises, however exceptional and rare he feels them to be, reveals a lot about the kind of knowledge that normal science is capable of bringing forth.

True, when Kuhn turns his gaze to crisis and theory change, he sees the problem as calling for a sociopsychological solution. It calls for, he thinks, a study of what *scientists consider* an unevadable anomaly and "what scientists will and will not give up" to gain other advantages (Kuhn 1977, 212). The problem, put this way, Kuhn says, "has scarcely even been stated before." For philosophers of science, the problem would usually be posed as asking after the nature and warrant of discrediting and testing theories. Nevertheless, what emerges from Kuhn's point of view is relevant for the more usual problem (which is, of course, the problem that interests me), provided we keep on our normative-epistemological spectacles.

What emerges from Kuhn's descriptive inquiry is that quantitative measurement and knowledge of quantitative effects are of key importance for both crisis and testing. This is particularly clear in Kuhn 1977. What makes an anomaly so "persistently obtrusive" and unevadable as to precipitate a crisis, Kuhn stresses, is that it be quantitatively determined:

> No crisis is, however, so hard to suppress as one that derives from a quantitative anomaly that has resisted all the usual efforts at reconciliation. . . . Qualitative anomalies usually suggest *ad hoc* modifications of theory that will disguise them, and . . . there is little way of telling whether they are "good enough." An established quantitative anomaly, in contrast, usually suggests nothing except trouble, but at its best it provides a razor-sharp instrument for judging the adequacy of proposed solutions. (P. 209)

This can be supplied with a very plausible normative basis: the quantitative anomaly identifies a genuine experimental effect, in particular, a discrepancy of a specified extent, and rather precise statistical criteria can determine whether a proposed solution adequately accounts for it. For example, if a two-standard-deviation discrepancy has been identified, then a hypothesis, say *H*, that can account for an effect of at most 0.1 standard deviation fails to explain the identified discrepancy. Thus the quantitative aspect of the discrepancy makes it clear that hypothesis *H fails to solve the problem.*

Here the Kuhnian function of quantitative anomalies sounds very Popperian, and it is. But for Kuhn the quantitative knowledge arose from a normal problem, an effort to learn something—not from a test of a large-scale theory.

True to his perspective on the problem, Kuhn sees himself as reporting on the behavior and attitudes of scientists. Yet the respect scientists show for quantitative effects is not just a sociopsychological fact about them. It is grounded in the fact that to try to blunt the sharp criteria quantitative information affords would be to forfeit accuracy and reliability, to forfeit learning. As Kuhn reports:

> I know of no case in the development of science which exhibits a loss of quantitative accuracy as a consequence of the transition from an earlier to a later theory. . . . Whatever the price in redefinitions of science, its methods, and its goals, scientists have shown themselves consistently unwilling to compromise the numerical success of their theories. (Kuhn 1977, 212–13)

Fortunately, our spectacles allow us to get beyond merely noticing that scientists appear unwilling to give up "quantitative accuracy" and "numerical success," and enable us to discern when and why this obstinacy is *warranted.*

This discernment is a task that will engage us throughout this book. Two things should be noted right off: First, it is not numerical success in the sense of doing a good job of "fitting the facts" that warrants clinging to. Only the special cases where numerical success corresponds to genuine experimental effects deserve this respect. One, but not the only, way of demonstrating such experimental knowledge is Hacking's favorite practice: intervening in phenomena. Second, if we press the normative "why" question about what makes quantitative effects so special and quantitative knowledge so robust, we see that what is desirable is not quantitative accuracy in and of itself. What is desirable is the strength and *severity of the argument* that is afforded by a special kind of experimental knowledge. As such, it makes sense to call all cases that admit of a specifiably severe or reliable argument "quantitative," so long as this special meaning is understood. (This is how I construe C. S. Peirce's notion of a "quantitative induction," a topic to be taken up in chapter 12.)

Quantitative knowledge teaches not only about the existence of certain entities but also about the properties of the process causing the effect. Such knowledge has primacy, as Kuhn recognizes, "whatever the price in redefinitions of science, its methods, and its goals."

Pause

Our recasting of normal science, I believe, substantiates the three highlights of Kuhn's contrast with Popper with which we began: (1)

"It is normal science, in which Sir Karl's sort of testing does not occur, rather than extraordinary science which most nearly distinguishes science from other enterprises" (p. 6); (2) "It is precisely the abandonment of critical discourse that marks the transition to a science" (p. 6); and (3) "Severity of test-criteria is simply one side of the coin whose other face is a puzzle-solving [i.e., a normal science] tradition" (p. 7).

Briefly, our gloss on them went as follows: The fundamental features of scientific inquiries are to be found in the criteria of normal testing, and these criteria demand stringent normal tests, not (uninformative) attacks on fundamental theory. Because anomalies that are reliably produced in normal tests indicate real effects that will not go away, they provide the most severe tests of theories—when these are warranted. This explains Kuhn's promise that scientists "can be confident that the pursuit of normal science will inform them when and where they can most usefully become Popperian critics" (p. 247), that normal science will tell them when and where to find fault with the underlying theory. But Kuhn, we shall see, reneges on his promise. Once having brought normal scientists to the crisis point, Kuhn still will not let them be Popperian testers!

This gets to the heart of my critique of what Kuhn says about theory testing. Before beginning that critique, I might warn the reader against a certain misconstrual of the relation of that critique to the account of experimental learning that I am after. My aim is not to provide an account of large-scale theory appraisal. (I am not looking to fulfill Wisdom's idea of turning Kuhn's rabbit [normal science] into Popper's duck. Nor do I want to turn Popper's duck into Kuhn's rabbit—which is Lakatos's way.) Theory appraisal and theory testing, as distinguished from theory choice and theory change, do turn on experimental knowledge grounded in normal testing, but that is not the main reason for acquiring such knowledge. In my view, scientific progress and growth is about the accumulation of experimental knowledge.[15]

Although experimental knowledge is not all there is to science, it holds the key to solving important philosophical problems about science. For instance, it is important to show why large-scale theory appraisal is objective and rational—in the very senses that Kuhn rejects. I think, however, that we should reject Kuhn's depiction of the problem of theory testing, namely, as a comparative appraisal of rival large-scale paradigm theories. Nevertheless, since I want to consider how our results about normal science force a revised view of Kuhn's own

15. An important reason Kuhn despaired of progress is his rejection of the Popperian syntactic measures of how theory change represents progress.

story about large-scale theory change, I am willing to be swept up in Kuhn's story a while longer.

2.7 Theory Choice versus Theory Appraisal: Gestalt Switches and All That

Why, having brought normal scientists to the crisis point, to the point of a warranted criticism of theory, will Kuhn still not let them be Popperian testers? (Granted I am being unclear about whether the theory at stake is medium-sized, large-scale, or a full disciplinary matrix, but that is because Kuhn is unclear.) According to Kuhn, the products of normal science are never going to be decisive for falsifying or for adjudicating between global theories. Testing and changing global theories or paradigms turns out not to be a matter of reasoned deliberation at all. Colorful passages abound in Kuhn's *Structure*. One such passage declares that

> the proponents of competing paradigms practice their trades in different worlds. . . . [T]he two groups of scientists see different things when they look from the same point in the same direction. . . . [B]efore they can hope to communicate fully, one group or the other must experience the conversion that we have been calling a paradigm shift. Just because it is a transition between incommensurables, the transition between competing paradigms cannot be made a step at a time, forced by logic and neutral experience. Like the gestalt switch, it must occur all at once (though not necessarily in an instant) or not at all. (Kuhn 1962, 149)

> What were ducks in the scientist's world before the revolution are rabbits afterwards. (Ibid., 110)

This picture of revolutionary science has been convincingly criticized by many authors (e.g., Laudan 1984b, 1990c; Scheffler 1982; Shapere 1984), and perhaps by now no further criticism is called for. But what I wish to consider, if only briefly, is how our recasting of normal science tells against Kuhn's view of global theory change.

Viewing global theory change as switching all elements of the paradigm, Kuhn supposes there to be no place to stand and scrutinize two whole paradigms, as a genuine paradigm test would require. Lacking an "empirically neutral system of language," Kuhn holds that "the proposed construction of alternate tests and theories must proceed from within one or another paradigm-based tradition" (1962, 145). This would be all right if it allowed that testing from within a paradigm could rest on something like our interparadigmatic canonical models of experiment. Failing to disentangle the experimental testing portion of the paradigm from immersion in its research program, Kuhn not surprisingly winds up viewing global theory change as arational—quite like the (experimentally) unwarranted critical discourse he attributes to nonsciences. It is as if the very process that allows practices to become scientific had shifted into reverse, until we are back to "mere" critical discourse:

> Critical discourse recurs only at moments of crisis when the bases of the field are again in jeopardy. Only when they must choose between competing theories do scientists behave like philosophers. That, I think, is why Sir Karl's brilliant description of the reasons for the choice between metaphysical systems so closely resembles my description of the reasons for choosing between scientific theories. (Pp. 6–7)

Indeed, the values Kuhn appeals to in theory change—simplicity, scope, fruitfulness, and the like—are precisely the criteria Popper claims we must resort to in appraising metaphysical systems.

The Circularity Thesis

Kuhn supposes that subscribers to competing global theories necessarily interpret and weigh these factors differently; hence, inevitably, one's own global theory gets defended. This *circularity thesis* is most clearly stated in *Structure*:

> Like the choice between competing political institutions, that between competing paradigms proves to be a choice between incompatible modes of community life. . . . When paradigms enter, as they must, into a debate about paradigm choice, their role is necessarily circular. Each group uses its own paradigm to argue in that paradigm's defense. (1962, 93)

> They will inevitably talk through each other when debating the relative merits of their respective paradigms. In the partially circular arguments that regularly result, each paradigm will be shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponent. (Ibid., 108–109)

While such circular defenses are possible, none of the requirements of paradigm theories, even as Kuhn conceives them, make their role in theory appraisal necessarily circular. On the contrary, there is much in what normal science requires that militates against such circularity—even in times of crisis. After all, "the criteria that [the paradigm] dictates for itself" are those in the experimental testing models and the exemplary arguments that go along with them. Those standards, if appropriate for their *own* goals, must condemn such question-begging arguments as failing utterly to probe a theory severely. At some point (i.e., with regard to some normal hypothesis), defending a global theory no matter what clashes with the requirements of normal testing. Such defenses can and do occur, but they do not count as warranted—by the strictures of good normal science. Why? Because they come down to a blanket refusal to acknowledge that a hypothesized solution to a normal problem fails, and that betrays an essential requirement of normal science.

Of course, nothing guarantees that actual science obeys the constraints of Kuhnian normal science. In fact, however, Kuhn's account of normal science is descriptively accurate for the bulk of important scientific episodes, episodes from which much has been learned. By and large, when these episodes appear to show otherwise it is due to mistaking as theory testing remarks that record individual biases expressed in hopes and fears, complaints, demands, name-calling, and bullying. Fortunately, the major and minor players in these cases discerned the difference. This is not surprising, given that they had been enjoying a normal science tradition. After all, and this is one of my main points, each such defense is just a hypothesized solution to a normal puzzle (and practitioners are well versed in scrutinizing alleged solutions).

To return to my criticism, let us sketch what happens, according to Kuhn's circularity thesis, when a global theory, T_1, slips into a crisis. Within T_1, which we assume to be a genuine science with a normal tradition and so on, genuine anomalies have been identified. These anomalies, especially if they are quantitative, identify genuine effects that need explaining. These give rise to normal puzzles, that is, normal testing, to scrutinize attempted solutions, to "set the situation right." The criteria of T_1, by dint of its enjoying a normal science tradition, severely constrain attempts to deal with such anomalies. A genuine crisis is afoot when, after considerable effort, T_1 is unable to explain away the anomalies as due either to initial conditions or background hypotheses.

All these points of Kuhn's story have already been told. Notice that it follows that *the experimental testing criteria of T_1 themselves warrant the existence of anomalies and crisis.* Incorporating as they must the general criteria of normal testing, they indicate when an anomaly really is unevadable (i.e., when to put blame elsewhere is tantamount to unwarranted criticism). Do they not, by the same token, indicate that any attempt to save a theory—if that defense depends upon evading the anomaly—violates the very norms upon which enjoying a normal "puzzle-solving" tradition depends? The norms would bar procedures of admitting hypotheses as solutions to puzzles, we saw, if they often would do so erroneously.

Of course, it may take a while until attempted defenses come up against the wall of normal testing strictures. But with a genuine crisis, it seems, that is exactly what happens. Moreover, from Kuhn's demarcation criterion, *it is possible to recognize* (even if not sharply) that a practice is losing its normal puzzle-solving ability. (Astrology is a kind of exemplar of a practice that falls over onto the nonscience line.)

These remarks should not be misunderstood. What normal science must condemn is not saving a global theory in the face of severe anomaly—although that is what Popperian spectacles might have us see. What it must condemn (recalling Kuhn's demarcation) is being prevented from *learning* from normal testing. In any particular case, the obstacles to learning that are condemned are very specific: having to reject experimentally demonstrated effects, contradict known parameter values, change known error distributions of instruments or background factors, and so on. Such moves are not always learning obstacles, but they are when they fly in the face of exemplary experimental models.

Consider what Kuhn calls for when scientists, having split off from global theory T_1 to develop some rival T_2, come knocking on the door of their less adventurous colleagues, who are still muddling through the crisis in T_1. Confronted with rival T_2, which, let's suppose, solves T_1's crisis-provoking problem, crisis scientists in T_1 necessarily defend T_1 circularly. This circularity thesis requires them to do a turnabout and maintain that T_1 will eventually solve this problem, or that the problem was not really very important after all. Once the members of rival T_2 go away (back to their own worlds, presumably), members of T_1 can resume their brooding about the crisis they have identified with their paradigm. Were they to do this, they would indeed be guilty of the unwarranted criticism and mere name-calling Kuhn finds typical of nonsciences. But Kuhn has given no argument to suppose that crisis scientists necessarily do this.

Even if the circularity thesis is rejected, as it should be, many of the more troubling allegations Kuhn raises about global paradigm appraisal persist. These follow, however, not from the high points of normal testing and crisis, but from two additional premises that we should also reject: first, that acceptance (rejection) of a paradigm is the same as or indistinguishable from taking up (stopping work on) its problems, and second, that global theory or paradigm change requires a conversion experience. Taken together, these assumptions render the problem of global theory change as the problem of what causes practitioners working in one paradigm to transfer their allegiance and become converted to another. Seen this way, it is no wonder Kuhn views the answer to be a matter of sociopsychology.

Discrediting a Global Theory versus Stopping Work on It

It would seem that the repeated anomalies that are supposed to bring on crisis provide grounds for thinking that the theory has got it wrong, at least for the anomalous area. Yet Kuhn denies that scientists take even the most well-warranted crisis as grounds to falsify or reject the global theory involved (even as a poor problem solver). For Kuhn, "once it has achieved the status of paradigm, a scientific theory is declared invalid only if an alternate candidate is available to take its place. . . . The decision to reject one paradigm is always simultaneously the decision to accept another" (Kuhn 1962, 77).

This thesis of comparative appraisal, for Kuhn, is not a result of an analysis of warranted epistemic appraisal. It is a consequence of Kuhn's first assumption—equating the acceptance of a global theory with pursuit of its research program. That is why Kuhn denies that scientists can reject a theory without adopting another: "They could not do so and still remain scientists" (1962, 78). Together with his assumption that science has to be done within a single paradigm, the comparative appraisal thesis follows logically. But as we have already said, working within a paradigm—pursuing normal problems and tests—is wholly distinct from according it certain epistemic credentials. A situation that through Kuhnian spectacles appears as "still working within theory *T*, despite severe crisis" may actually be one where *T* is discredited (e.g., key normal hypotheses found false), but not yet replaced.[16] Discrediting a theory is not the same as stopping work on it.

Good reasons abound for still working on theory *T*, despite anomalies. Nor need these reasons go away even if *T* is replaced. Different types of anomalies must be distinguished. Suppose that the anomaly truly indicts a hypothesis of the theory *T*, say hypothesis *H*; that is, "*H* is in error" passes reliable tests. Consider two cases:

a. Theory *T* at the moment has got it wrong so far as *H* goes, but it is the kind of problem we know a fair amount about clearing up. Here the anomaly counts against neither the correctness of *T* nor the value

16. Even when *T* is replaced, it may very well still enjoy a puzzle-solving tradition. If I am correct, this counts against Kuhn's claim that theories are replaced only when they fail to have a puzzle-solving tradition.

of pursuing T further. Indeed, it is likely to be a fruitful means of modifying or replacing H.

b. In a genuine crisis situation, theory T is found to have got it wrong so far as several key hypotheses go. Here the anomalies discredit the full correctness of T but not the value of continued work on T. Kuhn's idea of "the creative function of crises" should be taken seriously. The anomalies are extremely rich sources of better hypotheses and better normal tests, as well as better theories. After all, whatever passes severe tests, that is, experimental effects—including the anomalies themselves—are things for which any new theory should account: they do not go away.[17]

One reason the thesis about comparative testing appears plausible, even to philosophers who otherwise take issue with Kuhn, is that often the clearest grounds for discrediting a theory arise when a rival, T_2, is at hand. That is because the rival often supplies the experimental grounds for using anomalies to show that T_1 is genuinely in crisis. But the argument, whether it comes in the course of working on T_1 or T_2, must be made out in the experimental testing framework of T_1. More correctly, it must be made out by means of shared canonical models of experiment. Such an argument, where warranted, does not depend on already holding T_2.

An example, to be discussed more fully later, may clarify my point. The two theories are classical thermodynamics and the molecular-kinetic theory. Jean Perrin's experiments, while occurring within the molecular-kinetic account, demonstrated an unevadable anomaly for the classical account by showing that Brownian motion violates a nonstatistical version of the second law of thermodynamics. They did so by showing that Brownian movement exemplifies the type of random phenomenon known from simple games of chance. The canonical model here comes from random walk phenomena. It was well understood and did not belong to any one paradigm. Once the applicability of the canonical model to Brownian motion data was shown, the anomaly, which was quantitative, was unevadable.[18]

17. An important task for the experimental program I am promoting would be to identify canonical ways of deliberately learning from anomalies. Contributions from several sources would be relevant. One source would be some of the empirical work in cognitive science, such as Lindley Darden 1991.

18. Those seeking to save a nonstatistical account were not allowed to explain away the anomaly or defend their theory circularly. Once all their attempted explanations were shown wanting—on normal experimental grounds—they had to concede. It is not that they are bound by a sociological convention—doing otherwise violates canons of learning from experiment. I discuss this case in detail in chapter 7.

I do not assert that experimental arguments always exist to guide theory appraisal, but rather deny Kuhn's claim that they never do. Moreover, for experimental arguments to ground theory appraisal, the experimental testing frameworks of the rival large-scale theories need not be identical. It is sufficient for the needed arguments to be made out by appeal to the interparadigmatic canonical experimental models. How can we suppose such a shared understanding? It follows from taking seriously the criteria for good normal scientific practice, criteria that, for Kuhn, must hold for any practice that enjoys a normal scientific tradition. Moreover, the historical record reveals case after case where even the most ardent proponents are forced to relent on the basis of very local but very powerful experimental tests. The Kuhn of normal science can explain this consensus quite naturally; the Kuhn of revolutionary science cannot.

One is justly led to wonder why Kuhn holds to the curious position that the strictures of normal science can compel rejection of hypotheses within a large-scale theory, even to the extent of provoking a crisis, while supposing that when the crisis gets too serious or a rival theory is proposed, the normal practitioner abruptly throws the strictures of normal science out the window and declares (being reduced to aesthetic criteria now) that his or her theory is the most beautiful. One will look in vain for an argument for this position as well as for an argument about why Kuhn takes away what I thought he had promised us—that a crisis compelled by good normal science lets us finally be *warranted* Popperian testers and reject the theory (as being wrong at least so far as its key hypotheses go)—quite apart from stopping work on it. Instead one finds that, when turning his gaze to the problem of large-scale theory appraisal, Kuhn is simply wearing spectacles that necessarily overlook the role of the shared strictures and arguments of normal testing.

Kuhn seeks something that can effect a transfer of allegiance and the gestalt switch that allegedly goes with it. Impressive experimental demonstrations can at most pave the way for this conversion by making a scientist's mind susceptible to the new gestalt. And unlike the gestalt switch of psychology, the scientist cannot switch back and forth to compare global theories.

The Early Innovators

Not everyone switches at the same time, which is one reason why Kuhn supposes that there is no single argument that must rationally convince everyone. We should be glad of this, Kuhn maintains, because such innovations are generally mistaken!

> Most judgments that a theory has ceased adequately to support a puzzle-solving tradition prove to be wrong. If everyone agreed in such judgements, no one would be left to show how existing theory could account for the apparent anomaly as it usually does. (P. 248)

But this has a curious consequence for Kuhn. The innovators, with their daring value systems, switch early and proceed to work within the new theory T_2 (never mind how to explain their converting together). Where are they, it might be asked, while they are developing the new theory? Presumably, for Kuhn, they must be within T_2, since working within T_2 is equated with accepting it. It would seem to follow, however, that the early innovators would have to convert back to T_1 when, as happens most of the time pace Kuhn, the innovation is mistaken. Yet this, according to Kuhn, is impossible, or nearly always so.

Whichever paradigm theory they find themselves caught in, the early innovators must still be employing experimental testing tools from the earlier paradigm (in which we include the general pool of canonical models) to test their new hypotheses. Otherwise they could not demonstrate the quantitative experimental successes that Kuhn's own spectacles reveal to be central (if not determinative) in paradigm appraisal. Although their divergent paradigms result in their speaking different languages where translation is at most partial, they can learn how they differ via experiment. Kuhn himself says, "First and foremost, men experiencing communication breakdown can discover by experiment—sometimes by thought-experiment, armchair science—the area within which it occurs" (p. 277).

A Kuhnian may agree with my thesis about shared testing models, yet deny that the experimental arguments provided offer a basis for appraising global theories. Nevertheless, that is still no argument for Kuhn's thesis that global theory change cannot turn on experimental arguments, and, indeed, Kuhn fails to supply one. Rather, his thesis results from assuming that theory change is a conversion experience, that it requires one to "go native," and is complete only when the new theory establishes a grip on one's mind.

We can hold, with Kuhn, that experimental demonstrations and arguments at most allow one to accept a rival theory "intellectually" and yet reach the conclusion opposite from his. Far from downplaying the role of experimental argument, this just shows why scientific appraisal properly turns only on (epistemologically grounded) "intellectual" acceptance and not on psychological conversions. That evidential arguments are incapable of grounding theory change when defined as mind shifts is precisely why mind shifts have nothing to do with grounding theory assessment in science.

It is instructive to consider the new technology of "virtual reality" machines. Fitted with the appropriate helmet and apparatus, one can enter a 3-d world to learn history, medical procedures, and more. In the future a virtual reality program created by the members of one paradigm might well allow scientists from another paradigm to vicariously experience the world through the other's eyes (without the risk of being unable to convert back). This might even prove to heighten the capacity to find solutions to problems set by a research program. However, the ability to tell someone to "get into the machine and see for yourself" will never be an argument, will never substitute for an evidential grounding of the theory thereby "lived in."

2.8 Summary and Closing Remarks

We began by asking what philosophical mileage could be gotten from exploring Kuhn's contrast of his position with Popper's. How far have we gone and how much of it will be utilized in the project of this book?

On Kuhn's treatment of normal science, we can sensibly construe his comparison with Popper—and it turns out that Kuhn is correct. Normal scientists, in my rereading of Kuhn, have special requirements without which they could not learn from standard tests. They insist on stringent tests, reliable or severe. They could not learn from failed solutions to normal problems if they could always change the question, make alterations, and so on. That is what Kuhn says. That is what having a normal science tradition is all about. But then we have some curious consequences at the level of theory appraisal.

Via the criteria of normal science, Kuhn says, normal science may be led to crisis. It is recognized as crisis because of the stringency of its rules. Suddenly, when confronted with a rival theory, Kuhn says, normal scientists do an about-face, defending their theory and denying it is in crisis. Kuhn gives no argument for supposing this always happens, in fact, my point is that his view of normal science militates against this supposition.

Answering Kuhn does not require us to show that global theory testing is always a function of experimental knowledge, but merely that we deny the Kuhnian view that it cannot be. My solution is based on one thing normal practitioners, even from rival paradigms, have in common (by dint of enjoying a normal testing tradition): they can and do perform the tasks of normal science reliably. That is the thrust of

Kuhn's demarcation criterion. Later we will see how experimental knowledge functions in theory testing.

The problems with Kuhn's account of theory appraisal, however, are not the problems that my approach requires be overcome. I do not seek an account of the comparative testing of rival large-scale theories, as I deny that such a thing occurs (except as understood in an elliptical fashion, to be explained). I do not accept Kuhn's supposition that there are two kinds of empirical scientific activities, normal and revolutionary: there is just normal science, understood as standard testing.[19]

Nevertheless, I will retain several of the key theses I have gleaned from Kuhn's comparison with Popper: Taking Popperian aim at global theories when doing so is not constrained by severe normal testing is a poor strategy for obtaining experimental knowledge. The constraints and norms of normal testing provide the basis for severe tests and informative scientific inquiries. To understand the nature and growth of experimental knowledge, one must look to normal testing.

We can also retain a version of Kuhn's demarcation criterion. The relevant distinction, although it is not intended to be sharp and may well admit of degrees, is between inquiries that are scientific or informative and those that are not. Inquiries are informative to the extent that they enable experimental knowledge, that is, learning from normal science. For Kuhn, in a genuine science, anomalies give rise to research puzzles. In our recasting of Kuhn this translates as, in a genuinely scientific inquiry, anomalies afford opportunities for learning. This learning is tantamount to learning from error, as described in chapter 1 and in what follows. The aim of science is not avoiding anomaly and error but being able to learn from anomaly and error.

Finding things out is a lot like normal science being revisited in the manner discussed in this chapter. Nevertheless, even the idea of normal science as extending and filling in theories should be questioned. Although it is not too far from a description of what scientists generally do, it does not entirely capture knowledge at the forefront—it is still too tied to a theory-dominated way of thinking. In the gathering up of knowledge, it is typical not to know which fields will be called upon to solve problems. There need not even be a stable background theory in place.

Take, for example, recent work on Alzheimer's disease. Clumps of an insoluble substance called beta amyloid have been found in the brains of its victims, which presents problems in its relation to the dis-

19. By "empirical scientific activities" I am referring here to experimental activities in the broad sense in which I understand this.

ease and how it builds up in the brain. But what is the background theory or paradigm being extended? It could come from biology or neuroscience, from any one of their specialties—except that recent findings suggest that the solution may come from genetics.

The growth of knowledge, by and large, has to do not with replacing or amending some well-confirmed theory, but with testing specific hypotheses in such a way that there is a good chance of learning something—whatever theory it winds up as part of. Having divorced normal (standard) testing from the Kuhnian dependence upon background paradigms in any sense other than dependence upon an intertheoretic pool of exemplary models of error, it is easy to accommodate a more realistic and less theory-dominated picture of inquiry. In much of day-to-day scientific practice, and in the startling new discoveries we read about, scientists are just trying to find things out.