Falsification and the Methodology of Scientific Research Programmes[1]

Imre Lakatos

SCIENCE: REASON OR RELIGION?

For centuries knowledge meant proven knowledge—proven either by the power of the intellect or by the evidence of the senses. Wisdom and intellectual integrity demanded that one must desist from unproven utterances and minimize, even in thought, the gap between speculation and established knowledge. The proving power of the intellect or the senses was questioned by the sceptics more than two thousand years ago; but they were browbeaten into confusion by the glory of Newtonian physics. Einstein's results again turned the tables and now very few philosophers or scientists still think that scientific knowledge is, or can be, proven knowledge. But few realize that with this the whole classical structure of intellectual values falls in ruins and has to be replaced: One cannot simply water down the ideal of proven truth—as some logical empiricists do—to the ideal of "probable truth"[2] or—as some sociologists of knowledge do—to "truth by [changing] consensus."[3] . . .

[1] This paper is a considerably improved version of my [1968*b*] and a crude version of my [1970]. Some parts of the former are here reproduced without change with the permission of the Editor of the *Proceedings of the Aristotelian Society*. In the preparation of the new version I received much help from Tad Beckman, Colin Howson, Clive Kilmister, Larry Laudan, Eliot Leader, Alan Musgrave, Michael Sukale, John Watkins and John Worrall.

[2] The main contemporary protagonist of the ideal of "probable truth" is Rudolf Carnap. For the historical background and a criticism of this position, cf. Lakatos [1968*a*].

[3] The main contemporary protagonists of the ideal of "truth by consensus" are Polanyi and Kuhn. For the historical background and a criticism of this position, cf. Musgrave [1969a], Musgrave [1969*b*] and Lakatos [1970].

FALLIBILISM VERSUS FALSIFICATIONISM

Dogmatic (or Naturalistic) Falsificationism. The Empirical Basis

. . . *According to the "justificationists" scientific knowledge consisted of proven propositions.* Having recognized that strictly logical deductions enable us only to infer (transmit truth) but not to prove (establish truth), they disagreed about the nature of those propositions (axioms) whose truth can be proved by extralogical means. *Classical intellectualists* (or "rationalists" in the narrow sense of the term) admitted very varied—and powerful—sorts of extralogical "proofs" by revelation, intellectual intuition, experience. These, with the help of logic, enabled them to prove every sort of scientific proposition. *Classical empiricists* accepted as axioms only a relatively small set of "factual propositions" which expressed the "hard facts." Their truth-value was established by experience and they constituted the *empirical basis* of science. In order to prove scientific *theories* from nothing else but the narrow empirical basis, they needed a logic much more powerful than the deductive logic of the classical intellectualists: "*inductive logic.*" All justificationists, whether intellectualists or empiricists, agreed that a singular statement expressing a "hard fact" may *disprove* a universal theory; . . . but few of them thought that a finite conjunction of factual propositions might be sufficient to *prove* "inductively" a universal theory.[4]

[4] Indeed, even some of these few shifted, following Mill, the rather obviously insoluble problem of inductive proof (of uni-

From Sections 1, 2, and 3 of "Falsification and the Methodology of Scientific Research Programmes" by Imre Lakatos, in *Criticism and the Growth of Knowledge*, Imre Lakatos and Alan Musgrave, Eds. (1970), pp. 91–92, 94–125, 127–138, 189–195. Reprinted by permission of Cambridge University Press.

Justificationism, that is, the identification of knowledge with proven knowledge, was the dominant tradition in rational thought throughout the ages. Scepticism did not deny justificationism: It only claimed that there was (and could be) no proven knowledge and *therefore* no knowledge whatsoever. For the sceptics "knowledge" was nothing but animal belief. Thus justificationist scepticism ridiculed objective thought and opened the door to irrationalism, mysticism, superstition.

This situation explains the enormous effort invested by classical rationalists in trying to save the synthetical *a priori* principles of intellectualism and by classical empiricists in trying to save the certainty of an empirical basis and the validity of inductive inference. For all of them *scientific honesty demanded that one assert nothing that is unproven*. However, both were defeated: Kantians by non-Euclidean geometry and by non-Newtonian physics, and empiricists by the logical impossibility of establishing an empirical basis (as Kantians pointed out, facts cannot prove propositions) and of establishing an inductive logic (no logic can infallibly increase content). It turned out that *all theories are equally unprovable*.

Philosophers were slow to recognize this, for obvious reasons: Classical justificationists feared that once they conceded that theoretical science is unprovable, they would have also to concede that it is sophistry and illusion, a dishonest fraud. The philosophical importance of *probabilism* (or *"neojustificationism"*) lies in the denial that such a concession is necessary.

Probabilism was elaborated by a group of Cambridge philosophers who thought that although scientific theories are equally unprovable, they have different degrees of probability (in the sense of the calculus of probability) relative to the available empirical evidence.[5] *Scientific honesty then requires less than had been thought: It consists in uttering only highly probable theories; or even in merely specifying, for each scientific theory, the evidence, and the probability of the theory in the light of this evidence.*

Of course, replacing proof by probability was a major retreat for justificationist thought. But even this retreat turned out to be insufficient. It was soon shown, mainly by Popper's persistent efforts, that under very general conditions all theories have zero probability, whatever the evidence; *all theories are not only equally unprovable but also equally improbable*.[6]

Many philosophers still argue that the failure to obtain at least a probabilistic solution of the problem of induction means that we "throw over almost everything that is regarded as knowledge by science and common sense."[7] It is against this background that one must appreciate the dramatic change brought about by falsificationism in evaluating theories and, in general, in the standards of intellectual honesty. Falsificationism was, in a sense, a new and considerable retreat for rational thought. But since it was a retreat from utopian standards, it cleared away much hypocrisy and muddled thought, and thus, in fact, it represented an advance.

First I shall discuss a most important brand of falsificationism: dogmatic (or "naturalistic")[8] falsificationism. Dogmatic falsificationism admits the fallibility of *all* scientific theories without qualification, but it retains a sort of infallible empirical basis. It is strictly empiricist without being inductivist: It denies that the certainty of the empirical basis can be transmitted to theories. *Thus dogmatic falsificationism is the weakest brand of justificationism.*

It is extremely important to stress that admitting [fortified] empirical counterevidence as a final arbiter against a theory does not make one a dogmatic falsificationist. Any Kantian or inductivist will agree to such arbitration. But both the Kantian

versal from particular propositions) to the slightly less obviously insoluble problem of proving *particular* factual propositions from other *particular* factual propositions.

[5] The founding fathers of probabilism were intellectualists; Carnap's later efforts to build up an empiricist brand of probabilism failed. Cf. my [1968*a*], p. 367 and also p. 361, footnote 2.

[6] For a detailed discussion, cf. my [1968*a*], especially pp. 353 ff.

[7] Russell [1943], p. 683. For a discussion of Russell's justificationism, cf. my [1962], especially pp. 167 ff.

[8] For the explanation of this term, cf. *below*, footnote 14.

and the inductivist, while bowing to a negative crucial experiment, will also specify conditions of how to establish, entrench one unrefuted theory more than another. Kantians held that Euclidean geometry and Newtonian mechanics were established with certainty; inductivists held they had probability I. For the dogmatic falsificationist, however, empirical *counter*evidence is the *one and only* arbiter which may judge a theory.

The hallmark of dogmatic falsificationism is then the recognition that all theories are equally conjectural. Science cannot *prove* any theory. But although science cannot *prove*, it can *disprove:* It "can perform with complete logical certainty [the act of] repudiation of what is false,"[9] that is, there is an absolutely firm empirical basis of facts which can be used to disprove theories. Falsificationists provide new—very modest—standards of scientific honesty: They are willing to regard a proposition as "scientific" not only if it is a proven factual proposition, but even if it is nothing more than a falsifiable one, that is, if there are factual propositions available at the time with which it may clash, or, in other words, if it has potential falsifiers.[10]

Scientific honesty then consists of specifying, in advance, an experiment such that if the result contradicts the theory, the theory has to be given up.[11] The falsificationist demands that once a proposition is disproved, there must be no prevarication: The proposition must be unconditionally rejected. To (non-tautologous) unfalsifiable propositions the dogmatic falsificationist gives short shrift: He brands them "metaphysical" and denies them scientific standing.

Dogmatic falsificationists draw a sharp demarcation between the theoretician and the experimenter: The theoretician proposes, the experimenter—in the name of Nature—disposes. As Weyl put it: "I wish to record my unbounded admiration for the work of the experimenter in his struggle to wrest interpretable facts from an unyielding Nature who knows so well how to meet our theories with a decisive *No*—or with an unaudible *Yes*."[12] Braithwaite gives a particularly lucid exposition of dogmatic falsificationism. He raises the problem of the objectivity of science: "To what extent, then, should an established scientific deductive system be regarded as a free creation of the human mind, and to what extent should it be regarded as giving an objective account of the facts of nature?" His answer is: "The form of a statement of a scientific hypothesis, and its use to express a general proposition, is a human device; what is due to Nature are the observable facts which refute or fail to refute the scientific hypothesis . . . [In science] we hand over to Nature the task of deciding whether any of the contingent lowest-level conclusions are false. This objective test of falsity it is which makes the deductive system, in whose construction we have very great freedom, a deductive system of scientific hypotheses. Man proposes a system of hypotheses: Nature disposes of its truth or falsity. Man invents a scientific system, and then discovers whether or not it accords with observed fact."[13]

According to the logic of dogmatic falsificationism, science grows by repeated overthrow of theories with the help of hard facts. For instance, according to this view, Descartes's vortex theory of gravity was refuted—and eliminated—by the *fact* that planets moved in ellipses rather than in Cartesian circles; Newton's theory, however, explained successfully the then available facts, both those which had been explained by Descartes's theory and those which refuted it. Therefore Newton's theory replaced Descartes's theory. Analo-

[9] Medawar [1967], p. 144.

[10] This discussion already indicates the vital importance of a demarcation between provable factual and unprovable theoretical propositions for the dogmatic falsificationist.

[11] "*Criteria of refutation* have to be laid down beforehand: It must be agreed which observable situations, if actually observed, mean that the theory is refuted" (Popper [1963], p. 38, footnote 3).

[12] Quoted in Popper [1934], Section 85, with Popper's comment: "I fully agree."

[13] Braithwaite [1953], pp. 367–8. For the "incorrigibility" of Braithwaite's observed facts, cf. his [1938]. While in the quoted passage Braithwaite gives a forceful answer to the problem of scientific objectivity, in another passage he points out that "except for the straightforward generalizations of observable facts . . . complete refutation is no more possible than is complete proof" ([1953], p. 19). . . .

gously, as seen by falsificationists, Newton's theory was, in turn, refuted—proved false—by the anomalous perihelion of Mercury, while Einstein's explained that too. Thus science proceeds by bold speculations, which are never proved or even made probable, but some of which are later eliminated by hard, conclusive refutations and then replaced by still bolder, new and, at least at the start, unrefuted speculations.

Dogmatic falsificationism, however, is untenable. It rests on two false assumptions and on a too narrow criterion of demarcation between scientific and non-scientific.

The *first assumption* is that there is a natural, *psychological* borderline between theoretical or speculative propositions on the one hand and factual or observational (or basic) propositions on the other. (I shall call this—following Popper—the *naturalistic doctrine of observation.*)

The *second assumption* is that if a proposition satisfies the psychological criterion of being factual or observational (or basic) then it is true; one may say that it was *proved* from facts. (I shall call this the *doctrine of observational (or experimental) proof.*)[14]

These two assumptions secure for the dogmatic falsificationist's deadly disproofs an empirical basis from which proven falsehood can be carried by deductive logic to the theory under test.

These assumptions are complemented by a *demarcation criterion:* Only those theories are "scientific" which forbid certain observable states of affairs and therefore are factually disprovable. *Or, a theory is "scientific" is it has an empirical basis.*[15]

But both assumptions are false. Psychology testifies against the first, logic against the second, and, finally, methodological judgment testifies against the demarcation criterion. I shall discuss them in turn.

1. A first glance at a few characteristic examples already undermines the *first assumption.* Galileo claimed that he could "observe" mountains on the moon and spots on the sun and that these "observations" refuted the time-honoured theory that celestial bodies are faultless crystal balls. But his "observations" were not "observational" in the sense of being observed by the—unaided—senses: Their reliability depended on the reliability of his telescope—and of the optical theory of the telescope—which was violently questioned by his contemporaries. It was not Galileo's—pure, untheoretical—*observations* that confronted Aristotelian *theory* but rather Galileo's "observations" in the light of his optical theory that confronted the Aristotelians' "observations" in the light of their theory of the heavens. . . . This leaves us with two inconsistent theories, *prima facia* on a par. Some empiricists may concede this point and agree that Galileo's "observations" were not genuine observations; but they still hold that there is a "natural demarcation" between statements impressed on an empty and passive mind directly by the senses—only these constitute genuine "immediate knowledge"—and between statements which are suggested by impure, theory-impregnated sensations. . . . But it transpires from the work of Kant and Popper . . . that . . . there are and can be no sensations unimpregnated by expectations and therefore *there is no natural (i.e., psychological) demarcation between observational and theoretical propositions. . . .*

2. But even if there was such a natural demarcation, logic would still destroy the *second assumption* of dogmatic falsificationism. For the truth-value of the "observational" propositions cannot be indubitably decided: *No factual proposition can ever be proved from an*

[14] For these assumptions and their criticism, cf. Popper [1934], Sections 4 and 10. It is because of this assumption that—following Popper—I call this brand of falsificationism "naturalistic." Popper's "basic propositions" should not be confused with the basic propositions discussed in this section; cf. *below,* footnote 35.

It is important to point out that these two assumptions are also shared by many justificationists who are not falsificationists: They may add to experimental proofs "intuitive proofs"—as did Kant—or "inductive proofs"—as did Mill. Our falsificationist accepts experimental proofs *only.*

[15] The empirical basis of a theory is the set of its potential falsifiers: the set of those observational propositions which may disprove it.

experiment. Propositions can only be derived from other propositions, they cannot be derived from facts: One cannot prove statements from experiences—"no more than by thumping the table."[16] This is one of the basic points of elementary logic, but one which is understood by relatively few people even today.[17]

If factual propositions are unprovable, then they are fallible. If they are fallible, then clashes between theories and factual propositions are not "falsifications" but merely inconsistencies. Our imagination may play a greater role in the formulation of "theories" than in the formulation of "factual propositions",[18] but they are both fallible. Thus *we cannot prove theories and we cannot disprove them either.*[19] The demarcation between the soft, unproven "theories" and the hard, proven "empirical basis" is nonexistent: *All* propositions of science are theoretical and, incurably, fallible. . . .

3. Finally, even if there were a natural demarcation between observation statements and theories, and even if the truth-value of observation statements could be indubitably established, dogmatic falsificationism would still be useless for eliminating the most important class of what are commonly regarded as scientific theories. For even if experiments *could* prove experimental reports, their disproving power would still be miserably restricted: *Exactly the most admired scientific theories simply fail to forbid any observable state of affairs.*

To support this last contention, I shall first tell a characteristic story and then propose a general argument.

The story is about an imaginary case of planetary misbehaviour. A physicist of the pre-Einsteinian era takes Newton's mechanics and his law of gravitation (*N*), the accepted initial conditions, *I*, and calculates, with their help, the path of a newly discovered small planet, *p*. But the planet deviates from the calculated path. Does our Newtonian physicist consider that the deviation was forbidden by Newton's theory and therefore that, once established, it refutes the theory *N*? No. He suggests that there must be a hitherto unknown planet *p'* which perturbs the path of *p*. He calculates the mass, orbit, etc., of this hypothetical planet and then asks an experimental astronomer to test his hypothesis. The planet *p'* is so small that even the biggest available telescopes cannot possibly observe it: The experimental astronomer applies for a research grant to build yet a bigger one.[20] In three years' time the new telescope is ready. Were the unknown planet *p'* to be discovered, it would be hailed as a new victory of Newtonian science. But it is not. Does our scientist abandon Newton's theory and his idea of the perturbing planet? No. He suggests that a cloud of cosmic dust hides the planet from us. He calculates the location and properties of this cloud and asks for a research grant to send up a satellite to test his calculations. Were the satellite's instruments (possibly new ones, based on a little-tested theory) to record the existence of the conjectural cloud, the result would be hailed as an outstanding victory for Newtonian science. But the cloud is not found. Does our scientist abandon Newton's theory, together with the idea of the perturbing planet and the idea of the cloud which hides it? No. He sug-

---

[16] Cf. Popper [1934], Section 29.

[17] It seems that the first philosopher to emphasize this might have been Fries in 1837 (cf. Popper [1934], Section 29, footnote 3). This is of course a special case of the general thesis that logical relations, like probability or consistency, refer to *propositions.* Thus, for instance, the proposition "nature is consistent" is false (or, if you wish, meaningless), for nature is not a proposition (or a conjunction of propositions).

[18] Incidentally, even this is questionable. Cf. below, pp. 188ff.

[19] As Popper put it, "No conclusive disproof of a theory can ever be produced"; those who wait for an infallible disproof before eliminating a theory will have to wait forever and "will never benefit from experience" ([1934], Section 9).

[20] If the tiny conjectural planet were out of the reach even of the biggest *possible* optical telescopes, he might try some quite novel instrument (like a radiotelescope) in order to enable him to "observe it," that is, to ask Nature about it, even if only indirectly. (The new "observational" theory may itself not be properly articulated, let alone severely tested, but he would care no more than Galileo did.)

gests that there is some magnetic field in that region of the universe which disturbed the instruments of the satellite. A new satellite is sent up. Were the magnetic field to be found, Newtonians would celebrate a sensational victory. But it is not. Is this regarded as a refutation of Newtonian science? No. Either yet another ingenious auxiliary hypothesis is proposed or . . . the whole story is buried in the dusty volumes of periodicals and the story never mentioned again.[21]

This story strongly suggests that even a most respected scientific theory, like Newton's dynamics and theory of gravitation, may fail to forbid any observable state of affairs.[22] Indeed, *some scientific theories forbid an event occurring in some specified finite spatio-temporal region (or briefly, a "singular event") only on the condition that no other factor* (possibly hidden in some distant and unspecified spatio-temporal corner of the universe) *has any influence on it.* But then *such theories never alone contradict a "basic" statement:* They contradict at most a conjunction of a basic statement describing a spatio-temporally singular event and of a universal non-existence statement saying that no other relevant cause is at work anywhere in the universe. And the dogmatic falsificationist cannot possibly claim that such universal non-existence statements belong to the empirical basis—that they can be observed and proved by experience.

Another way of putting this is to say that some scientific theories are normally interpreted as containing a *ceteris paribus* clause[23]: in such cases it is always a specific theory *together* with this clause which may be refuted. But such a refutation is inconsequential for the *specific* theory under test because by replacing the *ceteris paribus* clause by a different one the *specific* theory can always be retained whatever the tests say.

If so, the "inexorable" disproof procedure of dogmatic falsificationism breaks down in these cases *even if* there were a firmly established empirical basis to serve as a launching pad for the arrow of the *modus tollens:* The prime target remains hopelessly elusive.[24] And as it happens, it is exactly the most important, "mature" theories in the history of science which are *prima facie* undisprovable in this way.[25] Moreover, by the standards of dogmatic falsificationism all probabilistic theories also come under this head, for no finite sample can ever *disprove* a universal probabilistic theory[26]; probabilistic theories, like theories with a *ceteris paribus* clause, have no empirical basis. But then the dogmatic falsificationist relegates the most important scientific theories *on his own admission* to metaphysics where rational discussion—consisting, by his standards, of proofs and disproofs—has no place, since a metaphysical theory is neither provable nor disprovable. The demarcation criterion of dogmatic falsificationism is thus still strongly antitheoretical.

(Moreover, *one can easily argue that* ceteris paribus *clauses are not exceptions, but the rule in science.* Science, after all, must be demarcated from a curiosity shop where funny local—or cosmic—oddities are collected and displayed. The assertion that "all Britons died from lung cancer between 1950 and 1960" is logically possible and might even have been true. But if it has been only an occurrence of an event with minute probability, it would have only curiosity value for the crankish fact-collector; it would have a macabre

---

[21] At least not until a new research program supersedes Newton's programme which happens to explain this previously recalcitrant phenomenon. In this case, the phenomenon will be unearthed and enthroned as a "crucial experiment". . . .

[22] Popper asks, "What kind of clinical responses would refute to the satisfaction of the analyst not merely a particular diagnosis but psychoanalysis itself?" ([1963], p. 38, footnote 3.) But what kind of observation would refute to the satisfaction of the Newtonian not merely a particular version but Newtonian theory itself?

[23] This *ceteris paribus* clause must not normally be interpreted as a separate premise. . . .

[24] Incidentally, we might persuade the dogmatic falsificationist that his demarcation criterion was a very naive mistake. If he gives it up but retains his two basic assumptions, he will have to ban theories from science and regard the growth of science as an accumulation of proven basic statements. This indeed is the final stage of classical empiricism after the evaporation of the hope that facts can prove or at least disprove theories.

[25] This is no coincidence. . . .

[26] Cf. Popper [1934], Chapter VIII.

entertainment value, but no scientific value. A proposition might be said to be scientific only if it aims at expressing a causal connection: Such connection between being a Briton and dying of lung cancer may not even be intended. Similarly, "All swans are white," if true, would be a mere curiosity unless it asserted that swanness *causes* whiteness. But then a black swan would not refute this proposition, since it may only indicate *other causes* operating simultaneously. Thus "all swans are white" is either an oddity and easily disprovable or a scientific proposition with a *ceteris paribus* clause and therefore undisprovable. *Tenacity of a theory against empirical evidence would then be an argument for rather than against regarding it as "scientific." "Irrefutability" would become a hallmark of science.)* . . .

To sum up: Classical justificationists only admitted proven theories; neoclassical justificationists probable ones; dogmatic falsificationists realized that in either case no theories are admissible. They decided to admit theories if they are disprovable—disprovable by a finite number of observations. But even if there were such disprovable theories—those which can be contradicted by a finite number of observable facts—they are still logically too near to the empirical basis. For instance, on the terms of the dogmatic falsificationist, a theory like "All planets move in ellipses" may be disproved by five observations; therefore the dogmatic falsificationist will regard it as scientific. A theory like "All planets move in circles" may be disproved by four observations; therefore the dogmatic falsificationist will regard it as still more scientific. The acme of scientificness will be a theory like "All swans are white," which is disprovable by one single observation. On the other hand, he will reject all probabilistic theories together with Newton's, Maxwell's, Einstein's theories, as unscientific, for no finite number of observations can ever disprove them.

If we accept the demarcation criterion of dogmatic falsificationism, *and* also the idea that facts can prove "factual" propositions, we have to declare that the most important, if not all, theories ever proposed in the history of science are metaphysical, that most, if not all, of the accepted progress is pseudo-progress, that most, if not all, of the work done is irrational. If, however, still accepting the demarcation criterion of dogmatic falsificationism, we deny that facts can prove propositions, then we certainly end up in complete scepticism: Then all science is undoubtedly irrational metaphysics and should be rejected. *Scientific theories are not only equally unprovable, and equally improbable, but they are also equally undisprovable.* But the recognition that not only the theoretical but *all* the propositions in science are fallible means the total collapse of *all* forms of dogmatic falsificationism as theories of scientific rationality.

Methodological Falsificationism. The "Empirical Basis."

The collapse of dogmatic falsificationism because of fallibilistic arguments seems to bring us back to square one. If *all* scientific statements are fallible theories, one can criticize them only for inconsistency. But then, in what sense, if any, is science empirical? If scientific theories are neither provable, nor probabilifiable, nor disprovable, then the sceptics seem to be finally right: Science is no more than vain speculation and there is no such thing as progress in scientific knowledge. Can we still oppose scepticism? *Can we save scientific criticism from fallibilism?* Is it possible to have a fallibilistic theory of scientific progress? In particular, if scientific criticism is fallible, on what ground can we ever eliminate a theory?

A most intriguing answer is provided by *methodological falsificationism*. Methodological falsificationism is a brand of conventionalism; therefore, in order to understand it, we must first discuss conventionalism in general.

There is an important demarcation between *"passivist" and "activist" theories of knowledge.* "Passivists" hold that true knowledge is Nature's imprint on a perfectly inert mind: Mental *activity* can only result in bias and distortion. The most influential passivist school is classical empiricism. "Activists" hold that we cannot read the book of Nature without mental activity, without interpreting it in the light of our expectations or theories.[27] Now *conservative "activists"* hold that we are born with our basic expectations; with them we turn the world into "our world" but must then live for ever in the prison of our world. The idea that we live and die in the prison of our "conceptual frameworks" was developed primarily by Kant; pessimistic Kantians thought that the real world is forever unknowable because of this prison, while optimistic Kantians thought that God created our conceptual framework to fit the world. . . . But *revolutionary activists* believe that conceptual frameworks can be developed and also replaced by new, *better* ones; it is *we* who create our "prisons" and we can also, critically, demolish them. . . .

New steps from conservative to revolutionary activism were made by Whewell and then by Poincaré, Milhaud and Le Roy. . . . Poincaré, Milhaud and Le Roy . . . preferred to explain the continuing historical success of Newtonian mechanics by a *methodological decision* taken by scientists: After a considerable period of initial empirical success scientists may *decide* not to allow the theory to be refuted. Once they have taken this decision, they solve (or dissolve) the apparent anomalies by auxiliary hypotheses or other "conventionalist stratagems."[28] This *conservative conventionalism* has, however, the disadvantage of making us unable to get out of our self-imposed prisons, once the first period of trial-and-error is over and the great decision taken. It cannot solve the problem of the elimination of those theories which have been triumphant for a long period. According to conservative conventionalism, experiments may have sufficient power to refute young theories, but not to refute old, established theories: *As science grows, the power of empirical evidence diminishes.*[29]

Poincaré's critics refused to accept his idea, that, although the scientists build their conceptual frameworks, there comes a time when these frameworks turn into prisons which cannot be demolished. This criticism gave rise to two rival schools of *revolutionary conventionalism*: Duhem's simplicism and Popper's methodological falsificationism.[30]

Duhem accepts the conventionalists' position that no physical theory ever crumbles merely under the weight of "refutations," but claims that it still may crumble under the weight of "continual repairs, and many tangled-up stays" when "the worm-eaten columns" cannot support "the tottering building" any longer[31]; then the theory loses its original simplicity and has to be replaced. But falsification is then left to subjective taste or, at best, to scientific fashion, and leaves too much leeway for dogmatic adherence to a favorite theory. . . .

Popper set out to find a criterion which is both more objective and more hard-hitting. He could not accept the emasculation of empiricism, inherent even in Duhem's approach, and proposed a methodology which allows experiments to be powerful even in "mature" science. Popper's

[27] This demarcation—and terminology—is due to Popper; cf. especially his [1934], Section 19 and his [1945], Chapter 23 and footnote 3 to Chapter 25.

[28] Cf. especially Poincaré [1891] and [1902]; Milhaud [1896]; Le Roy [1899] and [1901]. It was one of the chief philosophical merits of conventionalists to direct the limelight to the fact that any theory can be saved by "conventionalist stratagems" from refutations. (The term *conventionalist stratagem* is Popper's; cf. the critical discussion of Poincaré's conventionalism in his [1934], especially Sections 19 and 20.)

[29] Poincaré first elaborated his conventionalism only with regard to geometry (cf. his [1891]). Then Milhaud and Le Roy generalized Poincaré's idea to cover all branches of accepted physical theory. Poincaré's [1902] starts with a strong criticism of the Bergsonian Le Roy against whom he defends the empirical (falsifiable or "inductive") character of all physics *except for* geometry and mechanics. Duhem, in turn, criticized Poincaré: In his view there was a possibility of overthrowing even Newtonian mechanics.

[30] The *loci classici* are Duhem's [1905] and Popper's [1934]. Duhem was not a *consistent* revolutionary conventionalist. Very much like Whewell, he thought that conceptual changes are only *preliminaries* to the final—if perhaps distant—"natural classification":"The more a theory is perfected, the more we apprehend that the logical order in which it arranges experimental laws is the reflection of an ontological order." In particular, he refused to see Newton's mechanics *actually* "crumbling" and characterized Einstein's relativity theory as the manifestation of a "frantic and hectic race in pursuit of a novel idea" which "has turned physics into a real chaos where logic loses its way and commonsense runs away frightened" (Preface—of 1914—to the second edition of his [1905]).

[31] Duhem [1905], Chapter VI, Section 10.

methodological falsificationism is both conventionalist and falsificationist, but he "differs from the [conservative] conventionalists in holding that the statements decided by agreement are *not* [spatio-temporally] universal but [spatio-temporally] singular"[32]; and he differs from the dogmatic falsificationist in holding that the truth-value of such statements cannot be proved by facts but, in some cases, may be decided by agreement.[33]

The *conservative conventionalist* (or methodological justificationist, if you wish) makes unfalsifiable by *fiat* some (spatio-temporally) universal theories, which are distinguished by their explanatory power, simplicity, or beauty. Our *revolutionary conventionalist* (or "methodological falsificationist") makes unfalsifiable by *fiat* some (spatio-temporally) singular statements which are distinguishable by the fact that there exists at the time a "relevant technique" such that "anyone who has learned it" will be able to *decide* that the statement is "acceptable."[34] Such a statement may be called an "observational" or "basic" statement, but only in inverted commas.[35] Indeed, the very selection of all such statements is a matter of a decision, which is not based on exclusively psychological considerations. This decision is then followed by a second kind of decision concerning the separation of the set of *accepted* basic statements from the rest.

These *two decisions* correspond to the *two assumptions* of dogmatic falsificationism. But there are important differences. First, the methodological falsificationist is not a justificationist, he has no illusions about "experimental proofs" and is fully aware of the fallibility of his decisions and the risks he is taking.

The methodological falsificationist realizes that in the "experimental techniques" of the scientist fallible theories are involved,[36] "in the light of which" he interprets the facts. In spite of this he "applies" these theories, he regards them in the given context not as theories under test but as *unproblematic background knowledge*, "which we accept (tentatively) as unproblematic while we are testing the theory."[37] He may call these theories—and the statements whose truth-value he decides in their light—"observational," but this is only a manner of speech which he inherited from naturalistic falsificationism.[38] The methodological falsificationist *uses our most successful theories as extensions of our senses* and widens the range of theories which can be applied in testing far beyond the dogmatic falsificationist's range of strictly observational theories. For instance, let us imagine that a big radio-star is discovered with a system of radio-star satellites orbiting it. We should like to test some gravitational theory on this planetary system—a matter of considerable interest. Now let us imagine that Jodrell Bank succeeds in providing a set of space-time co-ordinates of the planets which is inconsistent with the theory. We shall take these statements as potential falsifiers. Of course, these basic statements are not "observational" in the usual sense but only " 'observational.' " They describe planets that neither the human eye nor optical instruments can reach. Their truth-value is arrived at by an "experimental technique." This "experimental technique" is based on the "application" of a well-corroborated theory of radio-optics. Calling these statements "observational" is no more than a manner of saying that, in the context of his problem, that is, in testing our gravitational theory, the methodological falsificationist uses radio-optics uncritically, as "background knowledge." *The need for decisions to demarcate the theory under test from unproblematic background knowledge is a characteristic*

---

[32] Popper [1934], Section 30.

[33] *In this section I discuss the "naive" variant of Popper's methodological falsificationism. Thus, throughout the section "methodological falsificationism" stands for "naive methodological falsificationism"; for this "naivety," cf. below*, pp. 181–182.

[34] Popper [1934], Section 27.

[35] Op cit. Section 28. For the non-basicness of these methodologically "basic" statements, cf. e.g. Popper [1934] *passim* and Popper [1959], p. 35, footnote *2.

[36] Cf. Popper [1934], end of Section 26 and also his [1968], pp. 291–2.

[37] Cf. Popper [1963], p. 390.

[38] Indeed, Popper carefully puts "observational" in quotes; cf. his [1934], Section 28.

feature of this brand of methodological falsificationism.[39] (This situation does not really differ from Galileo's "observation" of Jupiter's satellites; moreover, as some of Galileo's contemporaries rightly pointed out, he relied on a virtually non-existent optical theory—which then was less corroborated, and even less articulated, than present-day radio-optics. On the other hand, calling the reports of our human eye "observational" only indicates that we "rely" on some vague physiological theory of human vision.[40])

This consideration shows the conventional element in granting—in a given context—the (methodologically) "observational" status to a theory.[41] Similarly, there is a considerable conventional element in the decision concerning the actual truth-value of a basic statement which we take after we have decided which "observational theory" to apply. One single observation may be the stray result of some trivial error: In order to reduce such risks, methodological falsificationists prescribe some safety control. The simplest such control is to repeat the experiment (it is a matter of convention how many times). . . .

This is how the methodological falsificationist establishes his "empirical basis." (He uses quotation marks in order "to give ironical emphasis" to the term.[42]) This "basis" can be hardly called a "basis" by justificationist standards: There is nothing proven about it—it denotes "piles driven into a swamp."[43] Indeed, if this "empirical basis" clashes with a theory, the theory may be *called* "falsified", but it is not falsified in the sense that it is disproved. Methodological "falsification" is very different from dogmatic falsification. If a theory is falsified, it is proven false; if it is "falsified", it may still be true. If we follow up this sort of "falsification" by the actual "elimination" of a theory, we may well end up by eliminating a true, and accepting a false, theory (a possibility which is thoroughly abhorrent to the old-fashioned justificationist). . . .

The methodological falsificationist separates rejection and disproof, which the dogmatic falsificationist had conflated. . . . He is a fallibilist but his fallibilism does not weaken his critical stance: He turns fallible propositions into a "basis" for a hard-line policy. On these grounds he proposes a *new demarcation criterion*: Only those theories—that is, non-"observational" propositions—that forbid certain "observable" states of affairs, and therefore may be "falsified"and rejected, are "scientific"; or, briefly, *a theory is "scientific"* (or *"acceptable") if it has an "empirical basis."* This criterion brings out sharply the difference between dogmatic and methodological falsificationism.[44] . . .

But even these three decisions are not sufficient to enable us to "falsify" a theory which cannot explain anything "observable" without a *ceteris paribus* clause.[45] No finite number of "observations" is enough to "falsify" such a theory. However, if this is the case, how can one reasonably defend a methodology which claims to "interpret natural laws or theories as . . . statements which are partially decidable, i.e., which are, for logical reasons, not verifiable but, in an asymmetrical way, falsifiable . . . ?[46] How can we interpret theories like Newton's theory of dynamics and gravitation as "one-sidedly decidable"?[47] How can we make in such cases genuine "attempts to weed out false theories—to find the weak points of a theory in order to reject it if it is falsified by the test"?[48] How can we draw them

[39] This demarcation plays a role both in the *first* and in the *fourth* type of decisions of the methodological falsificationist. (For the *fourth* decision, cf. below, p. 180.)

[40] For a fascinating discussion, cf. Feyerabend [1969].

[41] One wonders whether it would not be better to make a break with the terminology of naturalistic falsificationism and rechristen observational theories *"touchstone theories."*

[42] Popper [1963], p. 387.

[43] Popper [1934], Section 30; also cf. Section 29: "The Relativity of Basic Statements."

[44] The demarcation criterion of the dogmatic falsificationist was: A theory is "scientific" if it has an empirical basis (see *above*, p. 173).

[45] Cf. *above*, pp. 175–176.

[46] Popper [1933].

[47] Popper [1933].

[48] Popper [1957], p. 133.

into the realm of rational discussion? The methodological falsificationist solves the problem by making a further (*fourth type*) *decision:* When he tests a theory together with a *ceteris paribus* clause and finds that this conjunction has been refuted, he must decide whether to take the refutation also as a refutation of the specific theory. For instance, he may accept Mercury's "anomalous" perihelion as a refutation of the treble conjunction N_3 of Newton's theory, the known initial conditions, and the *ceteris paribus* clause. Then he tests the initial conditions "severely[49] and may decide to relegate them into the "unproblematic background knowledge." This decision implies the refutation of the double conjunction N_2 of Newton's theory and the *ceteris paribus* clause. Now he has to take the crucial decision: whether to relegate also the *ceteris paribus* clause into the pool of "unproblematic background knowledge." He will do so if he finds the *ceteris paribus* clause well corroborated.

How can one test a *ceteris paribus* clause severely? By assuming that there *are* other influencing factors, by specifying such factors, and by testing these specific assumptions. If many of them are refuted, the *ceteris paribus* clause will be regarded as well corroborated. . . .

Thus, with the help of this fourth type of decision,[50] our methodological falsificationist has finally succeeded in interpreting even theories like Newton's theory as "scientific." . . .

To sum up: The methodological falsificationist offers an interesting solution to the problem of combining hard-hitting criticism with fallibilism. Not only does he offer a philosophical basis for falsification after fallibilism had pulled the carpet from under the feet of the dogmatic falsificationist, but he also widens the range of such criticism very considerably. By putting falsification in a new setting, he saves the attractive code of honour of the dogmatic falsificationist: that scientific honesty consists in specifying, in advance, an experiment such that if the result contradicts the theory, the theory has to be given up.[51]

Methodological falsificationism represents a considerable advance beyond both dogmatic falsificationism and conservative conventionalism. It recommends risky decisions. But the risks are daring to the point of recklessness and one wonders whether there is no way of lessening them.

Let us first have a closer look at the risks involved.

Decisions play a crucial role in this methodology—as in any brand of conventionalism. Decisions, however, may lead us disastrously astray. The methodological falsificationist is the first to admit this. But this, he argues, is the price which we have to pay for the possibility of progress.

One has to appreciate the dare-devil attitude of our methodological falsificationist. He feels himself to be a hero who, faced with two catastrophic alternatives, dared to reflect coolly on their relative merits and choose the lesser evil. One of the alternatives was sceptical fallibilism, with its "anything goes" attitude, the despairing abandonment of all intellectual standards, and hence of the idea of scientific progress. Nothing can be established, nothing can be rejected, nothing even communicated: The growth of science is a growth of chaos, a veritable Babel. For two thousand years, scientists and scientifically minded philosophers chose justificationist illusions of some kind to escape this nightmare. . . . Our methodological falsificationist proudly rejects such escapism: he dares to measure up to the full impact of fallibilism and yet escape skepticism by a daring and risky conventionalist policy, with no dogmas. He is fully aware of the risks but insists that *one has to choose between some sort of methodological falsificationism and irrationalism.* He offers a game in which one has little hope of winning, but claims that it is still better to play than to give up. . . .

But is not the firm strategy of the brand of methodological falsificationism hitherto discussed

[49] For a discussion of this important concept of Popperian methodology, cf. my [1968a], pp. 397 ff.

[50] This type of decision belongs, in an important sense, to the same category as the first decision: It demarcates, by decision, problematic from unproblematic knowledge. Cf. *above* p. 179, text to footnote 39.

[51] See *above*, p. 172.

too firm? Are not the decisions it advocates bound to be *too arbitrary?* Some may even claim that all that distinguishes methodological from dogmatic falsificationism is that *it pays lip-service to fallibilism!*

To criticize a theory of criticism is usually very difficult. Naturalistic falsificationism was relatively easy to refute, since it rested on an empirical psychology of perception: One could show that it was simply *false*. But how can methodological falsificationism be falsified? No disaster can ever disprove a non-justificationist theory of rationality. Moreover, how can we ever recognize an epistemological disaster? We have no means to judge whether the verisimilitude of our successive theories increases or decreases.[52] At this stage we have not yet developed a general theory of criticism even for scientific theories, let alone for theories of rationality[53]: therefore, if we want to falsify our methodological falsificationism, we have to do it before having a theory of how to do it.

If we look at history of science, if we try to see how some of the most celebrated falsifications happened, we have to come to the conclusion that either some of them are plainly irrational or that they rest on rationality principles radically different from the ones we just discussed. First of all, our falsificationist must deplore the fact that stubborn theoreticians frequently challenge experimental verdicts and have them reversed. In the falsificationist conception of scientific "law and order" we have described there is no place for such successful appeals. Further difficulties arise from the falsification of theories to which a *ceteris paribus* clause is appended.[54] Their falsification as it occurs in actual history is *prima facie* irrational by the standards of our falsificationist. By his standards, scientists frequently seem to be irrationally slow: For instance, eighty-five years elapsed between the acceptance of the perihelion of Mercury as an anomaly and its acceptance as a falsification of Newton's theory, in spite of the fact that the *ceteris paribus* clause was reasonably well corroborated. On the other hand, scientists frequently seem to be irrationally rash: For instance, Galileo and his disciples accepted Copernican heliocentric celestial mechanics in spite of the abundant evidence against the rotation of the Earth; or Bohr and his disciples accepted a theory of light emission in spite of the fact that it ran counter to Maxwell's well-corroborated theory.

Indeed, it is not difficult to see at least two crucial characteristics common to both dogmatic and our methodological falsificationism which are clearly dissonant with the actual history of science: that (1) *a test is—or must be made—a two-cornered fight between theory and experiment so that in the final confrontation only these two face each other; and* (2) *the only interesting outcome of such confrontation is (conclusive) falsification: "[the only genuine] discoveries are refutations of scientific hypotheses."*[55] However, history of science suggests that (1′) tests are—at least—three-cornered fights between rival theories and experiment and (2′) some of the most interesting experiments result, *prima facie,* in confirmation rather than falsification.

But if—as seems to be the case—the history of science does not bear out our theory of scientific rationality, we have two alternatives. One alternative is to abandon efforts to give a rational explanation of the success of science. Scientific method (or "logic of discovery"), conceived as the discipline of rational appraisal of scientific theories—and of criteria of *progress*—vanishes. We, may, of course, still try to explain *changes* in "paradigms" in terms of social psychology. . . . This is

[52] I am using here "*verisimilitude*" in Popper's sense: the difference between the truth content and falsity content of a theory. For the risks involved in estimating it, cf. my [1968a], especially pp. 395 ff.

[53] I tried to develop such a general theory of criticism in my [1970].

[54] The falsification of theories depends on the high degree of corroboration of the *ceteris paribus* clause. This, however, is not always the case. This is why the methodological falsificationist may advise us to rely on our "scientific instinct" (Popper [1934], Section 18, footnote 2) or "hunch" (Braithwaite [1953], p. 20).

[55] Agassi [1959]; he calls Popper's idea of science "scientia negativa" (Agassi [1968]).

Polanyi's and Kuhn's way.[56] The other alternative is to try at least to *reduce* the conventional element in falsificationism (we cannot possibly eliminate it) and replace the *naive* versions of methodological falsificationism—characterized by the theses (1) and (2) above—by a *sophisticated* version which would give a new *rationale* of falsification and thereby rescue methodology and the idea of scientific *progress*. This is Popper's way, and the one I intend follow.

Sophisticated versus Naive Methodological Falsificationism. Progressive and Degenerating Problemshifts.

Sophisticated falsificationism differs from naive falsificationism both in its rules of *acceptance* (or "demarcation criterion") and its rules of *falsification* or elimination. For the naive falsificationist any theory which can be interpreted as experimentally falsifiable is "acceptable" or "scientific."[57] For the sophisticated falsificationist a theory is "acceptable" or "scientific" only if it has corroborated excess empirical content over its predecessor (or rival), that is, only if it leads to the discovery of novel facts. This condition can be analysed into two clauses: that the new theory has excess empirical content ("*acceptability*"$_1$) and that some of this excess content is verified ("*acceptability*"$_2$). The first clause can be checked instantly . . . by *a priori* logical analysis; the second can be checked only empirically and this may take an indefinite time.

Again, for the naive falsificationist a theory is *falsified* by a ("fortified" . . .) "observational" statement which conflicts with it (or rather, which he decides to interpret as conflicting with it). The sophisticated falsificationist regards a scientific theory T as falsified if and only if another theory T' has been proposed with the following characteristics: (1) T' has excess empirical content over T; that is, it predicts *novel* facts, that is, facts improbable in the light of, or even forbidden, by T;[58] (2) T' explains the previous success of T, that is, all the unrefuted content of T is contained (within the limits of observational error) in the content of T'; and (3) some of the excess content of T' is corroborated.[59]

In order to be able to appraise these definitions we need to understand their problem background and their consequences. First, we have to remember the conventionalists' methodological discovery that no experimental result can ever kill a theory: any theory can be saved from counterinstances either by some auxiliary hypothesis or by a suitable reinterpretation of its terms. Naive falsificationists solved this problem by relegating—in crucial contexts—the auxiliary hypotheses to the realm of unproblematic background knowledge, eliminating them from the deductive model of the test-situation and thereby *forcing* the chosen theory into logical isolation, in which it becomes a sitting target for the attack of test-experiments. But since this procedure did not offer a suitable guide for a rational reconstruction of the history of science, we may just as well completely rethink our approach. Why aim at falsification at any price? Why not rather impose certain standards on the theoretical adjustments by which one is allowed to save a theory? Indeed, some such standards have been well known for centuries, and we find them expressed in age-old wisecracks against *ad hoc* explanations, empty prevarications, face-saving, linguistic tricks.[60] We have already seen that Duhem adumbrated such stan-

[56] Feyerabend, who contributed probably more than anybody else to the spread of Popper's ideas, seems now to have joined the enemy camp. Cf. his intriguing [1970].

[57] Cf. *above*, p. 179.

[58] I use "*prediction*" in a wide sense that includes "*postdiction*."

[59] *For a detailed discussion of these acceptance and rejection rules and for references to Popper's work*, cf. my [1968a], pp. 375–90. For some qualifications (concerning continuity and consistency as regulative principles), cf. *below*, pp. 190–191. . . .

[60] Molière, for instance, ridiculed the doctors of his *Malade Imaginaire*, who offered the *virtus dormitiva* of opium as the answer to the question as to why opium produced sleep. One might even argue that Newton's famous dictum *hypotheses non fingo* was really directed against *ad hoc* explanations—like his own explanation of gravitational forces by an aether-model in order to meet Cartesian objections.

[61] Cf. *above*, p. 177.

dards in terms of "simplicity" and "good sense."[61] But *when* does lack of "simplicity" in the protective belt of theoretical adjustments reach the point at which the theory *must* be abandoned?[62] In what sense was Copernican theory, for instance, "simpler" than Ptolemaic?[63] The vague notion of Duhemian "simplicity" leaves, as the naive falsificationist correctly argued, the decision very much to taste and fashion.[64]

Can one improve on Duhem's approach? Popper did. His solution—a sophisticated version of methodological falsificationism—is more objective and more rigorous. Popper agrees with the conventionalists that theories and factual propositions can always be harmonized with the help of auxiliary hypotheses: He agrees that the problem is how to demarcate between scientific and pseudoscientific *adjustments,* between rational and irrational changes of theory. According to Popper, saving a theory with the help of auxiliary hypotheses which satisfy certain well-defined conditions represents scientific progress; but saving a theory with the help of auxiliary hypotheses which do not, represents degeneration. Popper calls such inadmissible auxiliary hypotheses *ad hoc* hypotheses, mere linguistic devices, "conventionalist stratagems."[65] But then any scientific theory has to be appraised together with its auxiliary hypotheses, initial conditions, etc., and, especially, together with its predecessors so that we may see by what sort of *change* it was brought about. Then, of course, what we appraise is a *series of theories* rather than isolated *theories.*

Now we can easily understand why we formulated the criteria of acceptance and rejection of sophisticated methodological falsificationism as we did.[66] But it may be worth while to reformulate them slightly, couching them explicitly in terms of *series of theories.*

Let us take a series of theories, T_1, T_2, T_3, . . . where each subsequent theory results from adding auxiliary clauses to (or from semantical reinterpretations of) the previous theory in order to accommodate some anomaly, each theory having at least as much content as the unrefuted content of its predecessor. Let us say that such a series of theories is *theoretically progressive (or "constitutes a theoretically progressive problemshift")* if each new theory has some excess empirical content over its predecessor; that is, if it predicts some novel, hitherto unexpected fact. Let us say that a theoretically progressive series of theories is also *empirically progressive (or "constitutes an empirically progressive problemshift")* if some of this excess empirical content is also corroborated, that is, if each new theory leads us to the actual discovery of some *new fact.*[67] Finally, let us call a problemshift *progressive* if it is both theoretically and empirically progressive, and *degenerating* if it is not.[68] We *"accept"* problemshifts as "scientific" only if they are at least theoretically progressive; if they are not, we *"reject"* them as "pseudoscientific." Progress is measured by the degree to which a problemshift is progressive, by the degree to which the series of theories leads us to the discovery of novel facts. We regard a theory in the series

[62] Incidentally, Duhem agreed with Bernard that experiments alone—without simplicity considerations—can decide the fate of theories in physiology. But in physics, he argued, they cannot ([1905] Chapter VI, Section I).

[63] Koestler correctly points out that only Galileo created the myth that the Copernican theory was simple (Koestler [1959], p. 476); in fact, "the motion of the earth [had not] done much to simplify the old theories, for though the objectionable equants had disappeared, the system was still bristling with auxiliary circles" (Dreyer [1906], Chapter XIII).

[64] Cf. *above,* p. 177.

[65] Popper [1934], sections 19 and 20. I have discussed in some detail—under the heads "monster-barring," "exception-barring," "monster-adjustment"—such stratagems as they appear in informal, quasi-empirical mathematics; cf. my [1963–4].

[66] Cf. *above,* p. 182.

[67] If I already know P_1: "Swan *A* is white," P_ω: "All swans are white" represents no progress, because it may only lead to the discovery of such further similar facts as P_2: "Swan *B* is white." So-called "empirical generalizations" constitute no progress. A *new* fact must be improbable or even impossible in the light of previous knowledge. Cf. *above,* p. 182. . . .

[68] The appropriateness of the term "problemshift" for a series of theories rather than of problems may be questioned. I chose it partly because I have not found a more appropriate alternative—"theoryshift" sounds dreadful—partly because theories are always problematical, they never solve all the problems they have set out to solve. Anyway, in the second half of the paper, the more natural term "research programme" will replace "problemshifts" in the most relevant contexts.

"falsified" when it is superseded by a theory with higher corroborated content. . . .

This demarcation between progressive and degenerating problemshifts sheds new light on the appraisal of *scientific—or rather, progressive—explanations*. If we put forward a theory to resolve a contradiction between a previous theory and a counterexample in such a way that the new theory, instead of offering a content-increasing (scientific) *explanation*, only offers a content-decreasing (linguistic) *reinterpretation*, the contradiction is resolved in a merely semantical, unscientific way. *A given fact is explained scientifically only if a new fact is also explained with it.* . . .

Sophisticated falsificationism thus shifts the problem of how to appraise *theories* to the problem of how to appraise *series of theories*. Not an isolated *theory*, but only a series of theories can be said to be scientific or unscientific: to apply the term "scientific" to one *single* theory is a category mistake.[69]

The time-honoured empirical criterion for a satisfactory theory was agreement with the observed facts. Our empirical criterion for a series of theories is that it should produce new facts. *The idea of growth and the concept of empirical character are soldered into one.*

This revised form of methodological falsificationism has many new features. First, it denies that "in the case of a scientific theory, our decision depends upon the results of experiments. If these confirm the theory, we may accept it until we find a better one. If they contradict the theory, we reject it."[70] It denies that "what ultimately decides the fate of a theory is the result of a test, i.e., an agreement about basic statements."[71] Contrary to naive falsificationism, *no experiment, experimental report, observation statement or well-corroborated low-level falsifying hypothesis alone can lead to falsification. . . . There is no falsification before the emergence of a better theory.*[72] But then the distinctively negative character of naive falsificationism vanishes; criticism becomes more difficult, and also positive, constructive. But, of course, if falsification depends on the emergence of better theories, on the invention of theories which anticipate new facts, then falsification is *not* simply a relation between a theory and the empirical basis, but a multiple relation between competing theories, the original "empirical basis," and the empirical growth resulting from the competition. Falsification can thus be said to have a *"historical character."*[73] Moreover, some of the theories which bring about falsification are frequently proposed *after* the "counterevidence." This may sound paradoxical for people indoctrinated with naive falsificationism. Indeed, this epistemological theory of the relation between theory and experiment differs sharply from the epistemological theory of naive falsificationism. The very term "counterevidence" has to be abandoned in the sense that no experimental result must be interpreted directly as "counterevidence." If we still want to retain this time-honoured term, we have to redefine it like this: "Counterevidence to T_1" is a corroborating instance to T_2, which is either inconsistent with or independent of T_1 (with the *proviso* that T_2 is a theory which satisfactorily explains the empirical success of T_1). This shows that *"cru-*

[69] Popper's conflation of "theories" and "series of theories" prevented him from getting the basic ideas of sophisticated falsificationism across more successfully. His ambiguous usage led to such confusing formulations as "Marxism [as the core of a series of theories or of a research programme] is irrefutable" and, at the same time, "Marxism [as a particular conjunction of this core and some specified auxiliary hypotheses, initial conditions and a *ceteris paribus* clause] has been refuted." (Cf. Popper [1963].)

Of course, there is nothing wrong in saying that an isolated, single theory is "scientific" if it represents an advance on its predecessor, as long as one clearly realizes that in this formulation we appraise the theory as the outcome of—and in the context of—a certain historical development.

[70] Popper [1945], Vol. II, p. 233. Popper's more sophisticated attitude surfaces in the remark that "concrete and practical consequences can be *more* directly tested by experiment" (*ibid;* my italics).

[71] Popper [1934], Section 30.

[72] "In most cases we have, before falsifying a hypothesis, another one up our sleeves" (Popper [1959], p. 87, footnote *I). But, as our argument shows, we *must* have one. Or, as Feyerabend put it: "The best criticism is provided by those theories which can replace the rivals they have removed" ([1965], p. p. 227). He notes that in *some* cases "alternatives will be quite indispensable for the purpose of refutation" (*ibid*, p. 254). But according to our argument *refutation without an alternative shows nothing but the poverty of our imagination in providing a rescue hypothesis*. . . .

[73] Cf. my [1968a], pp. 387 ff.

cial counterevidence"—or "*crucial experiments*"—can be recognized as such among the scores of anomalies only *with hindsight*, in the light of some superseding theory.[74]

Thus the crucial element in falsification is whether the *new theory* offers any novel, excess information compared with its predecessor and whether some of this excess information is corroborated. Justificationists valued "confirming" instances of a theory; naive falsificationists stressed "refuting" instances; for the methodological falsificationists it is the—rather rare—corroborating instances of the *excess* information which are the crucial ones; these receive all the attention. We are no longer interested in the thousands of trivial verifying instances nor in the hundreds of readily available anomalies: The few crucial *excess-verifying instances* are decisive.[75] This consideration rehabilitates—and reinterprets—the old proverb: *Exemplum docet, exempla obscurant.*

"Falsification" in the sense of naive falsificationism (corroborated counterevidence) is not a *sufficient* condition for eliminating a specific theory: In spite of hundreds of known anomalies we do not regard it as falsified (that is, eliminated) until we have a better one.[76] Nor is "falsification" in the naive sense *necessary* for falsification in the sophisticated sense: A progressive problemshift does not have to be interspersed with "refutations." Science can grow without any "refutations" leading the way. Naive falsificationists suggest a linear growth of science, in the sense that theories are followed by powerful refutations which eliminate them; these refutations in turn are followed by new theories.[77] It is perfectly *possible* that theories be put forward "progressively" in such a rapid succession that the "refutation" of the *n*th appears only as the corroboration of the $n + 1$-th. The problem fever of science is raised by proliferation of rival theories rather than counterexamples or anomalies.

This shows that the slogan of *proliferation of theories* is much more important for sophisticated than for naive falsificationism. For the naive falsificationist science grows through repeated experimental overthrow of theories; new rival theories proposed before such "overthrows" may speed up growth but are not absolutely necessary . . . ; constant proliferation of theories is optional but not mandatory. For the sophisticated falsificationist proliferation of theories cannot wait until the accepted theories are "refuted" (or until their protagonists get into a Kuhnian crisis of confidence).[78] While naive falsificationism stresses "the urgency of replacing a *falsified* hypothesis by a better one,"[79] sophisticated falsificationism stresses the urgency of replacing *any* hypothesis by a better one. Falsification cannot "compel the theorist to search for a better theory,"[80] simply because falsification cannot precede the better theory.

The problem-shift from naive to sophisticated falsificationism involves a semantic difficulty. For the naive falsificationist a "refutation" is an experimental result which, by force of his decisions, is made to conflict with the theory under test. But according to sophisticated falsificationism one must not take such decisions before the alleged "refuting instance" has become the confirming instance of a new, better theory. Therefore when-

[74] In the distorting mirror of naive falsificationism, new theories which replace old refuted ones are themselves born unrefuted. Therefore they do not believe that there is a relevant difference between anomalies and crucial counterevidence. For them, anomaly is a dishonest euphemism for counterevidence. But in actual history new theories are born refuted: They inherit many anomalies of the old theory. Moreover, frequently it is *only* the new theory which dramatically predicts that fact which will function as crucial counterevidence against its predecessor, while the "old" anomalies may well stay on as "new" anomalies. . . .

[75] *Sophisticated falsificationism adumbrates a new theory of learning;* cf. *below,* p. 186.

[76] It is clear that the theory T' may have excess corroborated empirical content over another theory T even if both T and T' are refuted. Empirical content has nothing to do with truth or falsity. Corroborated contents can also be compared irrespective of the refuted content. Thus we may see the rationality of the elimination of Newton's theory in favour of Einstein's, even though Einstein's theory may be said to have been born—like Newton's—"refuted." We have only to remember that "qualitative confirmation" is a euphemism for "quantitative disconfirmation." (Cf. my [1968a], pp. 384–6.)

[77] Cf. Popper [1934], Section 85, p. 279 of the 1959 English translation.

[78] Also cf. Feyerabend [1965], pp. 254–5.

[79] Popper [1959], p. 87, footnote *I.

[80] Popper [1934], Section 30.

ever we see terms like "refutation," "falsification," "counterexample," we have to check in each case whether these terms are being applied in virtue of decisions by the naive or by the sophisticated falsificationist.[81]

Sophisticated methodological falsificationism offers new standards for intellectual honesty. Justificationist honesty demanded the acceptance of only what was proven and the rejection of everything unproven. Neojustificationist honesty demanded the specification of the probability of any hypothesis in the light of the available empirical evidence. The honesty of naive falsificationism demanded the testing of the falsifiable and the rejection of the unfalsifiable and the falsified. Finally, the honesty of sophisticated falsificationism demanded that one should try to look at things from different points of view, to put forward new theories which anticipate novel facts, and to reject theories which have been superseded by more powerful ones.

Sophisticated methodological falsificationism blends several different traditions. From the empiricists it has inherited the determination to learn primarily from experience. From the Kantians it has taken the activist approach to the theory of knowledge. From the conventionalists it has learned the importance of decisions in methodology.

I should like to emphasize here a further distinctive feature of sophisticated methodological empiricism: the crucial role of excess corroboration. For the inductivist, learning about a new theory is learning how much confirming evidence supports it; about refuted theories one *learns* nothing (learning, after all, is to build up proven or probable *knowledge*). For the dogmatic falsificationist, learning about a theory is learning whether it is refuted or not; about confirmed theories one learns nothing (one cannot prove or probabilify anything) about refuted theories one learns that they are disproved.[82] For the sophisticated falsificationist, learning about a theory is primarily learning which new facts it anticipated; indeed, for the sort of Popperian empiricism I advocate, the only relevant evidence is the evidence anticipated by a theory, and *empiricalness (or scientific character) and theoretical progress are inseparably connected*. . . .

This idea is not entirely new. Leibnitz, for instance, in his famous letter to Conring in 1678, wrote: "It is the greatest commendation of an hypothesis (next to [proven] truth) if by its help predictions can be made even about phenomena or experiments not tried."[83] Leibnitz's view was widely accepted by scientists. But since all appraisal of a scientific theory was before Popper appraisal of its degree of justification, this position was regarded by some logicians as untenable. Mill, for instance, complains in 1843 in horror that "it seems to be thought that an hypothesis . . . is entitled to a more favourable reception, if besides accounting for all the facts previously known, it has led to the anticipation and prediction of others which experience afterwards verified."[84] Mill had a point; this appraisal was in conflict both with justificationism and with probabilism: Why should an event *prove* more, if it was anticipated by the theory than if it was known already before? As long as *proof* was the only criterion of the scientific character of a theory, Leibnitz's criterion could only be regarded as irrelevant.[85] Also, the *probability* of a theory given

[81] . . . Possibly it would be better in future to abandon these terms altogether, just as we have abandoned terms like "*inductive* (or experimental) proof." Then we may call (naive) "refutations" anomalies, and (sophisticatedly) "falsified" theories "superseded" ones. Our "ordinary" language is impregnated not only by "inductivist" but also by falsificationist dogmatism. A reform is overdue.

[82] For a defense of this theory of "learning from experience," cf. Agassi [1969].

[83] Cf. Leibnitz [1678]. The expression in brackets shows that Leibnitz regarded this criterion as second best and thought that the best theories are those which are proved. Thus Leibnitz's position—like Whewell's—is a far cry from fully fledged sophisticated falsificationism.

[84] Mill [1843], vol. II, p. 23.

[85] This was J. S. Mill's argument (*ibid*). He directed it against Whewell, who thought that "consilience of inductions" or successful prediction of improbable events *verifies* (that is, *proves*) a theory. (Whewell [1858], pp. 95–6.) No doubt, *the basic contradiction both in Whewell's and in Duhem's philosophy of science is their conflation of heuristic power and proven truth. Popper separated the two.*

evidence cannot possibly be influenced, as Keynes pointed out, by *when* the evidence was produced; the probability of a theory given evidence can depend only on the theory and the evidence,[86] and not upon whether the evidence was produced before or after the theory.

In spite of this convincing justificationist criticism, the criterion survived among some of the best scientists, since it formulated their strong dislike of merely *ad hoc* explanations, which "though [they] truly express the facts [they set out to explain, are] not born out by any other phenomena."[87]

But it was only Popper who recognized that the *prima facie* inconsistency between the few odd, casual remarks against *ad hoc* hypotheses on the one hand and the huge edifice of justificationist philosophy of knowledge must be solved by demolishing justificationism and by introducing new, non-justificationist criteria for appraising scientific theories based on anti-adhocness.

Let us look at a few examples. Einstein's theory is not better than Newton's *because* Newton's theory was "refuted" but Einstein's was not: There are many known "anomalies" to Einsteinian theory. Einstein's theory is better than—that is, represents progress compared with—Newton's theory *anno 1916* (that is, Newton's laws of dynamics, law of gravitation, the known set of initial conditions, "minus" the list of known anomalies such as Mercury's perihelion) *because* it explained everything that Newton's theory had successfully explained, and it explained also *to some extent* some known anomalies and, in addition, forbade events like transmission of light along straight lines near large masses about which Newton's theory had said nothing but which had been permitted by other well-corroborated scientific theories of the day; moreover, *at least some* of the unexpected excess Einsteinian content was in fact *corroborated* (for instance, by the eclipse experiments). . . .

Let us finally consider how much conventionalism remains in sophisticated falsificationism. Certainly *less* than in naive falsificationism. We need *fewer* methodological decisions. The *"fourth-type decision"* which was essential for the naive version[88] has become completely redundant. To show this we only have to realize that if a scientific theory, consisting of some "laws of nature," initial conditions, auxiliary theories (but without a *ceteris paribus* clause) conflicts with some factual propositions we do not have to decide which—explicit or "hidden"—part to replace. We may try to replace *any* part and only when we have hit on an explanation of the anomaly with the help of some content-increasing change (or auxiliary hypothesis), and nature corroborates it, do we move on to eliminate the "refuted" complex. Thus sophisticated falsification is a slower but possibly safer process than naive falsification.

Let us take an example. Let us assume that the course of a planet differs from the one predicted. Some conclude that this refutes the dynamics and gravitational theory applied: The initial conditions and the *ceteris paribus* clause have been ingeniously corroborated. Other conclude that this refutes the initial conditions used in the calculations: Dynamics and gravitational theory have been superbly corroborated in the last two hundred years and all suggestions concerning further factors in play failed. Yet others conclude that this refutes the underlying assumption that there were no other factors in play except for those which were taken into account: These people may possibly be motivated by the metaphysical principle that any explanation is only approximative because of the infinite complexity of the factors involved in determining any single event. Should we praise the first type as *"critical,"* scold the second type as *"hack,"* and condemn the third as *"apologetic"*? No. We do not need to draw any conclusions about such "refutation." We never reject a specific theory simply by *fiat*. If we have an inconsistency like the one mentioned, we do not have

[86] Keynes [1921], p. 305. But cf. my [1968a], p. 394.

[87] This is Whewell's critical comment on an *ad hoc* auxiliary hypothesis in Newton's theory of light (Whewell [1858], Vol. II, p. 317.)

[88] Cf. *above*, p. 180.

to decide which ingredients of the theory we regard as problematic and which ones as unproblematic: We regard all ingredients as problematic in the light of the conflicting accepted basic statement and try to replace all of them. If we succeed in replacing some ingredient in a "progressive" way (that is, the replacement has more corroborated empirical content than the original), we call it "falsified." . . .

The first, second, and third type decisions of naive falsificationism[89] however, cannot be avoided, but as we shall show, the conventional element in the second decision—and also in the third—can be slightly reduced. We cannot avoid the decision which sort of propositions should be the "observational" ones and which the "theoretical" ones. We cannot avoid either the decision about the truth-value of some "observational propositions." These decisions are vital for the decision whether a problemshift is empirically progressive or degenerating.[90] But the sophisticated falsificationist may at least mitigate the arbitrariness of this second decision by allowing for an *appeal procedure*.

Naive falsificationists do not lay down any such appeal procedure. They accept a basic statement if it is backed up by a well-corroborated falsifying hypothesis,[91] and let it overrule the theory under test—even though they are well aware of the risk.[92] But there is no reason why we should not regard a falsifying hypothesis—and the basic statement it supports—as being just as problematic as a falsified hypothesis. Now how exactly can we expose the problematicality of a basic statement? On what grounds can the protagonists of the "falsified" theory appeal and win?

Some people may say that we might go on testing the basic statement (or the falsifying hypothesis) "by their deductive consequences" until agreement is finally reached. In this testing we deduce—in the same deductive model—further consequences from the basic statement either with the help of the theory under test or some other theory which we regard as unproblematic. Although this procedure "has no natural end," we always come to a point when there is no further disagreement.[93]

But when the theoretician appeals against the verdict of the experimentalist, the appeal court does not normally cross-question the basic statement directly but rather questions the *interpretative theory* in the light of which its truth-value had been established.

One typical example of a series of successful appeals is the Proutians' fight against unfavourable experimental evidence from 1815 to 1911. For decades Prout's theory *T* ("that all atoms are compounds of hydrogen atoms and thus 'atomic weights' of all chemical elements must be expressible as whole numbers") and falsifying "observational" hypotheses, like Stas's "refutation" *R* ("the atomic weight of chlorine is 35·5") confronted each other. As we know, in the end *T* prevailed over *R*.[94]

The first stage of any serious criticism of a scientific theory is to reconstruct, improve, its logical deductive articulation. Let us do this in the case of Prout's theory *vis à vis* Stas's refutation. First of all, we have to realize that in the formulation we just quoted, *T* and *R* were *not* inconsistent (Physicists rarely articulate their theories sufficiently to be pinned down and caught by the critic.) In order to show them up as inconsistent we have to put them in the following form. *T*: "the atomic weight of all pure (homogeneous) chemical elements are multiples of the atomic weight of hydrogen," and *R*: "chlorine is a pure (homogeneous) chemical element and its atomic weight is 35·5." The last statement is in the form of a falsifying hypothesis

[89] Cf. *above*, pp. 178 and 179.

[90] Cf. *above*, p. 183.

[91] Popper [1934], Section 22.

[92] Cf. e.g., Popper [1959], p. 107, footnote *2. Also cf. *above*, pp. 180–181.

[93] This is argued in Popper [1934], Section 29.

[94] Agassi claims that this example shows that we may "stick to the hypothesis in the face of known facts in the hope that the facts will adjust themselves to theory rather than the other way round" [1966], p. 18). But *how* can facts "adjust themselves"? Under which *particular* conditions should the theory win? Agassi gives no answer.

which, if well corroborated, would allow us to use basic statements of the form *B*: "Chlorine *X* is a pure (homogeneous) chemical element and its atomic weight is 35·5"—where *X* is the proper name of a "piece" of chlorine determined, say, by its space-time co-ordinates.

But how well-corroborated is *R*? The first component of it says that R_1: "Chlorine *X* is a pure chemical element." This was the verdict of the experimental chemist after a rigorous application of the "experimental techniques" of the day.

Let us have a closer look at the fine-structure of R_1. In fact R_1 stands for a conjunction of two longer statements, T_1 and T_2. The first statement, T_1, could be this: "If seventeen chemical purifying procedures $P_1, P_2 \ldots P_{17}$ are applied to a gas, what remains will be pure chlorine." T_2 is then: "*X* was subjected to the seventeen procedures $P_1, P_2 \ldots P_{17}$." The careful "experimenter" carefully applied all seventeen procedures: T_2 is to be accepted. But the conclusion that therefore what remained *must* be pure chlorine is a "hard fact" only in virtue of T_1. The experimentalist, while *testing T, applied* T_1. He *interpreted* what he saw in the light of T_1: The result was R_1. *Yet in the monotheoretical model of the explanatory theory under test this interpretative theory does not appear at all.*

But what if T_1, the interpretative theory, is false? Why not "apply" *T* rather than T_1 and claim that atomic weights *must be* whole numbers? Then *this* will be a "hard fact" in the light of *T*, and T_1 will be overthrown. Perhaps additional new purifying procedures must be invented and applied.

The problem is then *not* when we should stick to a "*theory*" in the face of "*known facts*" and when the other way round. The problem is *not* what to do when "theories" clash with "facts." Such a "clash" is only suggested by the "*monotheoretical deductive model.*" Whether a proposition is a "*fact*" or a "*theory*" in the context of a test-situation depends on our methodological decision. "Empirical basis of a theory" is a monotheoretical notion, it is *relative* to some monotheoretical deductive structure. We may use it as first approximation; but in case of "appeal" by the theoretician, we must use a *pluralistic model.* In the pluralistic model the clash is not "between theories and facts" but between two high-level theories: between an *interpretative theory* to provide the facts and an *explanatory theory* to explain them; and the interpretative theory may be on quite as high a level as the explanatory theory. The clash is then not any more between a logically higher-level theory and a lower-level falsifying hypothesis. The problem should not be put in terms of whether a "*refutation*" is real or not. The problem is how to repair an *inconsistency* between the "explanatory theory" under test and the—explicit or hidden—"interpretative" theories; or, if you wish, *the problem is which theory to consider as the interpretative one which provides the "hard" facts and which the explanatory one which "tentatively" explains them.* In a mono-theoretical model we regard the higher-level theory as an *explanatory theory to be judged by the "facts"* delivered from outside (by the authoritative experimentalist): In the case of a clash we reject the explanation.[95] In a pluralistic model we may decide, alternatively, to regard the higher-level theory as an *interpretative theory to judge the "facts"* delivered from outside: In case of a clash we may reject the "facts" as "monsters." In a pluralistic model of testing, several theories—more or less deductively organized—are soldered together.

This argument alone would be enough to show the correctness of the conclusion, which we drew from a different earlier argument, that experiments do not simply overthrow theories, that no theory forbids a state of affairs specifiable in advance.[96] It is not that we propose a theory and Nature may shout NO; rather, we propose a maze of theories, and Nature may shout INCONSISTENT. . . .

The problem is then *shifted* from the old prob-

[95] The decision to use some monotheoretical model is clearly vital for the naive falsificationist to enable him to reject a theory on the *sole* ground of experimental evidence. *It is in line with the necessity for him to divide sharply, at least in a test-situation, the body of science into two: the problematic and the unproblematic.* (Cf. above p. 178–179) *It is only the theory he decides to regard as problematic which he articulates in his deductive model of criticism.*

[96] Cf. *above*, p. 174.

lem of replacing a theory refuted by "facts" to the new problem of how to resolve inconsistencies between closely associated theories. Which of the mutually inconsistent theories should be eliminated? The sophisticated falsificationist can answer that question easily: One has to try to replace first one, then the other, then possibly both, and opt for that new set-up which provides the biggest increase in corroborated content, which provides the most progressive problemshift.[97]

Thus we have established an appeal procedure in case the theoretician wishes to question the negative verdict of the experimentalist. The theoretician may demand that the experimentalist specify his "interpretative theory,"[98] and he may then replace it—to the experimentalist's annoyance—by a better one in the light of which his originally "refuted" theory may receive positive appraisal.[99]

But even this appeal procedure cannot do more than *postpone* the conventional decision. For the verdict of the appeal court is not infallible either. When we decide whether it is the replacement of the "interpretative" or of the "explanatory" theory that produces novel facts, we again must take a decision about the acceptance or rejection of basic statements. But then we have only *postponed*—and possibly *improved*—the decision, not avoided it.[100] The difficulties concerning the empirical basis which confronted "naive" falsificationism cannot be avoided by "sophisticated" falsificationism either. Even if we regard a theory as "factual," that is, if our slow-moving and limited imagination cannot offer an alternative to it (as Feyerabend used to put it), we have to make, at least occasionally and temporarily, decisions about its truth-value. *Even then, experience still remains, in an important sense, the "impartial arbiter"*[101] *of scientific controversy*. We cannot get rid of the problem of the "empirical basis," if we want to learn from experience[102]: but we can make our learning less dogmatic—but also less fast and less dramatic. By regarding some observational theories as problematic we may make our methodology more flexible: but we cannot articulate and include *all* "background knowledge" (or "background ignorance"?) into our critical deductive model. This process is bound to be piecemeal and some conventional line must be drawn at any given time.

There is one objection even to the sophisticated version of methodological falsificationism which cannot be answered without some concession to Duhemian "simplicism." The objection is the so-called "tacking paradox." According to our definitions, adding to a theory completely disconnected low-level hypotheses may constitute a "progressive shift." It is difficult to eliminate such makeshift shifts without demanding that "the additional assertions must be connected with the contradicting assertion *more intimately* than by mere conjunction."[103] This, of course, is a sort of sim-

---

[97] For instance, in our earlier example (cf. *above*, p. 178 ff.) some may try to replace the gravitational theory with a new one and others may try to replace the radio-optics by a new one: We choose the way which offers the more spectacular growth, the more progressive problemshift.

[98] Criticism does not *assume* a fully articulated deductive structure: It creates it. (Incidentally, this is the main message of my [1963–4].)

[99] A classical example of this pattern is Newton's relation to Flamsteed, the first Astronomer Royal. For instance, Newton visited Flamsteed on 1 September 1694, when working full time on his lunar theory; told him to reinterpret some of his data since they contradicted his own theory; and he explained to him exactly how to do it. Flamsteed obeyed Newton and wrote to him on 7 October: "Since you went home, I examined the observations I employed for determining the greatest equations of the earth's orbit, and considering the moon's places at the times. . . ., I find that (*if, as you intimate, the earth inclines on that side the moon that is*) you may abate abt 20" from it . . . ." Thus Newton constantly criticized and corrected Flamsteed's observational theories. Newton taught Flamsteed, for instance, a better theory of the refractive power of the atmosphere; Flamsteed accepted this and corrected his original "data." One can understand the constant humiliation and slowly increasing fury of this great observer, having his data criticized and improved by a man who, on his own confession, made no observations himself: It was this feeling—I suspect—which led finally to a vicious personal controversy.

[100] The same applies to the third type of decision. If we reject a stochastic hypothesis only for one which, in our sense, supersedes it, the exact form of the "rejection rules" becomes *less* important.

[101] Popper [1945], Vol. II, Chapter 23, p. 218.

[102] Agassi is then wrong in his thesis that "observation reports may be accepted as false and hence the problem of the empirical basis is thereby disposed of" (Agassi [1966], p. 20).

[103] Feyerabend [1965], p. 226.

plicity requirement which would assure the continuity in the series of theories which can be said to constitute *one* problemshift.

This leads us to further problems. For one of the crucial features of sophisticated falsificationism is that it replaces the concept of *theory* as the basic concept of the logic of discovery by the concept of *series of theories*. *It is a succession of theories and not one given theory which is appraised as scientific or pseudo-scientific.* But the members of such series of theories are usually connected by a remarkable *continuity* which welds them into *research programmes*. This *continuity*—reminiscent of Kuhnian "normal science"—plays a vital role in the history of science; the main problems of the logic of discovery cannot be satisfactorily discussed except in the framework of a *methodology of research programmes*.

A METHODOLOGY OF SCIENTIFIC RESEARCH PROGRAMMES

I have discussed the problem of objective appraisal of scientific growth in terms of progressive and degenerating problemshifts in series of scientific theories. The most important such series in the growth of science are characterized by a certain *continuity* which connects their members. This continuity evolves from a genuine research programme adumbrated at the start.[104] The programme consists of methodological rules: Some tell us what paths of research to avoid (*negative heuristic*), and others what paths to pursue (*positive heuristic*).

Even science as a whole can be regarded as a huge research programme with Popper's supreme heuristic rule: "Devise conjectures which have more empirical content than their predecessors." Such methodological rules may be formulated, as Popper pointed out, as metaphysical principles.[105] For instance, the *universal* anticonventionalist rule against exception-barring may be stated as the metaphysical principle: "Nature does not allow exceptions." This is why Watkins called such rules "influential metaphysics."[106]

But what I have primarily in mind is not science as a whole, but rather *particular* research programmes, such as the one known as "Cartesian metaphysics." Cartesian metaphysics, that is, the mechanistic theory of the universe—according to which the universe is a huge clockwork (and system of vortices) with push as the only cause of motion—functioned as a powerful heuristic principle. It discouraged work on scientific theories—like [the "essentialist" version of] Newton's theory of action at a distance—which were inconsistent with it (*negative heuristic*). On the other hand, it encouraged work on auxiliary hypotheses which might have saved it from apparent counterevidence—like Keplerian ellipses (positive heuristic).[107]

Negative Heuristic: The "Hard Core" of the Programme.

All scientific research programmes may be characterized by their *"hard core."* The negative heuristic of the programme forbids us to direct the *modus tollens* at this "hard core." Instead, we must use our ingenuity to articulate or even invent "auxiliary hypotheses," which form a *protective belt* around this core, and we must redirect the *modus tollens* to *these*. It is this protective belt of auxiliary hypotheses which has to bear the brunt of tests and get adjusted and re-adjusted, or even completely replaced, to defend the thus-hardened core. A research programme is successful if all this leads to a progressive problemshift, unsuccessful if it leads to a degenerating problemshift.

[104] One may point out that the negative and positive heuristic gives a rough (implicit) definition of the "conceptual framework" (and consequently of the language). The recognition that the history of science is the history of research programmes rather than of theories may therefore be seen as a partial vindication of the view that the history of science is the history of conceptual frameworks or of scientific languages.

[105] Popper [1934], Sections II and 70. I use "metaphysical" as a technical term of naive falsificationism: A contingent proposition is "metaphysical" if it has no "potential falsifiers."

[106] Watkins [1958]. . . .

[107] For this Cartesian research programme, cf. Popper [1958] and Watkins [1958], pp. 350–1.

The classical example of a successful research programme is Newton's gravitational theory, possibly the most successful research programme ever. When it was first produced, it was submerged in an ocean of "anomalies" (or, if you wish, "counterexamples"[108]),and opposed by the observational theories supporting these anomalies. But Newtonians turned, with brilliant tenacity and ingenuity, one counter-instance after another into corroborating instances, primarily by overthrowing the original observational theories in the light of which this "contrary evidence" was established. In the process they themselves produced new counter-examples which they again resolved. They "turned each new difficulty into a new victory of their programme."[109]

In Newton's program the negative heuristic bids us to divert the *modus tollens* from Newton's three laws of dynamics and his law of gravitation. This "core" is "irrefutable" by the methodological decision of its protagonists: Anomalies must lead to changes only in the "protective" belt of auxiliary, "observational" hypotheses and initial conditions.[110]

I have given a contrived micro-example of a progressive Newtonian problemshift.[111] If we analyse it, it turns out that each successive link in this exercise predicts some new fact; each step represents an increase in empirical content: The example constitutes a *consistently progressive theoretical shift.* Also, each prediction is in the end verified; although on three subsequent occasions they may have seemed momentarily to be "refuted."[112] While "theoretical progress" (in the sense here described) may be verified immediately . . . , "empirical progress" cannot, and in a research programme we may be frustrated by a long series of "refutations" before ingenious and lucky content-increasing auxiliary hypotheses turn a chain of defeats—*with hindsight*—into a resounding success story, either by revising some false "facts" or by adding novel auxiliary hypotheses. We may then say that we must require that each step of a research programme be consistently content-increasing: that each step constitute a *consistently progressive theoretical problemshift.* All we need in addition to this is that at least every now and then the increase in content should be seen to be retrospectively corroborated: The programme as a whole should also display an *intermittently progressive empirical shift.* We do not demand that each step produce *immediately* an *observed* new fact. Our term *intermittently* gives sufficient *rational* scope for dogmatic adherence to a programme in face of *prima facie* "refutations."

The idea of "negative heuristic" of a scientific research programme rationalizes classical conventionalism to a considerable extent. We may rationally decide not to allow "refutations" to transmit falsity to the hard core as long as the corroborated empirical content of the protecting belt of auxiliary hypotheses increases. But our approach differs from Poincaré's justificationist conventionalism in the sense that, unlike Poincaré's, we maintain that if and when the programme ceases to anticipate novel facts, its hard core might have to be abandoned: that is, *our* hard core, unlike Poincaré's, may crumble under certain conditions. In this sense we side with Duhem who thought that such a possibility must be allowed for;[113] but for Duhem the reason for such crumbling is purely *aesthetic,*[114] while for us it is mainly *logical and empirical.*

Positive Heuristic: The Construction of the "Protective Belt" and the Relative Autonomy of Theoretical Science.

Research programmes, besides their negative heuristic, are also characterized by their positive heuristic.

---

[108] For the clarification of the concepts of "counterexample" and "anomaly" cf. *above,* p. 180. . . .

[109] Laplace [1796], Livre IV, Chapter ii.

[110] The actual hard core of a programme does not actually emerge fully armed like Athene from the head of Zeus. It develops slowly, by a long, preliminary process of trial and error. In this paper this process is not discussed.

[111] Cf. *above,* pp. 174–175. For *real* examples, cf. my [1970].

[112] The "refutation" was each time successfully diverted to "hidden lemmas"; that is, to lemmas emerging, as it were, from the *ceteris paribus* clause.

[113] Cf. *above,* p. 177.

[114] Ibid.

Even the most rapidly and consistently progressive research programmes can digest their "counter-evidence" only piecemeal: Anomalies are never completely exhausted. But it should not be thought that yet unexplained anomalies—"puzzles" as Kuhn might call them—are taken in random order, and the protective belt built up in an eclectic fashion, without any preconceived order. The order is usually decided in the theoretician's cabinet, independently of the *known* anomalies. Few theoretical scientists engaged in a research programme pay undue attention to "refutations." They have a long-term research policy which anticipates these refutations. This research policy, or order of research, is set out—in more or less detail—in the *positive heuristic* of the research programme. The negative heuristic specifies the "hard core" of the programme which is "irrefutable" by the methodological decision of its protagonists; the positive heuristic consists of a partially articulated set of suggestions or hints on how to change, develop the "refutable variants" of the research programme, how to modify, sophisticate, the "refutable" protective belt.

The positive heuristic of the programme saves the scientist from becoming confused by the ocean of anomalies. The positive heuristic sets out a programme which lists a chain of ever more complicated *models* simulating reality: The scientist's attention is riveted on building his models following instructions which are laid down in the positive part of his programme. He ignores the *actual* counterexamples, the available "*data*."[115] Newton first worked out his programme for a planetary system with a fixed point-like sun and one single point-like planet. It was in this model that he derived his inverse square law for Kepler's ellipse. But this model was forbidden by Newton's own third law of dynamics; therefore, the model had to be replaced by one in which both sun and planet revolved round their common centre of gravity. This change was not motivated by any observation (the data did not suggest an "anomaly" here) but by a theoretical difficulty in developing the programme. Then he worked out the programme for more planets as if there were only heliocentric but no interplanetary forces. Then he worked out the case where the sun and planets were not mass-points but mass-*balls*. Again, for this change he did not *need* the observation of an anomaly; infinite density was forbidden by an (inarticulated) touchstone theory, therefore planets *had* to be extended. This change involved considerable mathematical difficulties, held up Newton's work—and delayed the publication of the *Principia* by more than a decade. Having solved this "puzzle," he started work on *spinning balls* and their wobbles. Then he admitted interplanetary forces and started work on *perturbations*. At this point he started to look more anxiously at the facts. Many of them were beautifully explained (qualitatively) by this model, many were not. It was then that he started to work on *bulging* planets, rather than round planets, etc.

Newton despised people who, like Hooke, stumbled on a first naive model but did not have the tenacity and ability to develop it into a research programme, and who thought that a first version, a mere aside, constituted a "discovery." He held up publication until his programme had achieved a remarkable progressive shift.[116]

Most, if not all, Newtonian "puzzles," leading to a series of new variants superseding each other, were foreseeable at the time of Newton's first naive model and no doubt Newton and his colleagues *did* foresee them: Newton must have been

---

[115] If a scientist (or mathematician) has a positive heuristic, he refuses to be drawn into observation. He will "lie down on his couch, shut his eyes and forget about the data." (Cf. my [1963–4], especially pp. 300 ff., where there is a detailed case study of such a programme.) Occasionally, of course, he will ask Nature a shrewd question: he will then be encouraged by Nature's *YES*, but not discouraged by its *NO*.

[116] Reichenbach, following Cajori, gives a different explanation of what delayed Newton in the publication of his *Principia*: "To his disappointment he found that the observational results disagreed with his calculations. Rather than set any theory, however beautiful, before the facts, Newton put the manuscript of his theory into his drawer. Some twenty years later, after new measurements of the circumference of the earth had been made by a French expedition, Newton saw that the figures on which he had based his test were false and that the improved figures agreed with his theoretical calculation. It was only after this test that he published his law . . . . The story of Newton is one of the most striking illustrations of the method of modern science" (Reichenbach [1951], pp. 101–2). Feyerabend criticizes Reichenbach's account (Feyerabend [1965], p. 229), but does not give an alternative *rationale*.

fully aware of the blatant falsity of his first variants.[117] Nothing shows the existence of a positive heuristic of a research programme clearer than this fact: This is why one speaks of "models" in research programmes. A *"model"* is a set of initial conditions (possibly together with some of the observational theories) which one knows is *bound* to be replaced during the further development of the programme, and one even knows, more or less, how. This shows once more how irrelevant "refutations" of any specific variant are in a research programme: Their existence is fully expected, the positive heuristic is there as the strategy both for predicting (producing) and digesting them. Indeed, if the positive heuristic is clearly spelt out, the difficulties of the programme are mathematical rather than empirical.[118]

One may formulate the "positive heuristic" of a research programme as a "metaphysical" principle. For instance, one may formulate Newton's programme like this: "The planets are essentially gravitating spinning-tops of roughly spherical shape." This idea was never *rigidly* maintained: The planets are not *just* gravitational, they have also, for example, electromagnetic characteristics which may influence their motion. Positive heuristic is thus in general more flexible than negative heuristic. Moreover, it occasionally happens that when a research programme gets into a degenerating phase, a little revolution or a *creative shift* in its positive heuristic may push it forward again.[119] It is better therefore to separate the "hard core" from the more flexible metaphysical principles expressing the positive heuristic.

Our considerations show that the positive heuristic forges ahead with almost complete disregard of "refutations": It may seem that it is the *"verifications"*[120] rather than the refutations which provide the contact points with reality. Although one must point out that any "verification" of the $n + 1$-th version of the programme is a refutation of the nth version, we cannot deny that *some* defeats of the subsequent versions are always foreseen: It is the "verifications" which keep the programme going, recalcitrant instances notwithstanding.

We may appraise research programmes, even after their "elimination," for their *heuristic power:* How many new facts did they produce, how great was "their capacity to explain their refutations in the course of their growth"?[121]

(We may also appraise them for the stimulus they gave to mathematics. The real difficulties for the theoretical scientist arise rather from the *mathematical difficulties* of the programme than from anomalies. The greatness of the Newtonian programme comes partly from the development—by Newtonians—of classical infinitesimal analysis which was a crucial precondition of its success.)

Thus the methodology of scientific research programmes accounts for the *relative autonomy of theoretical science:* a historical fact whose rationality cannot be explained by the earlier falsificationists. Which problems scientists working in powerful research programmes rationally choose is determined by the positive heuristic of the programme rather than by psychologically worrying (or technologically urgent) anomalies. The anomalies are listed but shoved aside in the hope that they will turn, in due course, into corroborations of the programme. Only those scientists have to rivet their attention on anomalies who are either engaged in trial-and-error exercises . . . or who work in a degenerating phase of a research programme when the positive heuristic ran out of steam. (All this, of course, must sound repugnant to naive falsificationists who hold that once a theory is "refuted" by experiment (by *their* rule book), it is irrational (and dishonest) to develop it further: One has to replace the old "refuted" theory by a new, unrefuted one.) . . .

---

117 For a further discussion of Newton's research programme, cf. my [1970].

118 For this point cf. Truesdell [1960].

119 Soddy's contribution to Prout's programme or Pauli's to Bohr's (old quantum theory) programme are typical examples of such creative shifts.

120 A "verification" is a corroboration of excess content in the expanding programme. But, of course, a "verification" does not *verify* a programme: It shows only its heuristic power.

121 Cf. my [1963–4], pp. 324–30. Unfortunately in 1963–4 I had not yet made a clear terminological distinction between theories and research programmes, and this impaired my exposition of a research programme in informal, quasi-empirical mathematics. There are fewer such shortcomings in my [1971].

REFERENCES

Agassi [1959]: "How are Facts Discovered?" *Impulse*, **3,** No. 10, pp. 2–4.

Agassi [1966]: "Sensationalism," *Mind*, N.S. **75,** pp. 1–24.

Agassi [1968]: "The Novelty of Popper's Philosophy of Science." *International Philosophical Quarterly*, **8,** pp. 442–63.

Agassi [1969]: "Popper on Learning from Experience," in Rescher (*ed.*): *Studies in the Philosophy of Science*, 1969.

Braithwaite [1938]: "The Relevance of Psychology to Logic," *Aristotelian Society Supplementary Volumes*, **17,** pp. 19–41.

Braithwaite [1953]: *Scientific Explanation*, 1953.

Dryer [1906]: *History of the Planetary Systems from Thales to Kepler*, 1906.

Duhem [1905]: *La Théorie Physique, Son Objet et Sa Structure*, 1905. English translation of the second (1914) edition: *The Aim and Structure of Physical Theory*, 1954.

Einstein [1949]: "Autobiographical Notes," in Schilpp (*ed.*): *Albert Einstein, Philosopher-Scientist*, **1,** pp. 2–95.

Feyerabend [1965]: "Reply to Criticism," in Cohen and Wartofsky (*eds.*): *Boston Studies in the Philosophy of Science*, II, pp. 223–61.

Feyerabend [1969]: "Problems of Empiricism II," in Colodny (*ed.*): *The Nature and Function of Scientific Theory*, 1969.

Feyerabend [1970]: "Against Method," *Minnesota Studies for the Philosophy of Science*, **4,** 1970.

Hempel [1952]: "Some Theses on Empirical Certainty," *The Review of Metaphysics*, **5,** pp. 620–1.

Keynes [1921]: *A Treatise on Probability*, 1921.

Koestler [1959]: *The Sleepwalkers*, 1959.

Lakatos [1962]: "Infinite Regress and the Foundations of Mathematics," *Aristotelian Society Supplementary Volume*, **36,** pp. 155–84.

Lakatos [1963–4]: "Proofs and Refutations," *The British Journal for the Philosophy of Science*, **14,** pp. 1–25, 120–39, 221–43, 296–342.

Lakatos [1968*a*]: "Changes in the Problem of Inductive Logic," in Lakatos (*ed.*): *The Problem of Inductive Logic*, 1968, pp. 315–417.

Lakatos [1968*b*]: "Criticism and the Methodology of Scientific Research Programmes," in *Proceedings of the Aristotelian Society*, **69,** pp. 149–86.

Lakatos [1970]: *The Changing Logic of Scientific Discovery*, 1970.

Lakatos [1971]: *Proofs and Refutations and Other Essays in the Philosophy of Mathematics*, 1971.

Laplace [1796]: *Exposition du Système du Monde*, 1796.

Leibniz [1678]: Letter to Conring, 19.3.1678.

Le Roy [1899]: "Science et Philosophie," *Revue de Métaphysique et de Morale*, **7,** pp. 375–425, 503–62, 706–31.

Le Roy [1901]: "Un Positivisme Nouveau," *Revue de Métaphysique et de Morale*, **9,** pp. 138–53.

Medawar [1967]: *The Art of the Soluble*, 1967.

Milhaud [1896]: "*La Science Rationnelle*," *Revue de Métaphysique et de Morale*, **4,** pp. 280–302.

Mill [1843]: *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*, 1843.

Musgrave [1969*a*]: *Impersonal Knowledge*, Ph.D. Thesis, University of London, 1969.

Musgrave [1969*b*]: Review of Ziman's "Public Knowledge: An Essay Concerning the Social Dimensions of Science," in *The British Journal for the Philosophy of Science*, **20,** pp. 92–4.

Neurath [1935]: "Pseudorationalismus der Falsifikation," *Erkenntnis*, **5,** pp. 353–65.

Poincaré [1891]: "Les géométries non euclidiennes," *Revue des Sciences Pures et Appliquées*, **2,** pp. 769–74.

Poincaré [1902]: *La Science et l'Hypothèse*, 1902.

Popper [1933]: "Ein Kriterium des empirischen Charakters theoretischer Systeme," *Erkenntnis*, **3,** pp. 426–7.

Popper [1934]: *Logik der Forschung*, 1935 (expanded English edition: Popper [1959]).

Popper [1935]: "Induktionslogik und Hypothesenwahrscheinlichkeit," *Erkenntnis*, **5,** pp. 170–2.

Popper [1945]: *The Open Society and its Enemies*, I–II, 1945.

Popper [1957]: "The Aim of Science," *Ratio*, **1,** pp. 24–35.

Popper [1958]: "Philosophy and Physics"; published in *Atti del XII Congresso Internazionale di Filosofia*, Vol. **2,** 1960, pp. 363–74.

Popper [1959]: *The Logic of Scientific Discovery*, 1959.

Popper [1963]: *Conjectures and Refutations*, 1963.

Popper [1968]: "Remarks on the Problems of Demarcation and Rationality," in Lakatos and Musgrave (*eds.*): *Problems in the Philosophy of Science*, 1968, pp. 88–102.

Reichenbach [1951]: *The Rise of Scientific Philosophy*, 1951.

Russell [1943]: "Reply to Critics," in Schilpp (*ed.*): *The Philosophy of Bertrand Russell*, 1943, pp. 681–741.

Truesdell [1960]: "The Program toward Rediscovering the Rational Mechanics in the Age of Reason," *Archive of the History of Exact Sciences*, **1,** pp. 3–36.

Watkins [1957]: "Between Analytic and Empirical," *Philosophy*, **32,** pp. 112–31.

Watkins [1958]: "Influential and Confirmable Metaphysics," *Mind*, N.S. **67,** pp. 344–65.

Watkins [1960]: "When are Statements Empirical?" *British Journal for the Philosophy of Science*, **10,** pp. 287–308.

Whewell [1837]: *History of the Inductive Sciences, from the Earliest to the Present Time*. Three volumes, 1837.

Whewell [1840]: *Philosophy of the Inductive Sciences, Founded upon their History*. Two volumes, 1840.

Whewell [1858]: *Novum Organon Renovatum*. Being the second part of the philosophy of the inductive sciences. Third edition, 1858.

Wisdom [1963]: "The Refutability of 'Irrefutable' Laws," *The British Journal for the Philosophy of Science*, **13**, pp. 303–6.