A PHILOSOPHICAL APPROACH TO MOND

David Merritt

A PHILOSOPHICAL APPROACH TO MOND

Assessing the Milgromian Research Program in Cosmology

DAVID MERRITT
Rochester Institute of Technology, New York

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108492690
DOI: 10.1017/9781108610926

First published 2020

Printed in the United Kingdom by TJ International Ltd, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-49269-0 Hardback

Contents

Preface

Some species of ant, bee and termite are 'eusocial': they live in colonies of overlapping generations in which all the offspring are produced from one or a few individuals (the queen bee, for instance) while the other, non-reproducing members of the colony devote their lives to selfless behavior, protecting the colony and collectively rearing the young. Explanations for the origin of eusocial behavior start from the observation that eusocial insects share a large fraction of their DNA with the other members of the colony – in the case of honeybees, the degree of relatedness is 75%. Worker bees can pass on more of their genetic material by helping their 'sisters' than by having offspring of their own, and natural selection responds to this state of affairs by endowing them with the motivation to act altruistically toward the other bees in the colony.

In May 1976, the biologist Richard Alexander gave a lecture at Northern Arizona University on eusociality in which he tried to explain why it had never evolved in vertebrates. As a thought experiment, he speculated on what a eusocial mammal might be like. The need to accommodate a large and growing colony would favor subterranean rodents. He predicted that the ideal niche would be tropical and that the burrowing rodents would prefer to live in heavy clay soil that is inaccessible to most predators, and to feed on large tubers. And of course there would be one 'queen' rodent that gave birth to all of the offspring, who would behave altruistically toward each other. After the lecture, Alexander was surprised to be told by a member of the audience that he had just given a perfect description of the naked mole-rat, a species native to East Africa and which had just begun to be studied by biologists (Sherman et al., 1991, p. vii–viii).

Scientists tend to be very impressed by episodes like this. It is hard to believe that the theory on which Alexander's prediction was based – in this case, Darwin's theory of evolution by natural selection – could be very wrong if it correctly predicts something as *a priori* unlikely as a naked mole-rat.

There is a research tradition in cosmology that has been repeatedly successful in just this way, correctly predicting facts and relations – some quite surprising – in advance of their discovery. I am not referring here to the standard model of cosmology, the one that you will find in the textbooks. The theory I have in mind goes under a number of names; the most common is 'MOND,' which stands for 'MOdified Newtonian Dynamics,' but many researchers prefer the name 'Milgromian dynamics,' since the theory was originated by the astrophysicist Mordehai Milgrom.

Milgrom published the foundational postulates of his new theory in a set of three papers in 1983. At the time, the standard cosmological model was facing a major crisis – or, in the language of philosopher Karl Popper, a 'falsifying instance.' Observations of spiral galaxies like the Milky Way had revealed that the motion of stars and gas in their outskirts always fails to match with the predictions of Newtonian gravity: orbital speeds are always greater than predicted, sometimes much greater, and they are never observed to fall off with distance as Newton's equations generically predict. Anyone who was doing research in astrophysics at that time (as I was) will remember how quickly the community closed ranks and agreed on a consensus explanation for the anomaly: galaxies, it was postulated, are surrounded by 'dark matter,' which does not interact with radiation but which does generate gravitational force. Standard-model cosmologists still universally *assume* that the dark matter in every galaxy has whatever spatial distribution is needed to reconcile the observed motions with Newton's laws.

Milgrom was one of the scientists who expressed reservations about the dark matter hypothesis. What he found most impressive was not the anomalously high orbital speeds in the outskirts of galaxies, but rather the fact that every galactic rotation curve (the plot of orbital speed versus distance from the center) is 'asymptotically flat': it tends toward a constant value, different in each galaxy, and remains at this value as far out as observations permit. This is extremely hard to understand under the dark matter hypothesis, since every galaxy has a different history of formation and evolution and the dark matter would need to repeatedly redistribute itself to maintain that flat rotation curve. Milgrom proposed a different, and quite bold, response to the rotation-curve anomaly: a modification to Newton's laws. Many such modifications could do the trick, but Milgrom singled out one: he proposed that Newton's relation between gravitational force and acceleration should be modified in regions where the gravitational acceleration (that is, the gravitational force per unit of mass) falls below a certain, universal value. Milgrom labelled this new constant a_0 and estimated a value of order 10^{-10} m s^{-2} – so small that the proposed modification to the laws of motion would be practically undetectable anywhere in the solar system; it would only become important in regions of very small gravitational force like the outskirts of galaxies.

On its face, Milgrom's explanation for the rotation-curve anomaly is neither more nor less ad hoc than the dark matter postulate. Both are examples of what philosophers of science call 'auxiliary hypotheses': assumptions that are added to a theory in order to (in this case) reconcile it with falsifying data. There are likely to be an infinite number of auxiliary hypotheses that can account for any given anomaly. How do we decide which (if any) is correct?

This state of affairs has arisen many times in science, and philosophers of science have come up with a set of criteria. Clearly it is not enough for an auxiliary hypothesis to explain facts that are already known; as philosopher Elie Zahar (1973, p. 103) put it, theories can always be "cleverly engineered to yield the known facts." To be acceptable, an auxiliary hypothesis should also predict some *new* facts: the more unlikely the predictions (in the light of the pre-existing theory), the better. (Remember the naked mole-rat.) And ideally, at least some of those novel predictions should be confirmed by observation or experiment – this gives us confidence that we are moving in the direction of the *correct* theory, which, after all, *only* makes correct predictions.

How well does the galactic dark matter hypothesis meet these requirements? In explaining the rotation curve of the Milky Way galaxy, that hypothesis does make a novel prediction: that there should be dark matter near the Sun with a density (mass per unit volume) that is approximately known. The particles that make up the dark matter (if particles they be) are passing continuously through every laboratory on the Earth and could be detected. Attempts to verify this prediction (so-called 'direct detection' experiments) got underway shortly after the dark matter hypothesis was agreed upon, in the early 1980s, and they have continued unabated since then; the detectors currently in use are about ten million times more sensitive than those of the 1980s. But all attempts to detect the dark matter particles have failed: no one has ever observed anything that might reasonably be interpreted as the signal of a dark matter particle passing through their detector.

The situation is very different for Milgrom's hypothesis. Already in his first papers from 1983, Milgrom wrote down a number of novel predictions that follow from his postulates. For instance, he showed that his modification to Newton's laws predicts not only that rotation curves should be asymptotically flat (that result was built into the postulates, just as it is built into the dark matter hypothesis), but also that there should be a universal relation between the orbital speed in the outer parts of a galaxy and the galaxy's total (not dark!) mass. No one had even thought to look for such a relation before Milgrom predicted it; no doubt because – according to the standard model – it is the dark matter, not the ordinary matter, that sets the rotation speed. But Milgrom's prediction has been beautifully confirmed – a splendid example of a verified, novel prediction.[1]

Note that this prediction of Milgrom's is *refutable*: it could, in principle, have been found to be incorrect. By contrast, the standard-model prediction that dark matter particles are passing through an Earth-bound laboratory is *not* refutable, since nothing whatsoever is known about the properties of the putative dark

[1] This is the 'baryonic Tully–Fisher relation'. See Figure 4.1 and the discussion in Chapter 4.

particles. A failure to detect them might simply mean that their cross section for interaction with normal matter is very small (and that is, in fact, the explanation that standard-model cosmologists currently promote). On these grounds, as well, Milgrom's hypothesis 'wins': it is epistemically the preferred explanation.

Of course, that is not the same as saying that Milgrom's hypothesis is *correct*. I will not, in fact, be arguing that – although I know of nothing that would preclude such a conclusion. My goal is more modest: to assess the degree to which the Milgromian research program is progressive.

The terms 'research program' and 'progressive' will be familiar to philosophers of science but not to most scientists – at least, not with the specific meanings that philosophers attach to them. Both terms are due to Imre Lakatos, a student and colleague of Karl Popper. Lakatos recognized (as did Popper, and Thomas Kuhn) that scientific theories evolve, and they do so in characteristic ways. Typically there is a fixed set of assumptions, which Lakatos called the 'hard core' (and which Kuhn, at least sometimes, referred to as a 'paradigm'); for instance, in the standard cosmological model, the hard core contains the assumption that the general theory of relativity is correct. When a prediction of a theory is shown to be incorrect – when the theory is 'falsified', to use Popper's term – scientists, Lakatos said, rarely modify the hard core; instead they are likely to add an auxiliary hypothesis that targets the anomaly and 'explains' it, leaving the hard core intact.

Since theories change over time, Lakatos argued that the proper unit of appraisal is not a single theory, but rather the evolving *set* of theories that share the same hard core postulates over time – what Lakatos called a 'research program.' To the extent that this is correct, the central question for epistemologists of science is no longer "Has this theory been falsified?" ("All theories", said Lakatos, "are born refuted and die refuted") but rather "Is this research program progressing or degenerating?" Based on his analysis of the historical record, and being guided whenever possible by Popper's epistemic insights, Lakatos identified two conditions that characterize theory change in successful research programs. First, Lakatos found that changes to a theory should not be ad hoc: they should enlarge its scope and create the potential for new predictions – some of which, ideally, should be confirmed. Indeed, Lakatos argued that the *only* experiments or observations of any evidentiary value were those that targeted *novel* predictions ("the only relevant evidence is the evidence anticipated by a theory"). Second, Lakatos noted that successful research programs tend to develop autonomously, and not simply in response to anomalies. "A research programme is said to be *progressing* as long as its theoretical growth anticipates its empirical growth, that is, as long as it keeps predicting novel facts with some success," he wrote. Whereas a stagnating, or "degenerating," research program is one that "gives only *post hoc* explanations either of chance discoveries or of facts anticipated by, and discovered in, a rival programme" (Lakatos, 1971, p. 112).

Lakatos's criteria should give pause to anyone familiar with the history of the standard cosmological model since about 1980. The hypotheses in that model

1

The Epistemology of Science

> The concordance model [of cosmology] is now well established, and there seems little room left for any dramatic revision of this paradigm.
>
> *(Olive* et al. *(2014))*
>
> The evidence for the dark matter of the hot big bang cosmology is about as good as it gets in natural science.
>
> *(Peebles (2015))*
>
> The trouble about people – uncritical people – who hold a theory is that they are inclined to take everything as supporting or 'verifying' it, and nothing as refuting it.
>
> *(Popper (1983))*

There is a tendency, among both scientists and non-scientists, to assume that our current scientific theories are correct in some fundamental sense: that they embody deep and established truths about the physical universe. No one denies that the theories might benefit from further refinement, particularly in regimes where they have not been well tested, and everyone would acknowledge that there are things in the universe that remain to be discovered and understood. But it is widely assumed that the theories of physics, chemistry and biology that are set out in the current textbooks are unlikely to change in any fundamental way. After all, the argument goes, these theories are the basis for the spectacular material progress of the modern world: for the design of airplanes and computers, the production of serums and antibiotics, the manufacture of plastics and synthetic fibers, the successful prediction of spacecraft trajectories and the weather. It is almost impossible to imagine (the argument goes) that these theories could be so successful unless they were essentially correct.[1]

But the history of science suggests otherwise. Almost all of the theories that were at one time viewed as correct have been abandoned. And what is even more

[1] "we are strikingly good at making science-based interventions in nature. . . . this success in intervention is incomprehensible unless we suppose that the claims we are putting to work in our practical activities are correct (or, at least, approximately correct)" (Kitcher, 1995, p. 659); "it is reasonable to believe that the successful theories in mature science – the unified theories that explain the phenomena without ad hoc assumptions . . . are, if you like, approximately true" (Worrall, 2007, p. 153–154).

striking is the manner in which theories change. There are certainly periods, within any scientific discipline, when the dominant theory undergoes only gradual modifications, without much change to the underlying assumptions. But such periods tend to last only so long; they are separated by revolutions during which the old assumptions are thrown out and a radically new set are brought in. As every student of physics knows, there were a number of such episodes in the early part of the twentieth century: classical mechanics and electromagnetism were replaced by quantum electrodynamics, Newton's theory of gravity and motion was replaced by Einstein's. The new theories were not simply improvements over the theories they replaced. The changes were so radical that even basic concepts like mass and time altered their meanings in fundamental ways.

That is not to deny that there are aspects of scientific progress that are genuinely cumulative. The universe is vast, and the longer we observe it, the more we learn about its composition and structure. Additions to knowledge of this sort are what the popular science writers usually have in mind when they talk about 'scientific discoveries.' But what lends science its particular prestige is not the accumulation of knowledge about what exists: it is the (apparent) ability of science to make correct predictions about things that no one had previously observed. Scientific theories contain *universal hypotheses*: statements or laws (often presented in mathematical language) that are claimed to be valid at all places and for all times, and that can be used to generate predictions even in situations that have never been encountered before. For instance: 'The gravitational force between two point objects varies as the inverse square of their separation'; 'the entropy of an isolated system never decreases'; 'the wavelength of a particle varies inversely with its momentum.'

Where do such hypotheses come from? It is tempting to believe that they are arrived at through induction: that they are generalizations from what is observed. But a few minutes' thought shows that that can not possibly be correct. Discrete instances do not imply universal laws; a finite set of observations is always consistent with an infinite number of different theories. Not only is induction insufficient to the task: it is fair to say that induction does not exist. The fallacy of induction has been discovered and rediscovered many times, going back at least to the fourth century BCE and the Greek philosopher Pyrrho of Elis.[2] Modern discussions of the 'problem of induction' usually adopt the formulation of the eighteenth century philosopher David Hume: "we have no reason to draw any inference concerning any object beyond those of which we have had experience" (Hume, 1739–40/1978, Book I, Part III, section xii). As an illustration, Hume invoked the impossibility of predicting the future: "For all inferences from experience suppose, as their foundation, that the future will resemble the past" (Hume, 1748/1975, section 4.2, 37–38).

[2] Pyrrho of Elis (*c*. 360–275 BCE) left no writings; the sole surviving texts from the Pyrrhonian movement are those of Sextus Empiricus (*c*. 160–210 BCE). Richard H. Popkin (2003) traces the history of Pyrrhonian skepticism from its revival in fifteenth-century Europe until the early eighteenth century.

Just because the Sun rose yesterday, and on all previous days for which records exist, there is no logical basis to assume that it will rise tomorrow (and of course it may not).

Hume's 'problem of tomorrow', to adopt the phrase of Karl Popper (1983, Part I, 4.III) – the lack of any basis in logic for assuming the regularity of nature – is one aspect of the problem of induction. But what is equally relevant to the epistemology of science is a different aspect: the logical impossibility of generalizing from a limited set of observations to an unrestrictedly general law, and (what is almost the same thing) the impossibility of *verifying* a universal law (whatever its provenance) given known instances of its success.

The essential point here is that even an incorrect theory can generate correct predictions. Take a simple example: today is Saturday, and someone proposes the hypothesis "Today is Sunday." That hypothesis is false, but from it necessarily follow any number of true statements, including "Today is not Monday," "The English word for this day of the week begins with the letter S," "It is illegal to sell packaged liquor after 9:00 pm today in Milwaukee" etc. Anyone so inclined could confirm the correctness of an unlimited number of such predictions ("It is not noon on Monday," "It is not 12:01 on Monday" etc.). This example may seem too simple or contrived to be relevant to the justification of scientific theories. But then, consider the fact that for two hundred years Newton's theory of gravity and motion was found again and again to yield accurate predictions, even to the extent of correctly predicting the existence and location of a new planet (Neptune). And yet we now know (or at least believe) that Newton was wrong: not wrong in a minor or trivial way, but deeply, fundamentally, conceptually wrong. Einstein's theory correctly predicts the same facts as Newton's, but interprets them as instances of a quite different set of hypotheses. And it would be foolish to assume that Einstein's theory, as well-corroborated as it is,[3] will not itself be replaced one day by another theory, perhaps a theory that differs as much from Einstein's as Einstein's differs from Newton's.

These arguments are convincing enough, but they do not bring us any closer to explaining the success of science. And if the inductive method – which since the time of Francis Bacon was widely (though mistakenly) seen as the principle that separates science from non-science[4] – does not exist, then what basis do we have for calling some theories 'scientific' and others just speculation?

[3] Here and throughout this book, 'corroborate' has the meaning adopted by Karl Popper after about 1958, roughly, 'provide evidential support for' (Popper, 1983, section 29). 'Corroboration' differs from 'confirmation'; the latter implies demonstration or proof of correctness. Following Popper and Hume, it is reasonable to believe that a *prediction* of a theory can be confirmed, but theories themselves can only be corroborated, never confirmed. E.g. Magee (1997, p. 188): "it is possible sometimes to be sure of a direct observation, but not of the explanatory framework that explains it."

[4] E.g. Lakatos (1974, p. 161): "at least among philosophers of science, Baconian method [i.e. inductivist logic of discovery] is now only taken seriously by the most provincial and illiterate."

Karl Popper, in his *The Logic of Scientific Discovery* (1959), claimed to have found the answer to both questions.[5] Popper granted the correctness of Hume's analysis: induction, he said, does not exist, and therefore it can be invoked neither as a basis for the growth of knowledge, nor as a criterion of demarcation between science and pseudoscience. But, he said, induction is not needed. Popper began by emphasizing the logical asymmetry between proof and disproof. While no number of observations can ever prove the validity of a universal law, a single observation that *conflicts* with the law is sufficient to *dis*prove it. The hypothesis 'All swans are white' can not be true if even a single black swan exists.

Of course, this argument – what logicians call *modus tollens* – was well known to Hume. But Popper went a big step further. All knowledge, said Popper (still in agreement with Hume), is uncertain and must always remain so. But if a hypothesis is testable, there exists at least the possibility that it can be shown to be incorrect and replaced with another, better one: "For it may happen that our test statements may refute some – but not all – of the competing theories; and since we are searching for a true theory, we shall prefer those whose falsity has not been established" (Popper, 1972, p. 8). Popper emphasized that the most useful tests are those carried out with the *intent* of falsifying a theory; for instance, experiments that test a prediction that conflicts with the experimenter's prior expectations. As long as a new theory holds up to such tests, Popper said, we are justified in considering the theory viable. Whereas if a prediction is shown to be false, the theory has been disproved, and it can be replaced. In this manner, via "conjectures and refutations," knowledge can grow.

Popper's view of epistemology is called 'critical rationalism.'[6] Critical rationalism is opposed to – for instance – inductivism, and to logical positivism, the belief that the only meaningful statements are those that are *verifiable* through observation. Critical rationalists deny that theories are verifiable. They assert that theories should be judged on the basis of how well they stand up to attempts to refute them.

But where do the hypotheses that we are testing come from? Popper was adamant on this point: it simply does not matter. Theories can come from anywhere.[7] What

[5] The 1959 publication date of *The Logic of Scientific Discovery*, the English translation of *Logik der Forschung*, is misleading. The German text was published in 1934. The original manuscript, in two volumes, was titled *Die Beiden Grundprobleme der Erkenntnistheorie* and was completed in early 1932; it was scheduled for publication in 1933 but the publisher (Springer) objected to its length. A new manuscript, which consisted of extracts from the two unpublished volumes, was also rejected by the publisher. Popper (1974, p. 67) gives credit to his uncle, Walter Schiff, who "ruthlessly cut about half the text" resulting in the 1934 publication of *Logik der Forschung*. Popper (1972, p. 1, n. 1) has said that he discovered the solution to the problem of induction around 1927.

[6] Here Popper uses 'rationalism' to mean the opposite of 'irrationalism' (and not the opposite of 'empiricism'); he defines it as "an attitude of readiness to listen to critical arguments and to learn from experience" (Popper, 1945, p. 225). Paul Feyerabend (1975, p. 172), in a discussion of Popperian epistemology, writes: "rational discussion consists in the attempt to criticize, and not in the attempt to prove or to make probable."

[7] E.g. Popper (1959, p. 32): "there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process." Peter Urbach (1978, p. 102) notes that many philosophers and scientists have endorsed Popper's view of the irrationality of scientific theorizing, including Albert Einstein, Carl Hempel, William Whewell and Hans Reichenbach.

does matter, crucially, is that a theory be testable. And this argument led Popper to his famous 'criterion of demarcation': *falsifiability* is the quality that separates science from non-science. If no experiment can be imagined that will disprove a theory, then all observations are consistent with it: it might as well be true as false and there is no basis for calling it 'scientific.'[8] And equally, *any* hypothesis that makes testable predictions (and which also satisfies certain other basic conditions, such as consistency) can legitimately claim to be scientific, irrespective of (for instance) how wide or narrow its domain of applicability.

Popper was quite aware that falsifying a theory is not always a straightforward proposition. An experiment rarely tests one hypothesis in isolation. The prediction that a quantity will have a certain measured value almost always involves a set of assumptions about the measuring apparatus and the experimental design, and if the measured value differs from the prediction, one can never be completely certain where the fault lies. In addition, the interpretation of an experiment often requires assumptions about the validity of various other scientific hypotheses in addition to the hypothesis being tested; it may take a series of cleverly designed experiments to ferret out which of the hypotheses has been falsified by a conflicting measurement.[9] But Popper insisted that – in spite of practical problems like these – it is the responsibility of the scientist to adopt a methodology that maintains falsifiability: "*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, n. 3).

§§

Philosophers are divided over whether Popper's demarcation criterion – which requires that scientific theories be testable, or refutable, or falsifiable – is really the best way to distinguish science from non-science (Laudan, 1983; Grünbaum, 1989; Hull, 2010). But even philosophers who object to falsifiability as a criterion of demarcation are likely to acknowledge the usefulness of a falsificationist *approach* to the testing of scientific hypotheses. The essential point (which Popper often made) is that scientists who are looking for evidence to support a theory can always

[8] Of course one can ask whether this is anything more than a *definition* of science. David Miller (2014b) notes that Popper's goal was not to certify certain hypotheses as 'scientific' and others as 'unscientific.' Rather, it was to determine whether an empirical investigation is worth undertaking. Miller quotes from Popper (1983, p. 174): "my 'problem of demarcation' . . . was not a problem of classifying or distinguishing some subject matters called 'science' and 'metaphysics'. It was, rather, an urgent practical problem: under what conditions is a *critical appeal to experience* possible–one that could bear some fruit?"

[9] The idea that theories are related to experimental results via a web of auxiliary hypotheses is probably obvious to most practicing scientists. Philosophers of science, on the other hand, never seem to tire of reiterating this point, often in the context of a critique of Popper's demarcation criterion (e.g. Suppes, 1967; Schaffner, 1969; Hempel, 1973; Grünbaum, 1976). For Popper's view of these critiques see "Replies to my critics: difficulties of the demarcation proposal" in Book 2 of Schilpp (1974). Anthony O'Hear (1980, chapter VI) presents a balanced discussion and concludes sensibly: "A genuinely scientific method of investigation, then, is one which proposes testable theories and which takes the tests seriously."

image
not
available

disproving theories. But almost everyone, whether inductivist or not, wants to believe that it is possible for experiments to *support* theories, by showing that a theoretical prediction was correct. And in fact it is easy to find examples from the history of science where the experimental or observational confirmation of a prediction led scientists to strongly endorse a new theory; and in at least some of those cases, scientists made (we would now say, with the benefit of hindsight) the 'correct' inference: they endorsed the 'right' theory on the basis of its experimental success.

But if induction is a fallacy, then it is very hard, from a strictly logical point of view, to connect a theory's predictive success to its correctness. Even incorrect theories can make correct predictions, and there will always be an infinite number of theories (most of them yet undreamed of) that can correctly explain any finite set of observations. Only one theory, at most, from that infinite set can be correct, and so there is simply no basis, logically speaking, for claiming that one's pet theory is *the* theory that is supported by the data.

In fact the situation is far worse even than this. For not only can many theories explain the same experimental results. One can also show that *any* observation of *anything* that does not contradict a theory is equally confirming of it, regardless of whether the observation targets a prediction of the theory.

This surprising result is usually[10] attributed to the logician and philosopher Carl Hempel and it is sometimes called 'Hempel's paradox' or the 'paradox of confirmation'[11] – although in fact there is no paradox, in the sense of logical inconsistency; the result is simply extremely counter-intuitive. The proof is simple and goes as follows:

Consider a universal hypothesis such as 'All ravens are black.' This can be written symbolically as the conditional statement

H: If A then B,

where A = raven and B = black. By *modus tollens*, hypothesis H is precisely equivalent to hypothesis H', where

H' : If not B then not A,

i.e. 'All non-black things are non-ravens.'

[10] The origin of the theorem is not clear. A common reference is to Hempel (1937) but the theorem does not appear there; Hempel first presented the theorem in print some years later (Hempel, 1945). In the meantime, the 'paradox' had been pointed out by Janina Hosiasson-Lindenbaum (1940), who would seem to deserve at least partial credit. In her paper (p. 136), Hosiasson-Lindenbaum attributes the result to Hempel without giving a reference ("C. G. Hempel has stated the following paradox"). According to Hempel (1965, p. 20, n. 25), "Dr. Hosiasson's attention had been called to the paradoxes by my article "Le problème de la vérité" [i.e. (Hempel, 1937)] . . . and by discussions with me." Henry Kyburg (1970, p. 166) sums up this confusing set of circumstances as follows: "The oddities that are referred to as the "paradoxes of confirmation" were first noted by Janina Hosiasson-Lindenbaum in 1940; they were christened by Carl Hempel in 1945."

[11] Another name one sometimes sees is 'the paradox of the ravens.' This seems to be an instance of the rule that favorable results in logic are associated with swans, unfavorable results with ravens.

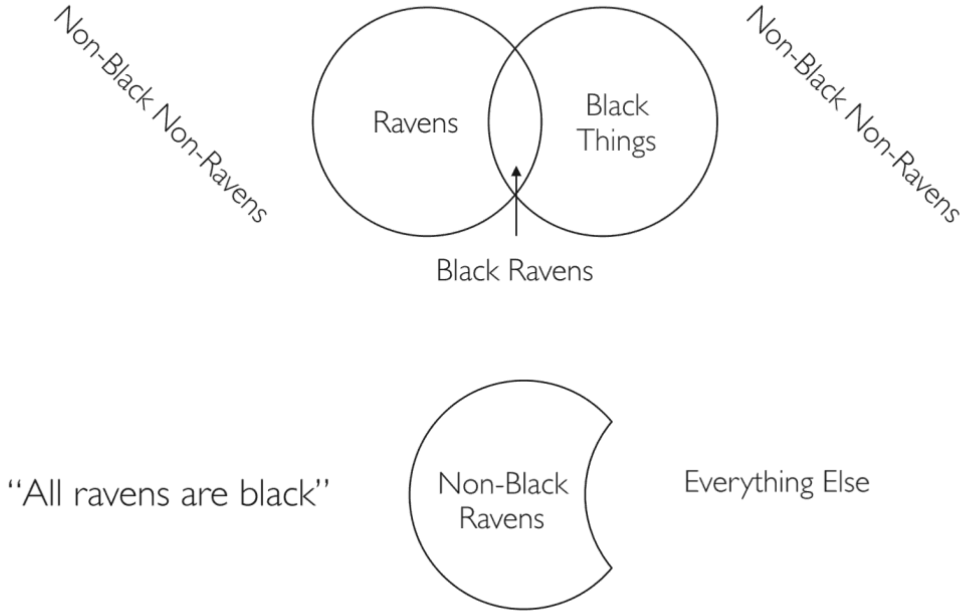

Figure 1.1 A graphical representation of Hempel's theorem (Hosiasson-Lindenbaum, 1940; Hempel, 1945). The top panel divides the universe of things into four groups. Black ravens occupy the intersection of 'Ravens' and 'Black Things.' The bottom panel illustrates the way in which observation of a thing supports the hypothesis 'All ravens are black.' Observation of a non-black raven falsifies the hypothesis; observation of anything else (black ravens, black crows, red robins, yellow mushrooms etc.) is equally confirming of the hypothesis. In other words: from a purely logical standpoint, a theory is supported by any observation that does not refute it. Hempel's theorem implies that – if scientists wish to claim that a theory is 'supported' by confirmed predictions – they need to give a reason why confirmed predictions are more evidentiarily valuable than any other sort of observation. Karl Popper argued that confirmed, *novel* predictions – prediction of facts that were first discovered in the process of attempting to falsify the theory – are the *only* kind that can lend support to a theory, and that the degree of support increases in proportion with the prior improbability of the predicted fact.

Now assume, as an inductivist would, that observation of a black raven supports hypothesis H. (Exactly what is meant here by 'support' is unimportant, as will become clear in a moment.) It must therefore also be true that observation of a non-black non-raven supports hypothesis H'. But H' is precisely equivalent to H. QED: observation of a non-black non-raven – for instance, a red robin, or a yellow mushroom – supports the hypothesis that all ravens are black.

This result is represented graphically in Figure 1.1. It takes only a few more lines of analysis to show that black non-ravens are no different than non-black non-ravens in terms of their ability to provide evidential support of H. In other words: *every hypothesis is supported by anything that does not contradict it; the only sort* *of observation that fails to support a hypothesis is one that disproves it*. Hempel's proof provides striking support for Popper's argument that the only, evidentiarily relevant sorts of observational facts are those that refute a hypothesis.

One is tempted to object to Hempel's theorem on the ground that 'all ravens are black' is a statement only about ravens, hence only the color of *ravens* can be relevant to assessing its truth. But as Figure 1.1 suggests, a universal conditional such as 'all ravens are black' in fact says something about every possible thing in the universe: namely that it is either not a raven, or it is black. We should therefore not be too surprised if every thing in the universe is equally capable of confirming it.

If this result doesn't strike you as devastating to the idea of theory corroboration, you may not have fully grasped its implications. Here is a concrete illustration. A colleague comes to you and says, proudly, "I just spent a year analyzing data from the *Planck* satellite observatory, and the data confirm a prediction that I made about the cosmic microwave background. What a tremendous success for my theory!" You would be perfectly justified, from a *logical* point of view, in responding, "That's very nice; but looking out my window right now, I can see a red robin; and since your theory does not rule out the existence of red robins, my observation is just as genuine a confirmation of the theory as yours. So I don't understand why you expect me to be impressed."

The logic of Hempel's proof is unassailable, but very few people – scientists or philosophers – are willing to accept the conclusion that verified *predictions* are no more corroborating of a theory than any other sort of data. And indeed there is an impressively large body of philosophical literature that is directed toward finding a way round the seeming paradox. One approach (endorsed by Hosiasson-Lindenbaum) starts from the premise that a confirmed prediction only increases the *probability* that the hypothesis on which the prediction was based is correct. Another approach (which Hempel endorsed) begins by supposing that the evidence implies the correctness of only a *weakened version* of the hypothesis.

These end runs around Hempel's theorem are essentially inductivist, or (what is a slightly better term in this context) verificationist. Popper, characteristically, did not even try to evade the implications of Hempel's proof: "Thus an observed white swan will, for the verificationist, support the theory that all swans are white; and if he is consistent (like Hempel), then he will say that an observed black cormorant also supports the theory that all swans are white" (Popper, 1983, p. 235). But Popper argued that there was one special set of circumstances under which observation of a white swan could be seen as supporting the all-swans-are-white hypothesis, while observation of a non-white non-swan would not. Suppose, Popper said, that one's background knowledge – one's earlier theory about swans, together with the existing corpus of data relating to cygnine coloration – had less to say about the color of swans than the new theory. That previous theory might have said that swans can be either white or brown; or that swans come in all colors with equal probability; or perhaps there was simply no basis in the existing theory or data to believe anything definite about the color of swans. Armed now with a bold new theory that claims 'all swans are white,' the scientist will try to falsify it. *Based on his existing knowledge*, he has every reason to expect that he will succeed in falsifying the new theory, since nothing in his prior experience would have led him to expect that the color white is so favored by swans. If the scientist then *fails* to falsify the new theory – if he finds nothing but white swan after white swan – he is justified, Popper argued, in being impressed, since his new theory has *correctly predicted something that previously would have been considered unlikely*.

Based on this argument, Popper (1959, Chapter 10; 1983, Chapter IV) was led to reformulate the question 'Does an observation E support a hypothesis H?' as 'Does E support H in the presence of background knowledge B?' The answer to the latter question, he said, is 'yes,' as long as two conditions are satisfied:

(i) E follows from the conjunction of H and B;
(ii) E is improbable based on B alone.

Popper defined 'background knowledge' as knowledge that existed "previous to the theory which was tested and corroborated."[12] Popper noted that "it is not so much the number of corroborating instances which determines the degree of corroboration as the *severity of the various tests* to which the hypothesis in question can be, and has been, subjected" (Popper, 1959, p. 267).[13] The more improbable a prediction – the greater its novelty based on the background knowledge – the more strongly the new hypothesis is corroborated when the prediction is verified. The *most* severe test would be one that has the potential to contradict, i.e. falsify, the previous theory. Popper called such a test a "crucial experiment" or a "crucial test." But even predictions that are not inconsistent with an old theory can still corroborate a new theory, as long as, and to the extent that, they are novel predictions.

How best to define qualifiers such as 'improbable' or 'novel' is, of course, not obvious. Popper required that an experimental test consist of a "genuine attempt" to refute the theory, but he acknowledged that there was no objective way of judging whether an attempt was genuine.[14] In the years since Popper's proposal, a number of more concrete definitions of novelty have been proposed; these are discussed in detail in Chapter 2. But there is one sort of test that many philosophers agree does *not* satisfy the novelty condition. That is when the theory being tested contains parameters, and the parameters are determined from the same observations

[12] If the reader is reminded here of Bayes's theorem, she is in good company. There is a substantial literature that, building on Popper's insights, attempts to find a Bayesian 'solution' to Hempel's 'paradox'; see e.g. Earman (1996), Howson and Urbach (2006), Crupi et al. (2010). Such attempts have been partially successful at best (e.g. Miller, 2014a). Popper himself was skeptical of Bayesian approaches to theory confirmation: in his words, Bayes's theorem "is not generally applicable to hypotheses which form an infinite set – as do our natural laws" (in Schilpp, 1974, Book 2, p. 1185–1186, n. 68).

[13] Miller (2014a, p. 106): "Sitting around complacently with a well-meant resolve to accept any refutations that happen to arise is a caricature of genuine falsificationism."

[14] Popper (1983, p. 236): "sincerity is not the kind of thing that lends itself to logical analysis."

that constitute the test.[15] An example is the standard, or 'concordance,' model of cosmology which contains roughly a half-dozen parameters that are adjusted in order to give the best correspondence between the theoretical predictions and a particular set (also numbering roughly a half-dozen) of observational results. There is, of course, nothing illegitimate in determining a theory's parameters from data, but it is problematic (a critical rationalist would likely say) to claim that such a procedure constitutes *corroboration* of the theory. Data that are used to set the parameters of a theory do not *corroborate* a theory; they only *complete* the theory; and of course those data can be considered part of the 'background knowledge' that was used in theory construction. Corroboration would consist (for instance) of using the theory – with its parameters now fixed – to generate *new* predictions and demonstrating that they are correct.

§§

Scientific theories contain universal laws, but they can also contain more prosaic elements; for instance, statements that something *exists*. In some cases, these statements refer to entities the existence of which is not (or at least, is no longer) debated; for instance, the list of fundamental particles (and their measured properties) that are part of the standard model of particle physics. But existential statements can also refer to entities that are conjectural. In such cases, existential statements take the form of hypotheses. Examples are statements in the standard model of cosmology about dark matter and dark energy.

There is an interesting logical symmetry between universal statements (such as scientific laws) and existential statements. As discussed above, universal statements are falsifiable (at least, if they are well formulated) but they are not verifiable. Existential statements, on the other hand, *are* verifiable: one can verify the statement 'positrons exist' by finding a positron. And existential statements are *not* falsifiable: one could look forever for a positron without finding one, but never be certain that a positron didn't exist in some place that hadn't yet been searched.

The symmetry goes deeper. The negation of a universal statement is an existential statement, and the negation of an existential statement is a universal statement. The negation of 'All swans are white' (a universal statement) is 'It is not true that all swans are white' or equivalently 'There exists at least one non-white swan.' That is an existential statement and it is capable of being verified. Or consider the existential statement 'There is an element with atomic weight 52.' The negation of this statement is 'There is no element with atomic weight 52' – a universal hypothesis that can be falsified, by finding such an element, but never verified.

[15] E.g. Worrall (1978b, p. 48): "one can't use the same fact twice: once in the construction of a theory and then again in its support"; Zahar (1973, p. 102–3): "Very often the parameters can be adjusted so as to yield a theory T^* which 'explains' the given facts . . . In such a case we should certainly say that the facts provide little or no evidential support for the theory, since *the theory was specifically designed to deal with the facts*."

spectrum of temperature fluctuations in the cosmic microwave background. As in the case of DM-1, the correctness of DM-2 is typically assumed by these scientists, regardless of what else they might postulate or believe about dark matter. Logically, the assumption that DM-2 is correct implies nothing about the correctness of DM-1, or *vice versa*, although standard-model cosmologists routinely ‘fuse’ the two hypotheses into one.

A third assumption is often (though not always) made by standard-model cosmologists:

DM-3: The dark matter component that appears in DM-1 (or in DM-2, or in both) consists of elementary particles.

Now consider how these hypotheses might be refuted or corroborated. One thing to notice is that testable statements derivable from DM-1 constitute a completely distinct set from testable statements derivable from DM-2. DM-1 can only be used to make predictions about actual, observed galaxies,[20] while DM-2 says nothing at all about individual galaxies. The two hypotheses are empirically, as well as logically, independent – which, of course, was one of the reasons for stating them in this way. It follows that observational data that corroborate, or falsify, DM-1 can contain nothing of any evidentiary relevance to DM-2, and *vice versa*. It is perfectly possible for a particular set of data to corroborate DM-2 (for instance), while another set of data falsifies, or fails to corroborate, DM-1. And given that the existence of dark matter (not to mention its nature) is speculative, such a situation can not be ruled out *a priori*.

Not only is this *possible*: it is a pretty fair description of the current state of affairs. Hypothesis DM-2 has been corroborated (though not, of course, confirmed) by observations of large-scale structure and of the cosmic microwave background.[21] But rotation-curve data can not corroborate hypothesis DM-1, since, as Mordehai Milgrom (1989a, p. 216) has noted, “The DMH [dark matter hypothesis, i.e. DM-1] simply states that dark matter is present in whatever quantities and space distribution is needed to explain away whichever mass discrepancy arises.” Another way to express this is via the argument mentioned above: data (that is, a galaxy rotation curve) that are used to determine the parameters of a hypothesis (in this case, the parameters that specify the distribution of the galaxy’s dark matter) do not provide evidential support for the hypothesis. Stated yet another way: one can not invoke a hypothetical entity to explain anomalous data, then turn around and claim that those same data constitute evidence for the existence of the entity.[22]

A potentially testable prediction *does* follow from the conjunction of DM-1 and DM-3. If the dark matter consists of elementary particles, then the mass density

[20] That is: about the *dark matter* in actual, observed galaxies
[21] See Chapter 6.
[22] I belabor the point, because textbooks and review articles on cosmology do routinely argue in just this way.

of those particles near the Sun is known (modulo degeneracies in reproducing the known Milky Way rotation curve using different assumed spatial distributions of dark matter) as is their approximate velocity distribution. Laboratory experiments on the Earth could therefore corroborate the joint hypothesis by detecting the particles. But any such experiments suffer from a lack of knowledge about the cross section for interaction of the putative particles with the normal matter in the detectors, rendering the prediction essentially unfalsifiable. And of course, all attempts to detect the putative particles have so far failed.[23]

In this respect, the dark matter hypothesis is in a state similar to that of the atomistic hypothesis at the end of the nineteenth century. Popper notes that the hypothesis that atoms exist was, for a long time, too vague to be refuted. "Failure to detect the corpuscles, or any evidence of them, could always be explained by pointing out that they were too small to be detected. Only with a theory that led to an estimate of the size of the molecules was this line of escape more or less blocked, so that refutation became in principle possible" (Popper, 1983, p. 191). Popper's statement is perfectly applicable to the (particle) dark matter hypothesis if one replaces 'size of the molecules' by 'cross section of interaction of the dark matter particles with normal matter.'

As we will see in Chapters 6 and 7, Milgromian theory postulates rather *different* explanations for the anomalous observations on galactic and cosmological scales, observations which in the standard model are explained by a single postulated entity, 'dark matter.'

§§

Popper's scheme of conjectures and refutations implies that theories will evolve along a sequence, with each refuted version replaced in turn by another version, hopefully having more empirical content – that is, capable of generating more testable predictions – than the version it replaced. But to the two requirements of falsifiability and greater content, Popper added a third desideratum: a new theory "should pass some new, and severe, tests" (Popper, 1963, p. 242). That is: some of its novel predictions should be experimentally verified.

The idea here is again very simple. Scientists should not be in the business, Popper said, of "merely producing theories so that they can be superseded" (Popper, 1963, p. 245). Consider, said Popper, a sequence of theories, each of which explains the observations or experiments that brought down its predecessor, and each of which makes some new predictions. Now suppose that the novel predictions are

[23] Liu et al. (2017, p. 215): "there has been no solid evidence of a real event yet . . . one cannot ignore the importance of those null searches which have been setting tighter constraints to many theoretical models and which may eventually direct us on a completely different path towards understanding this mysterious component of our Universe." Note that the failure to detect dark matter particles has not, apparently, shaken Liu et al.'s conviction that dark matter exists.

always immediately refuted. There would be no reason to believe, said Popper, that such a sequence of theories represents an approach to the truth. A true theory, after all, makes nothing *but* successful predictions, and it is reasonable to require some reassurance that we are moving in the direction of that theory. "If we are content to look at our theories as mere stepping stones", said Popper, "then most of them will not even be good stepping stones":

> Thus we ought not to aim at theories which are mere instruments for the exploration of facts, but we ought to try to find genuine explanatory theories: we should make genuine guesses about the structure of the world . . . if we should cease to progress in the sense of our third requirement – if we should only succeed in refuting our theories but not in obtaining some verifications of predictions of a new kind – we might well decide that our scientific problems have become too difficult for us because the structure (if any) of the world is beyond our powers of comprehension (Popper, 1963, p. 245).

As an example of a corroborated novel prediction, Popper cited the bending of starlight by the gravitational force from the Sun, which led many scientists to accept the correctness of Einstein's theory of general relativity.

In fact, Popper's conclusion about the privileged role of confirmed novel predictions is one that scientists have independently arrived at, again and again, though not necessarily via the same chain of reasoning as Popper's. Here are three examples, all pre-dating Popper; many more could be given. The astronomer John Herschel wrote in 1842:[24]

> The surest and best characteristic of a well-founded and extensive induction . . . is when verification of it springs up, as it were, spontaneously, into notice, from quarters where they might be least expected . . . Evidence of this kind is irresistible and compels assent with a weight which scarcely any other possesses.

William Whewell wrote in 1847 that if a theory

> of itself and without adjustment for the purpose, gives us the rule and reason of a class of facts not contemplated in its construction, we have a criterion of its reality, which has never yet been produced in favour of a falsehood (Whewell, 1847, Vol. 2, p. 67–68).

And the physicist Norbert Robert Campbell wrote in 1921:

> A true theory will not only explain adequately the laws that it was introduced to explain; it will also predict and explain in advance laws which were unknown before. All the chief theories in science (or at least in physics) have satisfied this test (Campbell, 1921, p. 87).

As we will see, there *is* a research tradition in cosmology – the one originated by Mordehai Milgrom in 1983 – that has repeatedly been successful in just this privileged way, predicting again and again (to use Whewell's words) "a class of

[24] Herschel (1842), Sect. 180. Quoted in *Theories of Scientific Method: The Renaissance Through the Nineteenth Century*, E. Madden (ed.) University of Washington Press, 1960, p. 177. Note Herschel's assumption that scientific theories are arrived at via induction.

facts not contemplated in its construction." By contrast, the standard cosmological model has rarely succeeded in making successful novel predictions; instead it has repeatedly been forced to 'play catch-up,' finding post hoc explanations for unexpected discoveries rather than predicting them in advance. (In many cases, those discoveries *were* predicted in advance by Milgromian researchers; they were 'unexpected' only from the standpoint of standard-model researchers.) Given scientists' supposed predilection for theories that (in Campbell's words) "predict and explain in advance laws which were unknown before," one might reasonably ask why the standard cosmological model is currently so dominant, while Milgrom's theory is so marginalized. I will return to this question in Chapter 9.

§§

It is clear from his writings that Popper had a sophisticated appreciation of the different ways that theories can evolve: sometimes via the addition of auxiliary hypotheses, as in the case of Ptolemy's equants or 'dark matter,' and sometimes via radical or revolutionary changes, as when Newton's theory of gravity and motion was replaced, wholesale, by Einstein's. But in his arguments about the criteria for scientific progress (or as he often called it, the "growth of knowledge"), Popper did not distinguish strongly between these different modes of theory change, and it seems likely that he intended his three criteria of progress to apply to all of them.

The approach to theory appraisal that will be followed in this book is due to Popper's colleague Imre Lakatos. Lakatos accepted many of Popper's arguments about evidential support and about the necessary conditions for scientific progress. For instance, Lakatos emphasized, as did Popper, that the success of a theory is measured not by the total number of successful predictions, but only by successful *new* predictions: "the only relevant evidence is the evidence anticipated by a theory" (Lakatos, 1970, p. 38). But Lakatos argued, based on the historical record (and in agreement with Thomas Kuhn), that scientists will go to extreme lengths to avoid modifying the fundamental assumptions that underlie their theories. Scientists, he said, almost always respond to anomalies – that is, falsifications – by adding auxiliary hypotheses, and leaving the "hard core" of the theory unchanged:

> Scientists have thick skins. They do not abandon a theory merely because facts contradict it. They normally either invent some rescue hypothesis to explain what they then call a mere anomaly or, if they cannot explain the anomaly, they ignore it, and direct their attention to other problems (Lakatos, 1973, p. 4).

The result is a series of *connected* theories, linked together by an essentially fixed set of fundamental assumptions (the hard core), which Lakatos called a "research program." Lakatos argued that the proper unit of appraisal for scientific progress is the research program, rather than the isolated theory. By failing to distinguish clearly between theories and research programs, Lakatos argued, Popper had been unable to explain the *continuity* of science: the fact that theories often retain a recognizable character over time in spite of changes.

Lakatos died in 1974 at the age of 51. Had his short life continued just a few years longer, Lakatos might have applied his method of appraisal to the standard cosmological model – or as he might have called it: the standard cosmological research program.[25] No one, it seems, has yet taken the time to do that, nor will the attempt be made in this book. But it is clear that the development of the standard cosmological model since about 1970 adheres quite nicely to Lakatos's basic template. There is a fixed core, which includes Einstein's theory of gravity and the standard model of particle physics.[26] When the predictions of the theory have been refuted,[27] the response has almost always been to add an auxiliary hypothesis to the theory, a hypothesis explicitly designed to maintain the integrity of the hard core in the face of the anomalous data. The postulates relating to 'dark matter' and 'dark energy' came about in this manner. Of course, nothing in the foregoing sentences should be read as implying that the evolution of the standard cosmological model has been *progressive* in the sense understood by Popper or Lakatos.

Lakatos emphasized that two or more competing research programs are often pursued at the same time in a given field, typically by different sets of scientists, before one research program finally succeeds in supplanting the other(s). For instance, with regard to theories of matter prior to the early twentieth century, there were continuity theories, atomistic theories, and theories that tried to combine the two. Much the same is true today in the field of cosmology. There is a research program, begun by the physicist Mordehai Milgrom in 1983, that has evolved side-by-side with the standard cosmological research program. Milgrom's research program is the topic of most of the remainder of this book.

[25] Nor did Popper, who outlived Lakatos, have much to say about theories of cosmology; in Helge Kragh's (2012, p. 332) words, "one looks in vain in [Popper's] main works for discussions of the science of the universe."

[26] Throughout this book, 'standard-model cosmologist' refers to an adherent of the standard, or concordance, or ΛCDM cosmological model. Such cosmologists typically assume the correctness of the standard model of particle physics as well.

[27] Pavel Kroupa (2012) gives a timeline showing the major failures of the standard cosmological model.

And elsewhere Kuhn declared that normal science "often suppresses fundamental novelties because they are necessarily subversive of its basic commitments" (Kuhn, 1962, p. 5).

Kuhn reached these dismal conclusions (he said) based on an examination of the historical record, but his argument was essentially a logical one. Science (said Kuhn), like all rational discourse, requires a common language and set of assumptions. Kuhn called this shared framework a "paradigm,"[2] and he argued (in much the same way as cultural relativists and postmodernists[3]) that meaningful communication is only possible between scientists who share, uncritically, the same paradigm: "it is precisely the abandonment of critical discourse that marks the transition to a science" (Kuhn, 1970, p. 6). Kuhn acknowledged that scientists frequently engage in 'testing,' but he said that these were almost never tests of a *theory*; rather they were tests of the scientist's skill at (re)interpreting the theory so as to reconcile it with the data. He called this activity "puzzle solving," and said that a failure to solve a puzzle reflected on the scientist, not the theory: "in the final analysis it is the individual scientist rather than current theory which is tested" (Kuhn, 1970, p. 5).

Most scientists will admit that Kuhn's description of science-as-practiced, although unflattering, rings true in many respects (as did Popper, although he suggested that Kuhn's claims for the dominance of 'normal science' applied much more to science as practiced after the First World War than before; in Schilpp, 1974, Book 2, p. 1146). And there is no question that Kuhn's observations forced philosophers of science to give greater attention to the day-to-day activities of real scientists. But Kuhn never stated clearly how his "new image of science" could differentiate the practice of science from the other human activities that also take place within a shared paradigm. As Popper remarked,

> Kuhn and I agree that astrology is not a science, and Kuhn explains why from his point of view it is not a science. This explanation seems to me entirely unconvincing: *from his point of view* astrology should be accepted as a science. For it has all the properties which Kuhn uses to characterize science: there is a community of practitioners who share a routine, and who are engaged in puzzle solving . . . we may find, in a couple of years' time, the great foundations supporting astrological research. From Kuhn's sociological point of view, astrology would then be socially recognised as a science (in Schilpp, 1974, Book 2, p. 1146).

Or as Paul Feyerabend (1970, p. 200) put it: "Every statement which Kuhn makes about normal science remains true when we replace 'normal science' by 'organized crime'."

[2] At least, that is one way to interpret Kuhn's term. Margaret Masterman (1970, p. 59) finds that "On my counting, [Kuhn] uses 'paradigm' in not less than twenty-one different senses."

[3] The editors of *A Postmodern Reader* include, between selections of Jean Baudrillard and Cornel West, Kuhn's "The Resolution of Revolutions" and preface Kuhn's article with the approving words: "Kuhn argues that what is at stake at such moments of change is learning to "see science and the world differently." . . . Seeing within a different paradigm means being in a place where we can make conceivable that which is not already presentable within our prevailing paradigm's "rules of the game"" (Natoli and Hutcheon, 1993, p. 307).

Kuhn's 'new image of science' is indistinct in other ways. It is very well to point out that scientists can be indifferent to experimental anomalies. But what constitutes *support* for a theory, and on what basis do scientists make this judgment? Recall that Popper gave a carefully reasoned answer to the first question: theories are supported by confirmed, novel predictions. One can disagree with Popper's conclusion, or argue based on the historical record that scientists use a different criterion; but in light of Hempel's theorem (anything that does not contradict a hypothesis is, logically, equally confirming of it), scientists (and philosophers of science) need *some* criterion for separating the wheat from the chaff, evidentiarily speaking. Kuhn has little to say on this essential question.[4] When listing the features of a theory that are relevant in corroborating it, Kuhn mentions "accuracy of prediction, particularly of quantitative prediction ... and the number of different problems solved"(Kuhn, 1962, postscript, p. 206). But Ptolemy's epicycles and equants were capable of making predictions with arbitrary accuracy. And with regard to solved problems, Kuhn does not distinguish between problems which a theory was specifically designed to solve, and those which only appear after a theory's construction; nor did it appear to matter, to him, whether a solution arises organically from the theory, or consists of (as Popper would say) a conventionalist stratagem.

In the words of Paul Feyerabend (1970, p. 198):

> Whenever I read Kuhn, I am troubled by the following question: are we here presented with *methodological prescriptions* which tell the scientist how to proceed; or are we given a *description*, void of any evaluative element, of those activities which are generally called 'scientific'? Kuhn's writings, it seems to me, do not lead to a straightforward answer.

Or as John Kadvany (2001, p. 151) put it: "Kuhn never clearly identified just what should be *done* with the new image of science."

§§

It fell to Imre Lakatos to develop a description of science that incorporated Popper's logical and epistemic insights; maintained a demarcation between science and pseudoscience; and accommodated (or neutralized) the apparent threats to scientific rationality posed by Kuhn's interpretation of the historical record.

Lakatos was born Imre Lipsitz, a name which he changed during the Nazi occupation of Hungary to the less Jewish-sounding Imre Molnar, and again after the start of the Russian occupation to the more working-class Imre Lakatos ("Locksmith" in Hungarian). After the war, Lakatos was politically active and was made a secretary in the Ministry of Education. On returning to Hungary from a visit to Moscow in 1949, he was arrested (exactly why is not clear) and spent more than three years

[4] Larry Laudan (1984, p.73) makes a similar point: "Kuhn has failed over the past twenty years [i.e. since 1964] to elaborate any coherent account of consensus formation, that is, of the manner in which scientists could ever agree to support one world view rather than another."

in prison. After his release in 1954, Lakatos began studying mathematics with Alfréd Renyi, and translated György Pólya's *How to Solve It* into Hungarian. He also began to turn away from Marxism, and when the Hungarian uprising of 1956 was put down by Soviet troops, Lakatos fled Hungary for Vienna. From 1960 until his early death in 1974 he taught alongside Karl Popper at the London School of Economics.

Before moving to London, Lakatos completed a doctoral thesis in King's College, Cambridge, on the philosophy of mathematics, entitled *Essays in the Logic of Mathematical Discovery*. Four papers based on this work were published during 1963–1964; but Lakatos considered the work unfinished, and it was not until after his death that John Worrall and Elie Zahar published a volume, *Proofs and Refutations* (1976), that included selections from the thesis, the four published papers, and commentary speculating how Lakatos might have further developed his arguments had he lived.

In *Proofs*, Lakatos argued that mathematical theorems are not derived simply by deductive reasoning starting from some set of fixed postulates, as the textbooks usually imply. When proving a theorem, he said, mathematicians will start from a general idea, or hunch, as to what it is they are trying to prove, then 'stretch' the definitions of fundamental terms, or modify the statement of the theorem, as needed to resolve difficulties and allow the proof to go forward. Lakatos borrowed the term 'heuristic' from Pólya to describe this process of conceptual growth through conjectures and refutations. As an example, Lakatos considered Euler's formula for simple polyhedra, $V - E + F = 2$ ($V =$ number of vertices, $E =$ number of edges, $F =$ number of faces). Lakatos pointed out that some geometrical objects fail to satisfy Euler's theorem; for instance, a solid cube with a cubic space inside it has $V - E + F = 4$. Given a counter-example like this, the mathematician can abandon the theorem; restate the theorem in such a way as to account for the confounding object; or change the definition of some term or terms, e.g. 'polyhedron' or 'vertex.' Lakatos showed that by considering objects for which $V - E + F$ does *not* equal two, one is led to a rule for the 'Euler characteristic' $V - E + F$ that works for a much larger class of solids than the simple polyhedra.

Pólya had argued in *How to Solve It* that there were rational methods, or "heuristics," available to the mathematician for generating theorem-candidates. For instance, theorems in plane geometry often have counterparts in higher dimensions. Lakatos used 'heuristic' to mean, roughly, the process of critical argument that leads to a change in mathematical concepts or language, or a shift to a new conceptual framework. He did not claim that there was a *unique* mode of mathematical discovery; instead he sought to extract the heuristic used, case by case, by examination of particular historical episodes. In this respect, Lakatos was departing from the Popperian view that discovery and justification are two quite distinct things, and that only the latter is subject to rational analysis. Lakatos also differed from Popper by adopting a more nuanced view of the role of falsification; as in the Euler proof described above, he argued that counter-examples can sometimes be useful, by indicating the way forward.

After joining Popper in London, Lakatos turned his attention to the philosophy of science. Like Popper, Lakatos was uncomfortable with Kuhn's view of science as an enterprise detached from criticism. In his words:

> The clash between Popper and Kuhn is not about a mere technical point in epistemology. It concerns our central intellectual values, and has implications not only for theoretical physics but also for the underdeveloped social sciences and even for moral and political philosophy (Lakatos, 1970, p. 9).

In two long essays – "Falsification and the methodology of scientific research programmes" (1970) and "History of science and its rational reconstruction" (1971)[5] – Lakatos showed how to solve Popper's problem of demarcation in a way that respects the historical record. In outline, his procedure was as follows:

Scientists and philosophers have a pretty good idea (Lakatos said) about which episodes of intellectual history correspond to 'good' science (Newton, Einstein) and which do not (Marx, Freud). Suppose that the goal is to test a general hypothesis about what constitutes the methodology associated with good, or rational, science. For instance, one might postulate that scientists produce their theories by generalizing from experimental data; that is, via induction. In the same way that scientific *theories* can be tested and potentially falsified, so can hypotheses about the methodology of science. The trick is to inspect the historical record and look for episodes during which a scientific theory evolved in the postulated manner (the 'rational' episodes) and the episodes during which it did not (the 'irrational' episodes). Lakatos called such an exercise "rational reconstruction," and, like all historiography, a certain amount of interpretation is involved. If it turns out that the rational episodes dominate the non-rational ones, then the postulated methodology is reasonable: it has survived the attempt to falsify it, and one can proceed to apply the test to another period of successful science. Whereas if the postulated methodology turns out to conflict with the historical record – if most episodes of what we consider 'good' science cannot be described as 'inductivist' without doing violence to the historical record – then induction is probably not a good description of how science actually works.

Lakatos tested three hypotheses about what constitutes rational scientific practice: inductivism, conventionalism (in the original sense of that term due to Poincaré), and Popper's falsificationism. (He did not test any hypothesis of Kuhn because Kuhn never provided one.) As historical episodes, Lakatos chose the Copernican and Newtonian revolutions – clearly episodes of 'successful'

[5] Reprinted as Chapters 1 and 2 of *The Methodology of Scientific Research Programmes*, Philosophical Papers Volume 1 (Cambridge University Press, 1978), edited by John Worrall and Gregory Currie. In citing Lakatos from these two essays, I will adopt the pagination of that edited volume. I am grateful to Cambridge University Press for their permission to quote liberally from this volume.

science. Lakatos argued that all three hypotheses failed the historiographical test. For instance, under falsificationism, one would expect scientists to immediately abandon a theory once it has been falsified; continuing to work on a falsified theory would be irrational. But Lakatos found (as had Kuhn) that theories are engulfed in an “ocean of anomalies” from the start, and that scientists nevertheless continue to work on them; in fact a scientist who abandoned falsified theories would not be able to do science at all. Thus, falsificationism was ‘falsified’ by the historical record.

Having shown that none of the extant hypotheses about what constitutes rational science could be made to fit the historical record, Lakatos's next step was to formulate a new hypothesis and test it. His proposal – the “methodology of scientific research programmes” – was still (like everything in Popper) essentially critical-rationalist: the rational scientist is assumed to postulate theories, generate predictions, and test them via experiment or observation. And following Popper, Lakatos defined evidential support purely in terms of novel predictions: “the only relevant evidence is the evidence anticipated by a theory” (Lakatos, 1970, p. 38). But Lakatos broke with Popper by arguing (as he had in *Proofs*) that failures of prediction are not fatal: what matters more is how scientists *develop* a theory in response to a falsifying instance. (Popper, it seems, never forgave Lakatos for this apostasy.) The sequence of theories that results from this developmental process Lakatos called a “research program,” and he argued that the proper unit of appraisal is the entire program, not any single theory taken from it. Lakatos also carried over from his earlier work in philosophy of mathematics the idea of a heuristic that guides the scientist's work and suggests (among other things) how theories should be modified in response to anomalies.

In his two long essays, Lakatos tested his proposed methodology against some well-known episodes from the history of science, and argued that he was able to correctly distinguish ‘good’ science from ‘bad’ science (or, as he termed it, “progressive” versus “degenerating” science). A number of more thorough historiographical appraisals of this sort were carried out by his students and published after his death, many of them in the volume *Method and Appraisal in the Physical Sciences* (C. Howson, ed., 1976). More appraisals have been published since then: not only in the physical sciences, but in economics, demographics, biology etc. This large body of work can be interpreted as corroboration (for the most part) of Lakatos's proposed methodology. But it is important to recognize the dual role played by the historical record in the original formulation of the *Methodology*. Not only did Lakatos appeal to history to test his proposed methodology. It is also clear that Lakatos was guided in *formulating* that methodology by his knowledge of the history of science, and of the ways in which the extant demarcation proposals failed when confronted with the historical record. Quite a bit of ‘conjecture and refutation’

Newton's theory was eventually replaced by Einstein's, but not, Lakatos argues, because Newton's theory was falsified; rather because Einstein's theory

> 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) (Lakatos, 1970, p. 39).

Lakatos saw reflected in these examples two ideas already in Popper: (*i*) a theory can be 'saved' by invoking auxiliary hypotheses, as long as they are not 'conventionalist,' i.e. as long as they constitute an increase in empirical content (degree of falsifiability); and (*ii*) one theory may be preferred over another even if the first has not been conclusively falsified. The first idea is illustrated by the persistence of Newton's theory in spite of anomalies; the second by the adoption of Einstein's theory, for reasons other than the falsification of Newton's. What makes for scientific progress, Lakatos argued, is the *manner in which a theory is developed* over time.[8] The proper entity to consider when appraising scientific progress is not a theory in isolation, but what Lakatos called a "research program": a *series* of theories in which

> 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 (Lakatos, 1970, p. 33).

Theories change, but a research program maintains its identity over time. Popper, Lakatos felt, had sometimes acknowledged a distinction between theories and series of related theories, but his focus on the former had kept him from producing a convincing explanation for the continuity of science over time.

But what is it that remains *fixed* in a research program, in spite of the changes? Lakatos (1974, p. 146) reminds us of Popper's insistence that a scientist be prepared to state under what conditions he would abandon his most basic assumptions; to the extent that a Freudian or a Marxist is unwilling to do this, said Popper, their theories are unscientific. But, Lakatos argued, much the same could be said of Newtonian scientists, who (at least until shortly before the replacement of Newton's theory by Einstein's) were equally unwilling to abandon the central tenets of their theory: Newton's laws of gravity and motion. Lakatos (1970, p. 48) described Newtonian gravitational theory as "possibly the most successful research programme ever"; clearly, Lakatos said, the maintenance by Newtonian scientists of a fixed, irrefutable, "hard core" to their research program was not sufficient to render it unscientific. Here Lakatos made a substantial break with Popper, for whom

[8] Note that Lakatos would consider Newton's and Einstein's theories as belonging to different research programs.

conventions determined the acceptance of singular statements only, not of universal ones (Popper, 1959, section 30). Lakatos argued that the decision (i.e. convention) to define certain *universal* statements as unfalsifiable is a generic feature of scientific research programs.[9]

Lakatos noted that Niels Bohr, already in a paper from 1913, explicitly set out the postulates constituting the hard core of his research program: in Bohr's own words (Bohr, 1913, p. 874–875),

1. That energy radiation [within the atom] is not emitted (or absorbed) in the continuous way assumed in the ordinary electrodynamics, but only during the passing of the systems between different "stationary" states.
2. That the dynamical equilibrium of the systems in the stationary states is governed by the ordinary laws of mechanics, while these laws do not hold for the passing of the systems between the different states.
3. That the radiation emitted during the transition of a system between two stationary states is homogeneous, and that the relation between the frequency ν and the total amount of energy emitted E is given by $E = h\nu$, where h is Planck's constant.
4. That the different stationary states of a simple system consisting of an electron rotating round a positive nucleus are determined by the condition that the ratio between the total energy, emitted during the formation of the configuration, and the frequency of revolution of the electron is an entire multiple of $\frac{h}{2}$. . . .
5. That the "permanent" state of any atomic system, i.e. the state in which the energy emitted is maximum, is determined by the condition that the angular momentum of every electron round the centre of its orbit is equal to $\frac{h}{2\pi}$.

In other research programs, said Lakatos (though without giving examples), the hard core "develops slowly, by a long, preliminary process of trial and error" (Lakatos, 1970, p. 48, n. 4).[10] But substantial change in the hard core would mean abandoning the research program altogether.

Lakatos made much of the fact that Bohr's hard core postulates were *inconsistent*; in effect, Bohr "grafted" his research program onto Maxwell's theory of electromagnetism. Lakatos saw inconsistencies also in the hard cores of Prout's theory of atomic weights and of Copernican astronomy, and even (in its early days) in Newtonian theory, before the acceptance of 'action-at-a-distance.' Of course, to the extent that a theory corresponds to reality, it must be consistent. Lakatos argued, based on examples from the historical record, that "As the young grafted

[9] Although Lakatos appears never to use the term, *fideism* is the epistemic principle that he is here ascribing to scientists. Richard H. Popkin notes that fideism can be usefully defined in a number of ways, but that "there is . . . a common core, namely that knowledge . . . is unattainable without accepting something on faith" (Popkin, 2003, p. xxii). Or in Pierre Jurieu's more succinct seventeenth century formulation, "Je le crois, dis-je de cette manière: parce que je le veux croire" (Jurieu, 1687, p. 248–249). That scientists behave in the way that Lakatos describes can hardly be gainsaid, but one is left wondering on what basis a scientific community decides to assign certain assumptions to their hard core and not others. This question is revisited in Chapter 7.

[10] An instance from the standard cosmological model (or rather, its associated research program) would be the dark matter hypothesis, which was not part of the model prior to about 1980, but which since then has acquired the status of an unchallengeable assumption (Merritt, 2017).

programme strengthens, the peaceful co-existence comes to an end, the symbiosis becomes competitive and the champions of the new programme try to replace the old programme altogether" (Lakatos, 1970, p. 56–57).

§

To the extent that the hard core of a research program is taken as invariant, experimental or observational anomalies (i.e. falsifications) must be dealt with via auxiliary hypotheses. "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" (Lakatos, 1970, p. 48). Here again, the idea has a basis in Popper, who noted that, logically, any falsification could be dealt with by an addition to the theory. As discussed in Chapter 1, Popper required that, to be acceptable, changes must satisfy two extra conditions (in addition to preserving falsifiability): they must be content-increasing, and at least some of the theory's new predictions should eventually be confirmed. Lakatos adopted essentially the same requirements in defining what he called "progressive problemshifts":[11]

Let us take a series of theories, $T_1, T_2, T_3, \ldots$ [in a given research program] 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. Finally, let us call a problemshift *progressive* if it is both theoretically and empirically progressive, and *degenerating* if it is not. We '*accept*' problemshifts as 'scientific' only if they are at least theoretically progressive; if they are not, we '*reject*' them as 'pseudoscientific' (Lakatos, 1970, p. 33–34).[12]

That theories can be, and will be, falsified, in the sense understood by Popper, is taken for granted by Lakatos. But "We regard a theory in the series 'falsified' when it is superseded by a theory with higher corroborated content" (Lakatos, 1970, p. 34). Lakatos emphasized – as did Popper – that the success of a theory is measured not by the *total* number of successful predictions, but only by its success

[11] "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" (Lakatos, 1970, p. 34, n. 2).

[12] Note that Popper's definition of empirical content (the class of potential falsifiers; see Chapter 1) differs from that of Lakatos; see e.g. Popper (1963, p. 385) for a critical comparison of the two definitions. Lakatos's usage (which many authors adopt) is retained in what follows.