PDF version of the talk
"The Scientific Method from a Philosophical Perspective"
David Merritt
https://doi.org/10.5281/zenodo.6336021

An invited talk at the ESO on-line conference

The Present and Future of Astronomy

February 14-18, 2022

Cite as: Merritt, David. (2022, February 25). The scientific method from a philosophical perspective. Zenodo. https://doi.org/10.5281/zenodo. 6336021

The Scientific Method from a Philosophical Perspective

First published Fri Feb 25, 2022; revised Mon Mar 8, 2022

A methodology of science must satisfy two requirements: (i) It must be *ampliative*: the theories which it generates must make statements that go far beyond any data or observations that may have motivated those theories in the first place. (ii) It must be *epistemically probative*: it must provide a warrant for believing that the theories so produced are correct, or at least partially correct, even if they can never be fully confirmed. These two requirements pull in opposite directions, and attempts to specify the "scientific method" often focus on one to the exclusion of the other. On a few points there now exists something approaching a consensus. (i) Scientific hypotheses — including, particularly, statements about unobserved or unobservable entities or mechanisms — remain conjectural, no matter how frequently predictions based on those hypotheses are found to coincide with data. (ii) A good (best?) indicator of a theory's verisimilitude is its ability to successfully predict phenomena which it was not specifically designed to predict. I discuss these ideas with particular reference to cosmological theories.

Whatever scientific methodology *is*, it should satisfy two requirements if it is to explain the success of science:

— It should be **ampliative**: it should be able to take us from finite data and knowledge to theories which can make predictive statements even about situations which have never been observed or experienced

—It should be **epistemically probative**: it should give us confidence that our theories or hypotheses are correct, even if we can never hope to confirm all of their predictive content

These two requirements pull in opposite directions!

The only statements that are capable of confirmation are trivial ones (e.g. tautologies, or singular statements like “the cat is on the mat”) which say little or nothing about the world

Whereas any theory that contains universal statements (e.g. Einstein’s field equations; the laws of thermodynamics) entails infinite numbers of consequences (in time, and space) and can never be fully confirmed

Prior to the 20th century, the accepted model of scientific methodology was **induction**. According to this view (which can be traced back at least as far as Francis Bacon in the 16th century), scientists observe the world with an unprejudiced mind, and (somehow) convert this sensory information into complex and informative theories.

An influential group of philosophers (the "Vienna Circle") attempted to formalize this methodology starting in the 1920s. They called their philosophical school **logical positivism**. Their goal (among others) was to understand what differentiates science from other intellectual enterprises, and (hopefully) to import the "scientific" methodology into other disciplines, such as philosophy.

Their starting point was the **verification principle**: roughly speaking, that the only meaningful statements are those that can be verified by direct observation. Their task, as they saw it, was to establish how scientists could construct hypotheses from combinations of such verifiable statements.

Logical positivism failed: by the 1960s, essentially no philosopher endorsed this approach. And the person most responsible for the demise of logical positivism was Karl Popper.

Karl Popper (1902-1994) was a young student in Vienna in the 1920s, and while never a formal member of the Vienna Circle (he was described, by one of its members, as the "official opposition"), he paid close attention to their work.

Already by the end of that decade, Popper had generated his own, fully-formed ideas about what differentiates science from non-science, and about the methodology that scientists use (or could use, or should use) for generating knowledge.

His critiques of logical positivism included the following, strongly reasoned arguments:

— The "verification principle" would label as meaningless all scientific laws, since only a negligible fraction of a theory's predictive content can ever hope to be verified.

— Scientific theories contain entities (forces, fields) that can not be "seen" in data, yet these entities seem to be necessary for predictive success.

— The idea that one can "read the book of nature" without pre-conceptions is a myth. An observation conveys no information unless one knows beforehand the kind of thing one expects to see, but those same pre-conceptions will affect the interpretation of the observations, often in fundamental ways.

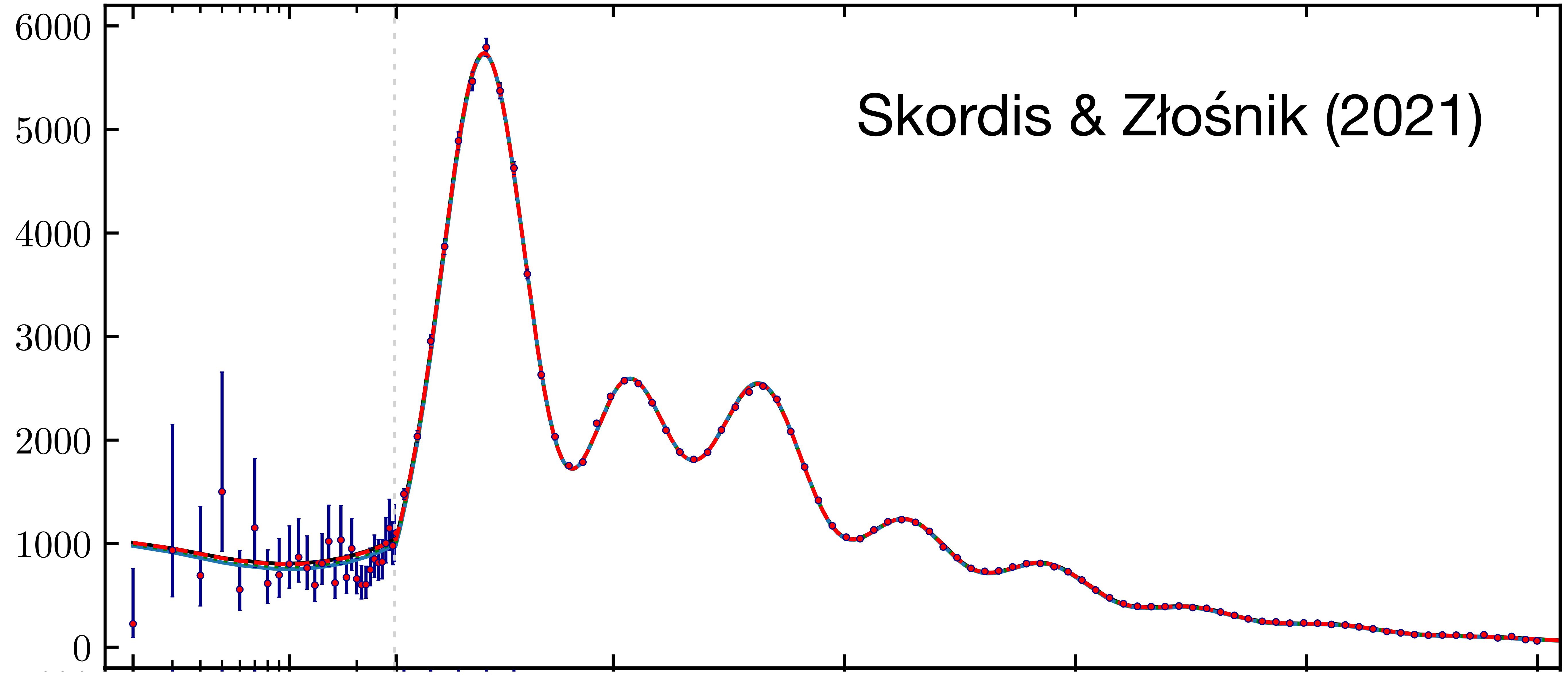
But Popper's strongest criticism of logical positivism was the following:
The attempt to generalize from discrete data to universally valid laws is, by definition, inductivist, and logically, ***induction does not exist***: it is a fallacy, a fake, a fraud.

The fallacy of induction had been established many times, going back at least as far as David Hume in the 18th century, and in fact by the final decades of the 19th century, most philosophers and scientists acknowledged that induction was logically invalid.

But a belief in inductive reasoning had re-surfaced again and again, possibly because no one prior to Popper could think of any *other* way for scientists to use observational or experimental data in the construction of theories.

Popper argued — not only that inductive *derivation* of theories is impossible — but that induction was useless even when it came to ranking theories in terms of probable truth (what the logical positivists called "inductive confirmation"). A theory might make correct predictions for centuries (as did Newton's theory of gravity), but even a perfect record of success never raises above zero the theory's probability of being correct.

In fact Popper argued that one expects always to be able find a large (formally, an infinite) number of theories that can explain any finite set of data.



Power spectrum of temperature fluctuations in the cosmic microwave background (CMB)

— $\mathbf{\Lambda}$CDM (Einstein gravity + dark matter)

— [R]MOND (no dark matter)

Data

Here is an example that illustrates Popper's point. The CMB spectrum is reproducible (in the sense that it can be fit, by adjusting a set of parameters) under the standard, or $\mathbf{\Lambda}$CDM, theory of cosmology. That theory postulates the existence of dark matter.

But the CMB data can equally well be fit by an alternative theory ([R]MOND) that makes no such assumption. That theory postulates instead a modification of Einstein's theory of gravity.

Popper argued that indeterminacy of this sort was the usual circumstance, supporting his contention that universal statements (for instance, about dark matter) could never reliably be inferred from data.

Given the inherent underdetermination of theory by data, how does one proceed? Popper's answer was: by trial and error — or (as he put it) by **conjectures and refutations**.

False theories can make successful predictions. But a false theory will also make at least some *incorrect* predictions. And if a single prediction is refuted by data, then the predicting theory is wrong — it has been **falsified**.

Popper argued that methodologically, this was all one needed. Falsification has the logical stature attributed by the logical positivists to verification — except, of course, that falsifying a theory is a *deductive* operation (one generates a theory's predictions via a deductive procedure) and so there is no need to tie the methodology to some nebulous, and of course invalid, justification of induction.

Popper argued that one should never seek to judge a scientific hypothesis (as the positivists did) on the basis of its origin, on how it was produced. Theories can come from anywhere, he said, and in fact many accepted hypotheses began their lives as metaphysical ideas, long before they had any empirical support; examples include the atomistic hypothesis, and the corpuscular theory of light. (Einstein was in full agreement with Popper on this point.) The genesis of a scientific hypothesis does not matter: what matters is only how well the idea, once proposed, stands up to criticism.

Popper called this attitude **critical rationalism**. A critical rationalist judges theories (or ideas generally) on the basis of how well they stand up to critical appraisal. Popper opposed critical rationalism to **verificationism**, in which

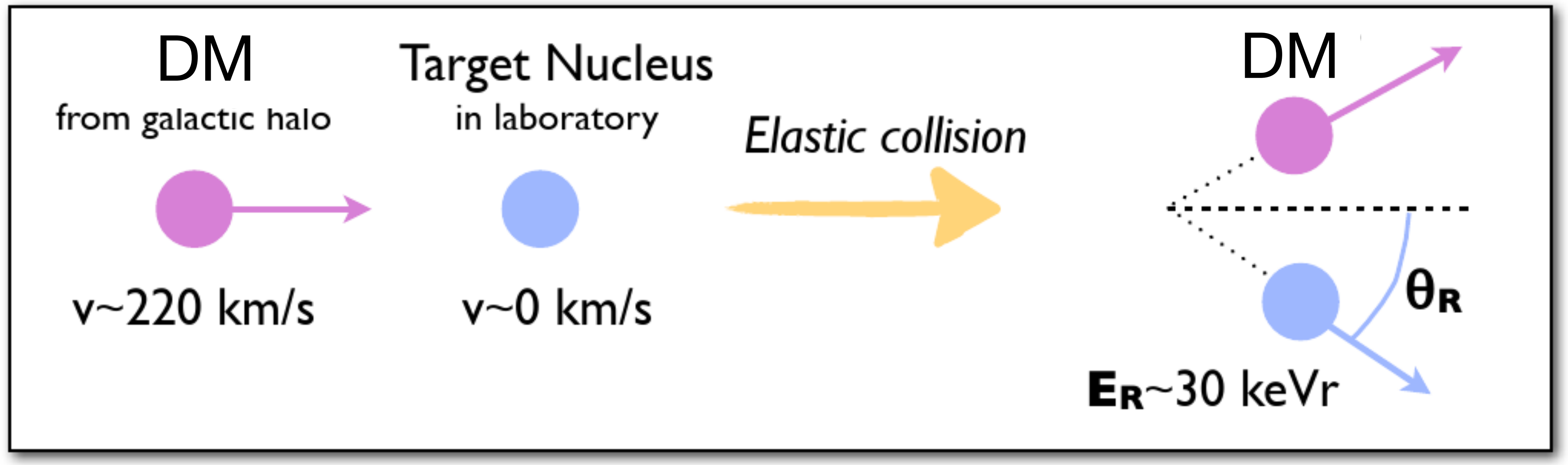
> one looks out for ‘verification’ or ‘confirmation’ or ‘instantiation’, and one finds it, as a rule. Every observed ‘instance’ of the theory is thought to ‘confirm’ the theory.

But in order for critical rationalism to work, *theories must be falsifiable*. That is: there must be some predictions of the theory that are amenable to testing, and which, if refuted by data, would justify abandoning the theory. Otherwise there is no hope of progress.

This requirement is as much a *methodological* one as a *logical* one. As an example, Popper cited the case of Sigmund Freud’s psychoanalytic theory. Freud, Popper noted, could have stated what observable situations (regarding human behavior) would be inconsistent with his theory. But instead Freud (and his followers) always saw confirmation of the theory wherever they looked. Every conceivable behavior was interpreted as verification. With such an attitude, Popper noted, there was no hope of ever *improving* Freud’s theory — and continued growth of theories was (to Popper) an essential aspect of science

The importance of falsifiability can be illustrated by a more recent example.

`Direct detection' experiments seek to test the hypothesis `Dark matter particles are present in my laboratory' by looking for evidence of the particles' passage through a detecting medium.



Dark matter particles could be detected via well-established techniques based on scintillation, ionization or calorimetry. Examples of such experiments include XENONnT, DarkSide-50, and PandaX.

Now, a statement like `Dark matter particles exist' is an example of what philosophers call an **existential hypothesis.** *In isolation*, an existential hypothesis is neither confirmable nor refutable (What is a dark matter particle? How can you tell if you've seen it? Or that you haven't seen it?). An existential hypothesis becomes *testable* only to the extent that it is *embedded in a theory*, which says something about the universal qualities or dispositions of the putative entity: where it can be found, how it behaves etc. And the *degree* to which the hypothesis (including the embedding theory) is testable depends on how complete or detailed the theoretical description of the entity is.

In the early days of the direct detection experiments (c. 1980-1985), there *was* such an embedding theory. Dark matter particles were identified with the lowest-mass particles ("WIMPs") in so called super-symmetric (SUSY) theories of particle physics. Those theories (which had, and continue to have, no empirical support) could be used to calculate the properties of the putative particles: their mass, their interaction rate with normal matter etc. These predictions were definite enough that experimenters could test them, which they did, and found that they were wrong.

Theorists accepted that their "minimal" version of a SUSY theory had been falsified, and constructed new versions, which predicted only particles of smaller interaction cross section. Again these predictions were refuted and again the theories falsified.

By the early 2000s, most particle theorists had chosen to abandon this effort: they saw little prospect that SUSY theories could yield anything resembling dark matter. Once this happened, the hypothesis `dark matter particles exist' became decoupled from a theory that was able to predict the particles' properties. That hypothesis is still (potentially) *confirmable,* by detecting a new particle that has a mass and a number density whose product is consistent with the extra mass needed to explain the Milky Way's rotation curve. But to the extent that it is divorced from a predictive theory, the hypothesis is no longer *falsifiable*: one can always `explain away' a non-detection by supposing (for instance) that interaction of the particles with normal matter is so weak as to be undetectable. And this non-falsifiability means that there is a possibility that the experiments will never either confirm, nor disconfirm, the existence of the particles: that is: that they may tell us nothing at all about dark matter.

Popper argued that falsifiability was the best criterion for distinguishing science from non-science. But whatever one thinks of his **criterion of demarcation**, Popper's insights about the consequences of falsifiability (or its absence) are remarkable, profound, and far-reaching:

Popper pointed out that the ability of a theory to make testable predictions is *entirely* a consequence of its ability to *rule things out* — to say that certain things never occur. In fact he defined what he called the **empirical content** of a theory — roughly speaking, its ability to make testable predictions — purely in terms of its **potential falsifiers**: statements which, if experimentally confirmed, would falsify the theory.

A theory that explains everything, explains nothing. (Think: Freud's theory) The more restrictive a theory is — the more events that it excludes from the realm of the possible — the more it says about the world: the greater its predictive and explanatory power. And the more things a theory excludes, the greater its potential for falsification. Thus, Popper argued, the following virtues of a theory rise and fall together:

Falsifiability — testability — predictive power — explanatory power

Popper also argued, convincingly, that all of the above virtues vary *inversely* with a theory's probability of being correct. This is precisely opposed to the currently popular view that one can use Bayes's theorem to find the `best' (i.e. most probable) of competing theories.

Suppose, then, that we are dealing with a falsifiable theory, and that the theory is faced with an anomaly: a refuted prediction: a falsification. In its simplest form, Popper's proposed methodology consists of the following steps:

1. Propose a new theory that explains the anomaly, and that is *as bold as possible*; that is, which goes far beyond the original theory in terms of its empirical content.

   — Here Popper satisfied the first of our two requirements for a successful scientific methodology: his method is **ampliative**, by fiat.

2. Derive a testable, novel prediction from the new theory. Focus on ***surprising*** predictions: predictions that conflict with prior expectations. These are just the predictions that are most likely to be wrong, and most easily refutable.

3. Test the prediction. If it is refuted, reject the theory, and go to step 1.

4. If the prediction is confirmed, accept the theory (subject to further tests) and return to step 2.

What about the second requirement of a successful methodology — that it be **epistemically probative** — that it give us confidence that the new theory is correct?

Popper argued that a modification to a theory that does nothing more than target an anomaly is ***ad hoc***: there is no warrant for believing that the explanation is correct. (Think: Ptolemy's epicycles and equants.) Corroboration, he argued, *only* occurs if the new theory has *new, untested* consequences, and if *some of those novel predictions can be shown to be correct*. In other words:

Failed refutation = confirmed novel prediction = corroboration

He argued further that the *degree* of corroboration depends on how unlikely the new prediction is, when viewed from the standpoint of the old theory and other "background knowledge". For instance: the new theory might imply a relation between two variables (e.g. pressure and temperature of a gas), while the old theory predicted no such relation.

Note that, according to Popper's argument, **simple agreement of a theory with data, by itself, counts for nothing**: claims that a theory is *supported* by data require that a higher standard be met. Again, here, Popper departed strongly from the inductivists, who argued that *any* agreement of theory with data is probative.

Here is a schematic diagram of how corroboration occurs according to Popper:

New theory *[makes]* **new prediction** *[which is confirmed by]* **new data**

time - - - > .... - - - >

As originally envisioned by Popper, this is a *temporal sequence*. But as many philosophers soon pointed out, there are famous examples from history where data that came to be seen as strongly supportive of a new theory, were available *before* that theory was proposed. For instance: Kepler's laws were known to Newton; Einstein knew about Mercury's anomalous orbital precession; the emission spectrum of hydrogen was known to Niels Bohr.

The more essential requirement, it was argued, is that *data expose a theory to potential falsification*, whether or not those data were available before the theory was proposed. Stated differently: only if a theory was not specifically engineered to explain a fact, can that fact be taken to support the theory.

There is now widespread agreement among philosophers of science that — even if a theory correctly predicts or explains a set of data or an observed phenomenon — the following two conditions should also be met, if the data are to be seen as *supporting* the theory:

1. No information about the phenomenon was used in the construction of the theory (sometimes called the "independence condition")

   As philosopher John Worrall put it: "One can't use the same fact twice: once in the construction of a theory and then again in its support"

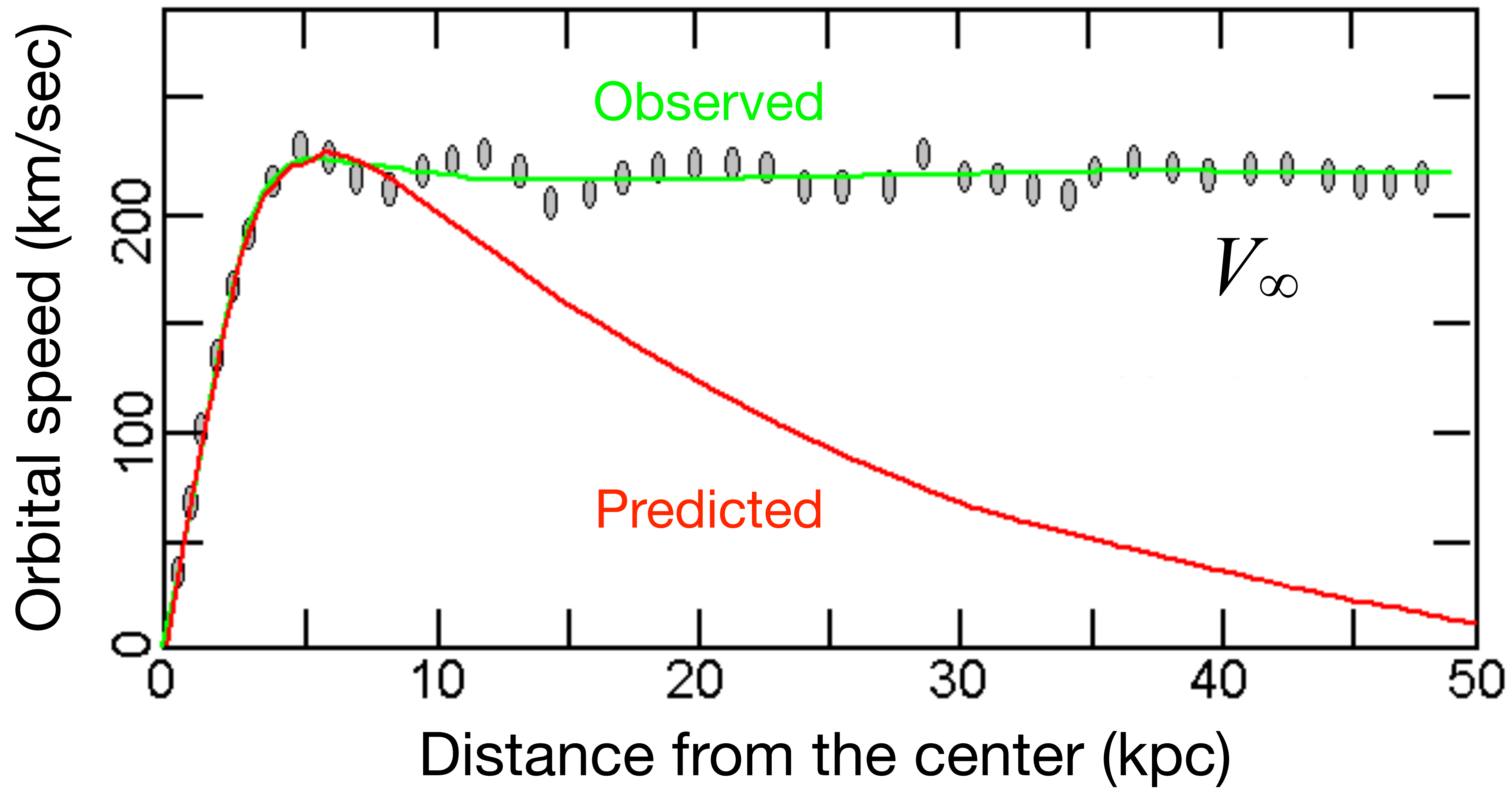
2. There exists no alternative theory that provides a non-contrived explanation of the data (sometimes called the "uniqueness condition")

   As philosopher Jarrett Leplin put it: "Truth is not to be attributed to a theory in explanation of its explanatory success if the result explained can also be explained another way"

Here is an example that illustrates all of these points.

The MOND theory postulates a modification to Newton's laws of gravity: the effective gravitational force from any distribution of matter is postulated to be greater than predicted by Newton's laws in regimes where the gravitational acceleration (force per unit of mass) falls below a certain universal value, called $\boldsymbol{a_0}$, or ***Milgrom's constant***.

Milgrom *designed* his proposed modification in order to reproduce a well-known (in 1983), and apparently universal, phenomenon: the fact that the orbital speed in spiral galaxies (like the Milky Way) tends to a constant value at large distances from the center of the galaxy.



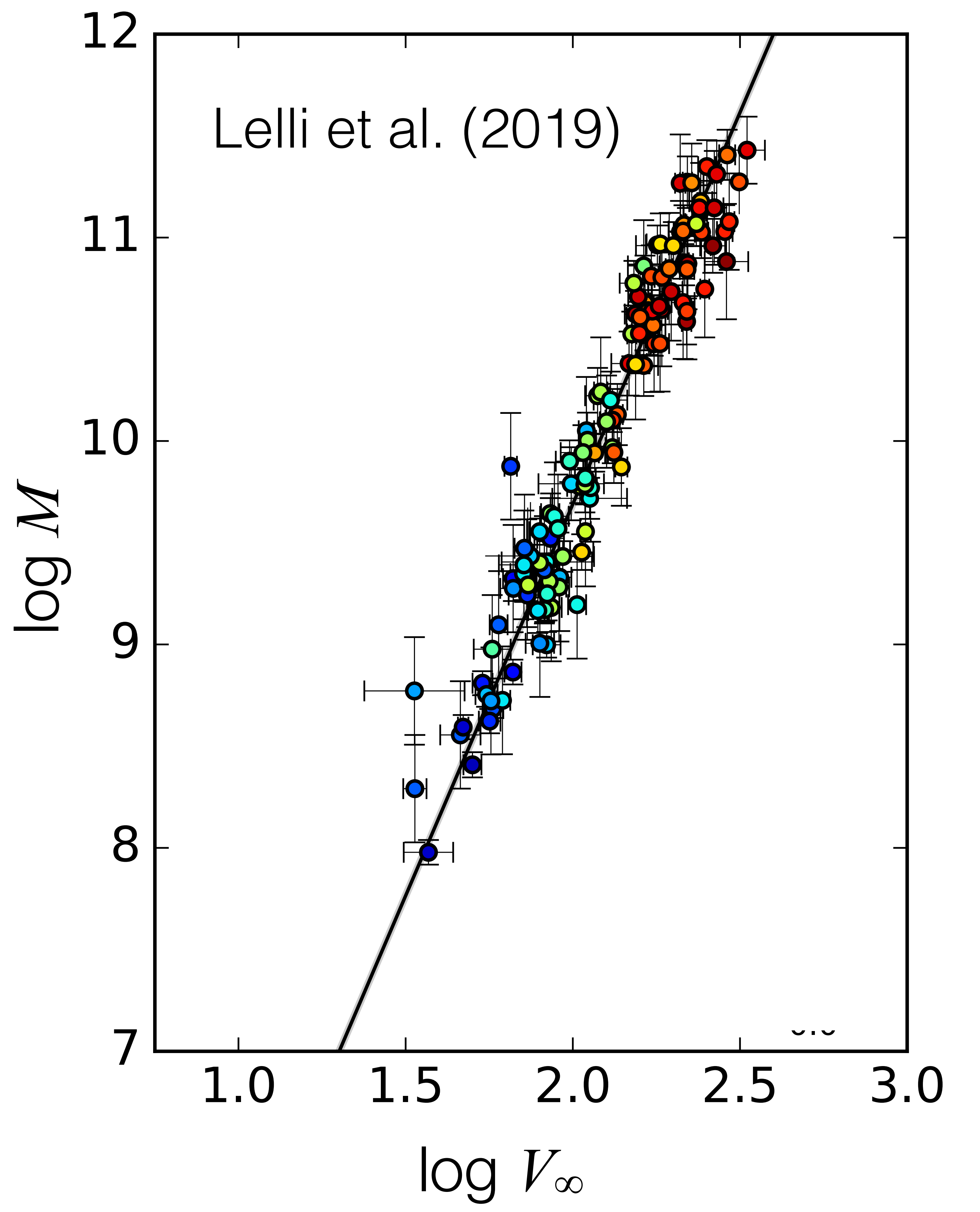
The "asymptotic flatness" of galaxy rotation curves is illustrated by this idealized figure.

Because Milgrom's theory was designed to explain this feature, the "independence" condition above implies that the observed asymptotic flatness does not support Milgrom's proposed changes to Newton's laws.

But the same MONDian postulates that imply asymptotic flatness of rotation curves, also contain a great deal of additional empirical content: they entail a number of novel predictions—and hence satisfy Popper's requirement that theory modifications should do more than simply explain a known anomaly. For instance, MOND predicts the following universal relation between the asymptotic orbital speed, V_∞, and the galaxy's total mass, M:

$$V_\infty = (GMa_0)^{1/4}$$

(MONDian astronomers call this the MASSR, or "mass-asymptotic speed relation"; standard-model cosmologists call it the BTFR, or "baryonic Tully-Fisher relation".)



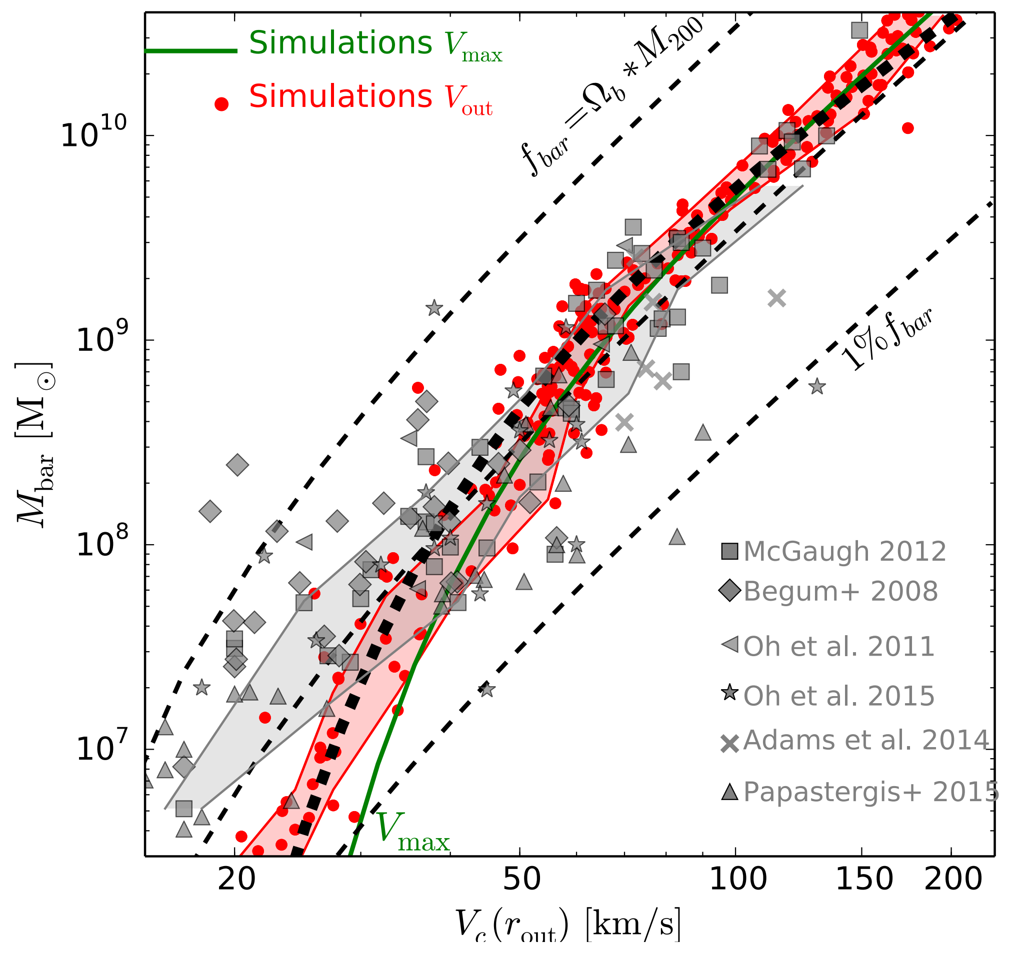
As shown by the figure, this novel prediction of Milgrom's theory has been confirmed; and since MOND was not *designed* to explain the phenomenon (indeed, no such relation was known to exist in 1983), it fully satisfies the "independence" condition. The confirmed existence of the relation therefore provides support for the correctness of Milgrom's postulates.

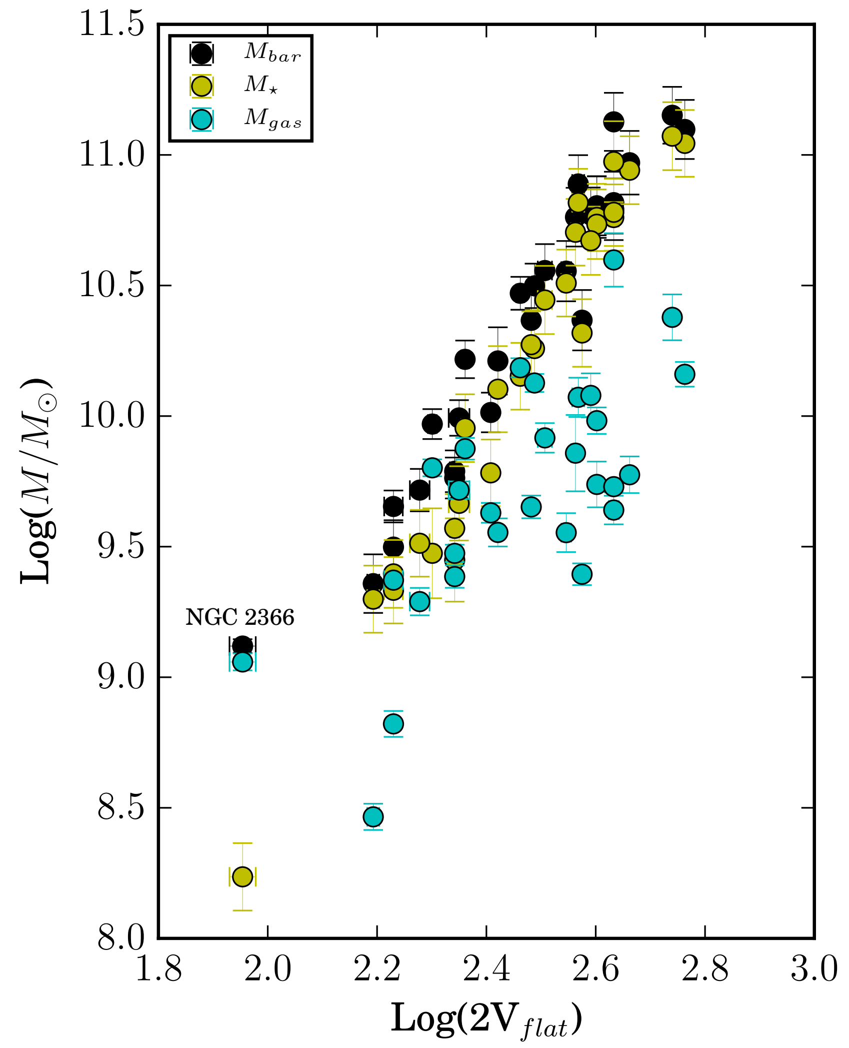
The MASSR also fully satisfies Popper's preference that a novel prediction be "surprising". Under the ΛCDM theory, the quantity V_∞ should be determined almost entirely by the *dark* mass of a galaxy, not by its *normal* mass M. And yet it is.

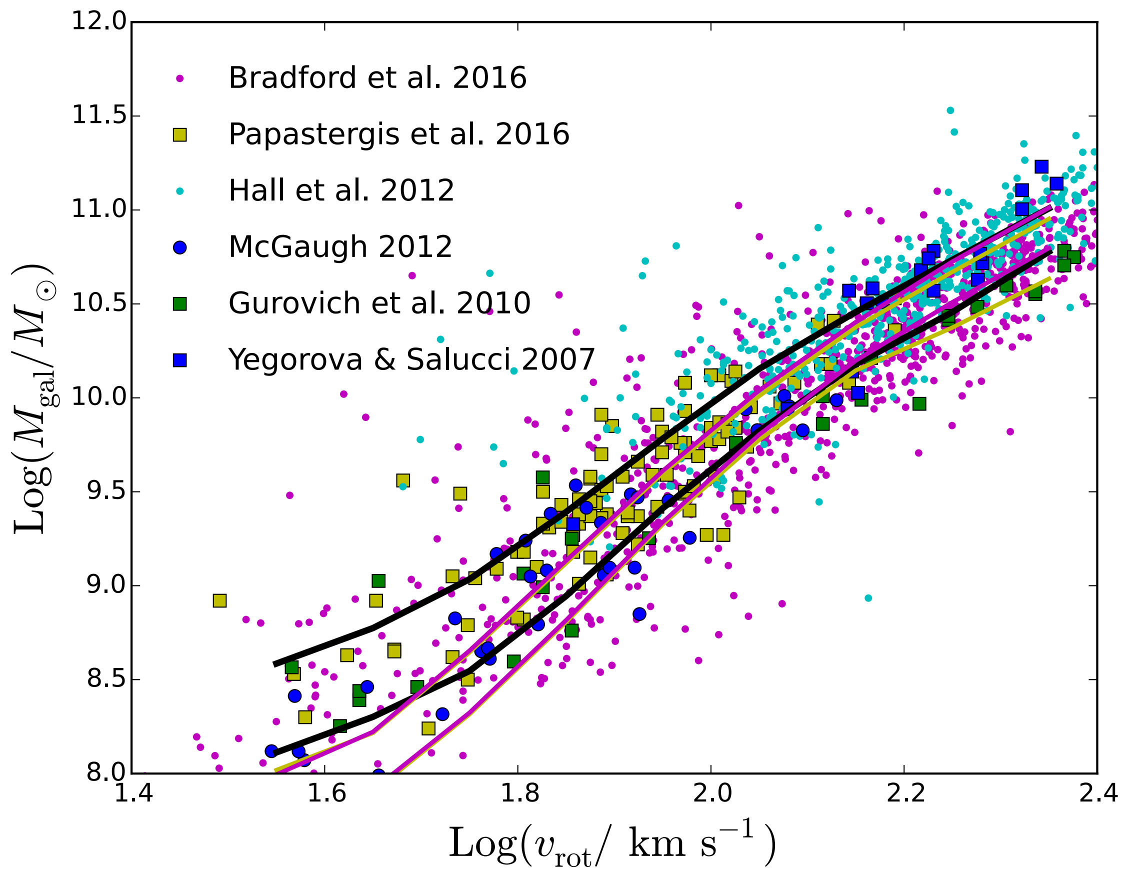
What about the second of our two conditions for data to support a theory — the "uniqueness" condition — the requirement that the competing (ΛCDM) theory *not* provide a natural explanation for the same data?

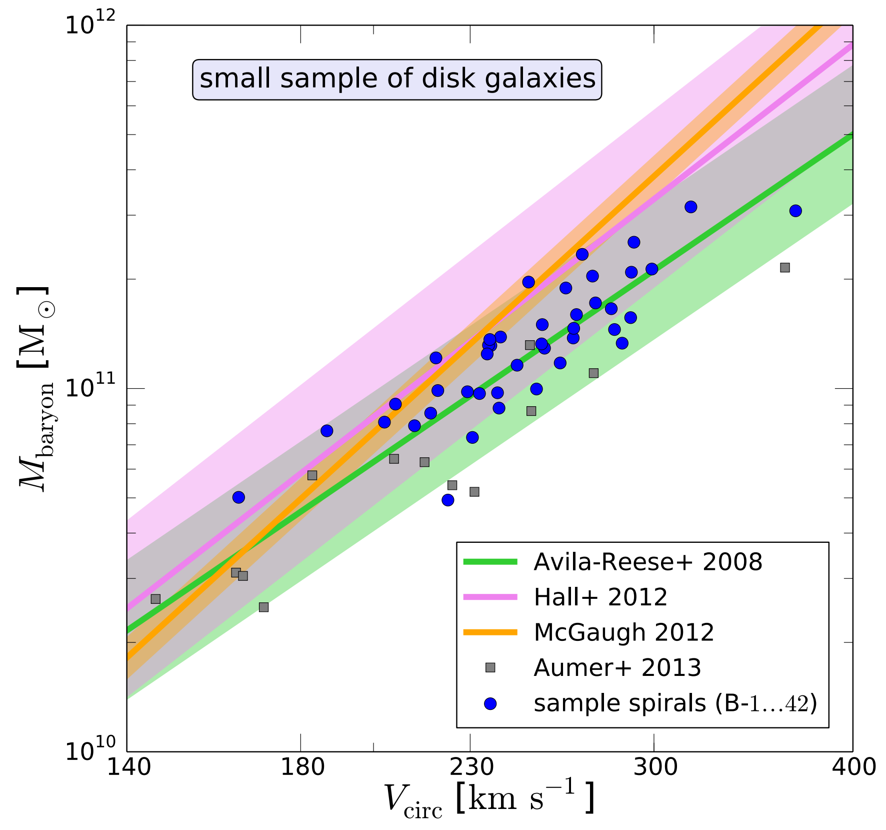
The only route available to standard-model cosmologists for explaining a relation like the MASSR is via large-scale computer simulations of the formation and evolution of galaxies. And as the examples on the right illustrate, every such study yields a different result — some of which are reasonably close to the observed relation; some of which are not.

ΛCDM









Using large-scale simulations to explain relations like the MASSR is very much like predicting the weather: each simulator will obtain a different result, and it is typically not possible to identify exactly why. (As the philosopher Eric Winsberg put it: “The end results of simulations often do not bear a simple, straightforward relation to the theories from which they stem.”) The ΛCDM theory’s prediction depends on who asks the question. Whereas Milgrom’s theory makes a definite, falsifiable prediction—confirmable if correct, refutable if incorrect.

As usual, Popper provides the best language for discussing situations like this. Recall his argument that the following virtues of a theory are closely linked:

Falsifiability — testability — predictive power — explanatory power

The MONDian prediction (the MASSR) is easily testable, and having passed the test, that theory explains why the two variables (M, V_∞) should be functionally related, as they are observed to be. Whereas the standard-model explanation in terms of dark matter ranks low in each of the categories. Since it is not clear exactly *what* that theory predicts, its prediction is only weakly testable: it is impossible to say either that the theory does, or does not, entail the observed relation, and thus the theory gives no compelling explanation for why a relation between those two variables should exist, or why it should have the form that it is observed to have.

Although they may use language that differs from Popper's, philosophers are in good agreement that prior prediction of a fact (in this case, via MOND) is epistemically more probative than post-hoc accommodation of that fact (in this case, via ΛCDM).

For instance, philosopher Peter Lipton writes:

"When data need to be accommodated… the scientist knows the answer she must get, and she does whatever it takes to get it ... In the case of [prior] prediction, by contrast, there is no motive for fudging … As a result, if the prediction turns out to have been correct, it provides stronger reason to believe the theory that generated it."

John Worrall writes:

"when one theory has accounted for a set of facts by parameter-adjustment, while a rival accounts for the same facts directly and without contrivance, then the rival does, but the first does not, derive support from those facts."

And of course we can cite Popper here:

"We choose the theory which best holds its own in competition with other theories … This will be the one which not only has hitherto stood up to the severest tests, but the one which is also testable in the most rigorous way."

All of these statements favor (in this instance) MOND over ΛCDM.

The logical complement to corroboration of a theory is falsification. Nowadays, many philosophers seek to deny the importance of falsifiability, on the basis of the following argument: If data conflict with theory, one never knows for certain where the problem lies. It might be in the theory, or it might be due to some extraneous factor: a mis-interpreted experiment (e.g. the spurious claims to have observed magnetic monopoles), a mis-interpreted theory (recall how long it took theorists to agree on what Einstein's theory actually says about gravitational waves), or stupid mistakes of one sort of another.

This argument is correct in principle. But scientists are quite aware of the issue, and (as in the two examples just given) will devote enormous effort to identifying any confounding factors that complicate the interpretation of an experimental result. (In so doing, they are acting as critical rationalists.) And it is undeniable that falsifications *have* occurred, and that they have sometimes been of enormous importance in influencing the development of theories.

Here are two well-known falsifications of the ΛCDM cosmological model:

- The demonstration, *c*. 1975-80, that galaxy rotation curves do not behave as predicted by Newton's laws
- The demonstration, *c*. 1997-98, that the expansion of the universe does not behave as predicted by Einstein's laws

Both of these falsifying instances occurred when the ΛCDM model was tested in a new regime with respect to the included theory of gravity. One might therefore suppose that cosmologists responded to the falsifications by modifying their theory of gravity. But they (at least, most of them) did not. Their collective response, in each case, was to leave the gravitational theory untouched, and to postulate the existence of a new entity: `dark matter' in the first case, `dark energy' in the second.

When seeking to understand why most cosmologists chose to respond in this way, the philosophers have an important insight to provide.

The philosopher Imre Lakatos (1922-1974) noted that scientific theories tend to evolve in characteristic ways in response to falsifications. Typically there is a fixed set of assumptions, which Lakatos called the **hard core** (and which Thomas Kuhn, at least sometimes, referred to as a "paradigm"). When a theory has been falsified, 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. In effect, the adherents of a particular theory choose, collectively (and perhaps tacitly), to 'shield' some components of their theory from falsification. One consequence of a fixed hard core is that theories maintain a recognizable character over time in spite of being gradually modified. Lakatos called the sequence of theories that share a fixed hard core a **research program**.

Lakatos never discussed cosmological theories (he died in 1974) but his insight applies perfectly to the standard cosmological model, if we identify Newton's/Einstein's descriptions of gravity as part of that theory's (or rather, that research program's) hard core.

Lakatos (in agreement with Popper) believed that scientists' choices about which elements of their theory to shield from falsification were arbitrary and essentially unjustifiable. We can find some support for Lakatos's position by considering the following facts:

By assuming the correctness of Einsteinian gravity, standard-model cosmologists are led to postulate the existence of dark matter. No known, elementary particle has the right properties to constitute the dark matter, hence it is assumed that some *new* particle — which, necessarily, lies outside of the standard model of particle physics — must exist, and hence that the currently accepted theory of particle physics is wrong.

Thus, by assuming the correctness of Einstein's theory, standard-model cosmologists are led to assume the *in*correctness of the standard model of particle physics.

MONDian researchers, on the other hand, have no quibble with the standard model of particle physics. They choose instead to modify the theory of gravity.

Is it possible to make a case that one or the other of these two choices for the ‘hard core’ is more warranted? Here is an argument that philosophers sometimes make when discussing theory choice:

There are, and have long been, many alternatives to Einstein’s gravitational theory. Review articles that do no more than summarize the viable alternative theories can run to hundreds of pages. Whereas there currently exists no comparably viable alternative to the standard model of particle physics.

Suppose (as a crude approximation) we assume that the probability that a given theory is correct is equal to $1/N$, with N the number of viable alternative theories. Then we would conclude that of these two theories — the current standard model of gravity (Einstein’s), and the current standard model of particle physics — the one more likely to be correct is the latter. And it would follow that the choice by ΛCDM cosmologists to assign Einstein’s theory to their hard core is the less warranted choice.

This argument is probably not compelling enough to change anyone’s mind (and I would not advocate it in any case). But at the very least, it makes a *prima facie* case that neither of these two choices of what to assign to the hard core is obviously the correct one; and of course, it is possible (as Popper would have insisted) that *both* Einstein’s theory, and the standard model of particle physics, are incorrect.

Whatever one's choice of elements to assign to the hard core, Popper's prescription for judging the success of a (new) theory is clear: one assesses how well the theory stands up to critical tests; that is: how successfully it makes novel predictions that turn out to be correct. Lakatos used the word **progressive** to describe theories that evolve via predictively successful modifications, and **degenerating** to describe theories that evolve primarily via post-hoc accommodations in response to whatever new data happen to come along.

It would be inappropriate, in a general talk on methodology, to say much more about which of these two cosmological theories is most successful according to Lakatos's criteria. But if one had to make a quick choice, the front-runner would be MOND, which has made a string of surprising, and successful, novel predictions. Again and again, MOND has exposed itself to potential falsification and survived the test. The standard cosmological model, by contrast, has dealt with anomalies almost exclusively via post-hoc accommodation. The assumed properties of dark matter and dark energy, for instance, have been modified again and again since these entities were added to the ΛCDM model *c*. 1978 and 1998 respectively.

Now, Lakatos's view of the arbitrariness of the hard-core assumptions, and of the need to subject those assumptions to critical testing, was generally accepted by philosophers throughout the latter half of the 20th century. But it turns out that the *current* generation of philosophers often espouse a very different view: to put it crudely: that the majority of scientists know what they are doing and should not be second-guessed!

For instance, here is a 2012 statement by the philosophers Nora Berenstain and James Ladyman:

> In rough terms, realists are those who take the theoretical claims of science at face value and who regard scientific theory choice as the final word in the epistemology of particular scientific theories and statements.

By “realism”, these authors mean something rather more specific than the belief that there is an objective reality. `Scientific realism’, since about the mid-1980s, means (roughly speaking) that *reality is accurately described by our current, `best’ theories of science*. For instance, Richard Boyd writes:

> Scientific theories, interpreted realistically, are confirmable and in fact are often confirmed as approximately true by ordinary scientific evidence interpreted in accordance with ordinary methodological standards.

And John Worrall writes:

> 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.

One implication is that unobservable (or unobserved) entities like dark matter and dark energy — which philosophers sometimes refer to as `theoretical terms’ — should be accepted as real, since they are components of our current, `best’ theories, and those theories are unlikely to change in any fundamental ways going forward.

The currently fashionable view that “history is over” when it comes to scientific theories would have been strenuously resisted by Karl Popper, who argued that “the rational and empirical character of science would vanish if it ceased to progress.” Why then have so many philosophers moved toward a realist — that is, a Panglossian — view of scientific theorizing? I don’t know the answer; although the cynic in me wonders whether it is not simply a reaction to Popper.

In any case: if one believes that “mature" theories in the sciences are (at least approximately) true, then one has little use for a methodology that encourages bold theorizing. Bold conjectures, of the sort that led to the theory of relativity or to quantum mechanics, can (if successful) require abrupt changes in ontology: that is: changes in our basic assumptions about what the world consists of. And so scientific realists are motivated to search for alternative, `gentler’ methodologies that never take theories very far beyond their current (presumed nearly true) forms, and which are unlikely to require the addition of entities beyond those whose existence has already been agreed upon. To quote another realist philosopher: “If our best science is not our best guide to our ontological commitments, then nothing is.”

To find a satisfactory methodology, realist philosophers have had to reach far back in time — to the mid 19th century at least, well before the time of Karl Popper. And their currently favored stand-in for Popper is the American philosopher **Charles Sanders Peirce** (1839 – 1914).

Peirce operated in a world that had not yet experienced the early 20th century revolutions due to Einstein and Bohr that so strongly influenced later philosophers like Popper, Kuhn, and Lakatos. In his proposals for a scientific methodology, Peirce attempted to transcend inductivism, but he never managed to progress very far beyond it. He is credited with coining the terms **abduction** and **abductive reasoning**, which he defined, roughly, as follows: denoting evidence by E and a hypothesis by H, argue that

1. H explains E
2. ∴ There is reason to believe H

This mode of reasoning was not, of course, original with Peirce; for instance, Thomas Burnet (1635-1750), in his *Sacred Theory of the Earth* (1684-90), supported his claims for the origin and evolution of the Earth purely on the basis that they were consistent with the evidence. He did not seem to appreciate that his hypothesis about what the Earth was like in ancient times was just one possible explanation among many, and that deciding between them would require submitting them to comparative tests.

Peirce followed Burnet — and, for that matter, all inductivists — in claiming that one could argue for the correctness of a hypothesis *simply on the basis that it entails the desired outcome*.

Twentieth- and 21st century admirers of Peirce have recognized that multiple hypotheses will usually be consistent with any finite set of data, and hence that there is a need to select between them. But rather than follow Popper's advice (bold conjectures followed by critical testing), they have chosen to modify "abduction" into what is usually called **inference to the best explanation**, or IBE:

1. H explains E, *and*
2. H is the *best* explanation for E
3. ∴ H is (*likely, probably,* …) true

Note that, just as in the case of abduction (or induction), there is *no requirement that a hypothesis be subjected to testing*. One can choose the most likely explanatory hypothesis (we are told) simply by inspection of the problem-situation that motivated the hypothesis.

Now, the claim that scientists follow the mode of reasoning just described is rampant in the current philosophical literature; for instance, Anjan Chakravartty writes, in a recent review article: "Inference to the best explanation ... seems ubiquitous in scientific practice." Scientists are likely to react to such a statement with "That is news to me!" followed by "Please tell me what is meant by the *best* explanation."

Many philosophers acknowledge that the idea of a `best' explanation is nebulous. But typically, the `best' explanation is defined as the *most likely* among the set of explanations that are consistent with the available data. And likelihood, in turn, is to be evaluated on the basis of background knowledge or assumptions. For instance, Psillos writes:

> Suppose there are two potentially explanatory hypotheses H_1 and H_2 but the relevant background knowledge favours H_1 over H_2. Unless there are specific reasons to challenge the background knowledge, H_1 should be accepted as the best explanation.

Under "background knowledge", Psillos includes both known facts, and accepted theories ("the laws that govern certain phenomena"). According to this interpretation, the `best' explanation is one that targets the anomaly and explains it in a manner that requires *as little change as possible to accepted theories*.

Of course, such a methodology — one that seeks explanations that go as little as possible beyond the facts to be explained — is precisely what Popper called *ad hoc*, and advised strongly against. But it is easy to see how this alternative methodology fits hand-in-glove with scientific realism, which posits that major changes in accepted theories are no longer to be expected. And if an explanatory hypothesis has little content beyond that which is needed to explain the anomalous data, there is little need (one can plausibly argue) for subjecting that hypothesis to critical testing.

What then are the “accepted theories” that constitute the background knowledge? The answer seems to be: Whatever the majority of scientists think they are. Or stated in Lakatosian terms: whatever theory(ies) the scientists have assigned to their hard core. Here is how one prominent abductivist, Ilkka Niiniluoto, justifies an explanation of the rotation curve anomaly in terms of dark matter:

> Here the [background] theory T is Newton’s mechanics … Anomalous evidence E has led to the explanatory hypothesis that the universe consists of only about 5% of ordinary matter and the rest is dark matter. The majority of physicists accept this abductive inference to the best explanation.

This, in its entirety, is Niiniluoto’s argument for dark matter. Like Thomas Burnet in the 17th century, he does not seem to worry that — because his “best” explanation is not unique — it needs to be critically tested. His argument, apparently, is “Dark matter is the best explanation because it is based on the theory of gravity which the majority of physicists believe to be correct”. Epistemology has been reduced, here, to the uncritical acceptance of whatever the majority of scientists believe.

Some philosophers trace the antecedents of abductive reasoning as far back as Aristotle’s *epagōgē* (*ἐπᾰγωγή*). I am tempted to go back just a little farther. Recall that Plato, in his *Republic*, asked the questions: Who should rule? Who should decide what is just or true? His answer was: the intellectual elite; the wise men (not *women*); the “philosopher-kings”. (“The wise shall lead and rule, and the ignorant shall follow.”) It is remarkable that, in side-stepping Popper, philosophers of science have seen fit to resurrect an epistemology that, much like Plato’s in the fifth century BCE, is authoritarian, dogmatic, and maximally protective of the scientific status quo.

Suggestions for further reading

For anyone interested in scientific methodology and epistemology, Karl Popper's work is an essential resource. All of his books are still in print and all have gone through multiple editions (some with new appendices or notes by the author) and, often, multiple publishers. Dates and publishers given here are for the original printings.

His 1934 *Logik Der Forschung*, translated into English as

The Logic of Scientific Discovery (London: Hutchinson and Company, 1959)

is without question the single most important and original contribution to the philosophy of science ever written. The *LSD* was followed by three important books containing essays that Popper viewed as "postscripts" to the *LSD*, but which can be read and understood independently. These are:

Conjectures and Refutations: The Growth of Scientific Knowledge (London: Routledge and Kegan Paul, 1963)
Objective Knowledge: An Evolutionary Approach (Oxford: Clarendon Press, 1972)
Realism and the Aim of Science (London: Hutchinson, 1983)

Popper's important theory of corroboration is to be found mostly in Chapter IV of *Realism*. That chapter is essential reading for any scientist engaged in theory testing (which is to say: every scientist everywhere).

The two-volume

The Philosophy of Karl Popper, ed. P. A. Schilpp (La Salle: Open Court, 1974)

Is a treasure trove. About one-half of the 1323 pages are by Popper; the remaining pages are mostly critical essays by other philosophers and scientists, to which Popper responds.

The most important insights on scientific methodology after Popper were those of Imre Lakatos, at one time a student of Popper's and later a colleague. Lakatos confronted Popper's falsificationism with the historical record, and in two long essays — "Falsification and the methodology of scientific research programmes" (1970) and "History of science and its rational reconstruction" (1971) — worked out his "methodology of scientific research programs", presenting a number of new and important ideas in the process: about hard cores, positive heuristics, and progressive vs. degenerating research programs. His two essays are included as Chapters 1 and 2 in the posthumous

The Methodology of Scientific Research Programmes, Philosophical Papers Volume 1, ed. J. Worrall and G. Currie (Cambridge: Cambridge University Press, 1978)

Among his other contributions, Lakatos's essay "Changes in the problem of inductive logic", printed as the final chapter in

The Problem of Inductive Logic, ed. I. Lakatos (Amsterdam: North-Holland, 1968)

is a fascinating and brilliantly insightful history of the failed efforts of R. Carnap and others to develop logical positivism.

For anyone interested in the currently fashionable trend toward "scientific realism", these three books are as good as any:

A Novel Defense of Scientific Realism by Jarrett Leplin (Oxford: Oxford University Press, 1997)
Scientific Realism: How Science Tracks Truth by Stathis Psillos (London: Routledge, 1999)
Critical Scientific Realism by Ilkka Niiniluoto (Oxford: Oxford University Press, 1999)

Warning: anyone who cut their teeth on Popper may experience serious disorientation when reading books like these. Their arguments in defense of inductive reasoning (yes, you read that correctly) range from problematic to risible, and they often present points of view that were first strongly reasoned and fully argued (for or against) by Popper, without citing him.

Suggestions for further reading (cont.)

Following are additional books and articles that were cited, or quoted from, in the talk:

Berenstain, N. & Ladyman, J. (2012). Ontic Structural Realism and Modality. Pages 149-168 of Berenstain, N. & Ladyman, J. (eds.), *Structural Realism*. Springer.

Boyd, R. N. (1984). The current status of scientific realism. In J. Leplin (Ed.), *Scientific Realism*. University of California Press.

Burnet, T. (1684-90). *The Sacred Theory of the Earth: Containing an Account of the Origin of the Earth, And all the General Changes which it hath already undergone, or is to undergo, till the Consummation of all Things : In Four Books.*

Chakravartty, A. (2017). Scientific Realism*. Stanford Encyclopedia of Philosophy*.

Lipton, P. (2004). *Inference to the Best Explanation*. Routledge.

Milgrom, M. (1983). A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *The Astrophysical Journal*, **270**, 365–370.

Niiniluoto, I. (2018). *Truth-Seeking by Abduction*. Springer.

Plato (*c*. 375 BCE). *The Republic* (Πολιτεία).

Psillos, S. (2009). *Knowing the Structure of Nature: Essays on realism and explanation*. Palgrave Macmillan.

Suggestions for further reading (cont.)

Skordis, C. & Złośnik, T. (2021), New relativistic theory for modified Newtonian dynamics. *Physical Review Letters*, **127**, article id.161302

Worrall, J. (1978). The ways in which the methodology of scientific research programmes improves on Popper's methodology. Pages 45–70 of: Radnitzky, G., and Andersson, G. (eds), *Progress and Rationality in Science*. Boston Studies in the Philosophy of Science, vol. 58. D. Reidel.

Worrall, J. (1985). Scientific discovery and theory-confirmation. Pages 201-331 of: J. C. Pitt (ed.), *Change and Progress in Modern Science*. Dordrecht.

Worrall, J. 2007. Miracles and models: Why reports of the death of structural realism may be exaggerated. Pages 125–154 of: O'Hear, A. (ed.), *Philosophy of Science. Royal Institute of Philosophy Supplement Series*, vol. 61. Cambridge University Press.