Turning Old Anomalies Into New Evidence

Evidential Reconfiguration and Dark Matter in 1974

Simon Allzén[1,2,3,‡]

[1]Department of Philosophy, Stockholm University
[2]Institute of Physics, University of Amsterdam
[3]Vossius Center for the History of Humanities and Sciences, University of Amsterdam
[‡]Correspondence: simon.allzen@philosophy.su.se

October 28, 2026

Abstract

The 1974 papers by Ostriker, Peebles, and Yahil (1974) and Einasto, Kaasik, and Saar (1974) are regarded as pivotal in establishing the evidential case for dark matter, although both relied on already existing observational data. By what means did they transform known observational anomalies into evidential support for dark matter? I argue that this transformation arose through *evidential reconfiguration*: extended galactic halos supplied a physical connection between galaxy rotation curves and cluster mass discrepancies, expanding the context of assessment so that results previously investigated locally could constrain one another. The significance of this connection was assessed against differing cosmological commitments and unresolved local dynamical assumptions, which help explain both the emphases of the two papers and their reception. To explicate its contribution to confirmation, I draw on Myrvold's Bayesian account of unification and use a toy model in which taking one result into account changes the assessment of a shared halo parameter, thereby altering expectations about the other. Comparison with an uncoupled model makes the resulting support explicit, while variations of the model show why a shared parameter alone is insufficient: the contribution depends on the connecting assumptions and what competing models predict. The novelty relevant to the old-evidence problem thus lies in what familiar observations became jointly capable of supporting. This reconstruction assesses that support once the connection is specified, while leaving a full account of historical belief change open.

1 Introduction

The nature of $\approx 80\%$ of the matter content in the universe remains unknown. Its presence is inferred from gravitational anomalies across a range of cosmic scales, where the observed motions and structures cannot be accounted for by ordinary matter alone. This putative mass component, known as dark matter, is a core posit of the standard model of cosmology, the ΛCDM (concordance) model.[1] Dark matter is invoked to explain the motions of galaxies, the dynamics of galaxy clusters, and the

[1]Λ denotes dark energy, and CDM stands for cold dark matter, where 'cold' refers to the negligible free-streaming length of the dark matter component.

large-scale distribution of matter in the universe. To date, it has eluded non-gravitational detection, leaving its nature radically underdetermined and making it one of the major open problems in contemporary physics. It nevertheless remains deeply embedded in ΛCDM and its existence is widely accepted by cosmologists.[2]

Textbook histories suggest that dark matter's central role is the result of a largely linear progression of accumulating evidence. Such accounts can make it appear that each observation was treated as direct evidence for an already established hypothesis. In practice, however, the relevant observations were initially often regarded as isolated, context-specific anomalies, not systematically integrated within a clearly defined framework or working hypothesis. The philosophical question is how these local discrepancies became coherent *evidence* for an ontological posit, and under what conditions such a change is justified. From a Bayesian perspective, this question becomes especially pressing when the observations were already accepted and used in constructing the hypothesis: how could their synthesis strengthen its evidential support without introducing new observational data? This paper provides a Bayesian explanation of that transformation in the 1974 dark matter synthesis.

I build on the historical work by de Swart, Bertone, and van Dongen (2017) and de Swart (2020), who identify two seminal papers from 1974 as pivotal for the acceptance and establishment of the dark matter hypothesis. I argue that their proposed halo framework connected phenomena previously treated as local and disconnected, allowing the available observations to jointly constrain a missing-mass hypothesis. I call this transformation *evidential reconfiguration*, and reconstruct its confirmatory contribution by comparing coupled and uncoupled models of the same results. Understanding why scientists pursued the connection also requires attention to their theoretical commitments. For some cosmologists, non-empirical considerations made the cosmic mass budget especially pressing. Their role in motivating the synthesis raises a question about justification alongside the question about confirmation: when were those commitments defensible, and what did the resulting connection give scientists reason to believe?

1.1 Overview

Two features of the reasoning in the period highlighted by de Swart et al. (2017) and de Swart (2020) structure the two main parts of the paper.

The first concerns *theory assessment*. In the mid–1970s, researchers had access to the same observational results, but differed in how strongly they judged these observations to support a missing-mass hypothesis. For some cosmologists, background commitments, including a preference for a closed or near-critical universe, made the mass-budget question especially consequential. Researchers concentrating on local astronomical problems placed greater weight on unresolved dynamical assumptions. These concerns overlapped across disciplines and could affect both prior assessments and judgments about which cross-domain connections were defensible.[3] Whether the resulting differences were *rationally permissible* depends on the justification of the commitments

[2]This tension makes dark matter a fertile case for philosophers of science. See, e.g., Vanderburgh (2003, 2005, 2014a, 2014b), Sus (2014), Jacquart (2021), Merritt (2021a, 2021b), De Baerdemaeker and Boyd (2020), De Baerdemaeker (2021), De Baerdemaeker and Dawid (2022), Antoniou (2023, 2025), Duerr and Wolf (2023), Wolf and Read (2025), Allzén (2021, 2026), Martens (2022), Martens, Carretero Sahuquillo, Scholz, Lehmkuhl, and Krämer (2022), and Vaynberg (2024).

[3]For accounts of indirect or meta-empirical assessment, see Dawid (2013, 2019) and Dawid, Hartmann, and Sprenger (2015).

involved. To assess a background constraint, I suggest asking whether it:

(a) is independently motivated within an established modeling context;

(b) has an established history of combining separate phenomena and models into a globally coherent framework that is empirically supported; and

(c) is treated as defeasible rather than as dogma.

The second concerns *unification*. The 1974 studies proposed extended galactic halos whose masses and extents also mattered to cluster and cosmological mass accounting. What justified treating this connection as support for missing mass? Myrvold (2003, 2017) gives us a Bayesian way to answer that question by distinguishing a common-origin explanation from a hypothesis that makes distinct results informative about one another: *the connection changed what the results could jointly constrain.* I specify the probabilistic requirements for this contribution, including the independence assumption used to represent an uncoupled comparison.

Evidential reconfiguration names the historical transformation of local anomalies into jointly informative evidence, and *structural novelty* identifies the new connection responsible for it. The discussion of community differences explains the background against which the connection was pursued and assessed. The Bayesian reconstruction specifies its possible contribution to support, allowing us to distinguish that contribution from the further problem of modeling historical changes in credence when we return to old evidence.

2 The standard historical narrative

Textbook and popular narratives of dark matter often begin with Zwicky's application of the virial theorem to the Coma Cluster in the 1930s, then follow a small number of later breakthroughs presented as progressively establishing the hypothesis. If the cluster was bound and approximately equilibrated, its galaxy velocities required more mass than the luminosities appeared to supply. The inference depended on that dynamical interpretation, not on directly observing a bound state. In retrospect, this is often treated as the earliest observational basis for non-luminous mass.[4]

The narrative then emphasizes the rotation-curve work of Rubin and Ford (1970) in the 1970s: rotational velocities in spiral galaxies did not decline with distance from the galactic center as expected by Keplerian dynamics. Subsequent observations by Roberts and Rots (1973), Bosma (1978), and others strengthened the case that many spiral galaxies exhibit approximately flat rotation curves (See Fig.1). Stars far from the galactic center orbit nearly as quickly as those closer in, in tension with the decline expected from visible matter alone under simple dynamical assumptions. Textbook presentations often treat these results as decisive vindication of dark matter, although Rubin herself was cautious about any causal interpretation of her measurements (Rubin, 2004), and many astronomers initially explored explanations that did not require positing a new exotic mass component.

The story is completed by incorporating later developments such as gravitational lensing, large-scale structure formation, and the Bullet Cluster observation (Clowe et al., 2006). These are

[4]At the time, Zwicky (1933, 1937) did use the term *"Dunkle Materie"* for unseen mass. However, its conceptual content bears little resemblance to the current concept of dark matter (see Allzén (2024) for a historical overview).

presented as extensions of the Zwicky–Rubin sequence, progressively strengthening the hypothesis until it became a central posit of contemporary cosmology. This narrative has an implicitly, and perhaps accidentally, *Bayesian* structure: an early hypothesis is followed by a linear series of results treated as evidence suitable for conditionalization. I do not claim that historical presentations organized in this way are illegitimate. They serve important pedagogical purposes, using hindsight to select episodes that connect to the current concept and explanatory role of dark matter. They nevertheless obscure an epistemically prior question: when, and how, did particular results come to function as evidence for missing mass rather than as local anomalies or modeling artifacts?

The 1974 episode examined by de Swart et al. (2017) and de Swart (2020) brings this question into focus. The two papers contributed to establishing the dark matter hypothesis without introducing new observational data. My aim is to complement the historical claim *that* the 1974 papers mattered with an epistemic account of *why* they mattered. Because the relevant observations were already accepted, ordinary Bayesian conditionalization on them cannot explain this transformation. We need to examine changes in which assumptions were held fixed, and how connecting existing observations transformed anomalies into evidence.

3 Constraints and the mass deficit

To understand why the dark matter (or missing-mass) hypothesis gained traction in both cosmology and astronomy, it is necessary to reconstruct the cosmological background that shaped theory assessment in the decades leading up to the 1970s. In the early and mid-20th century, cosmology was often guided more by theoretical considerations than by the observation-driven practices of astronomy, and debates frequently appealed to methodological and philosophical commitments (de Swart, 2020). One example is the early acceptance of the cosmological principle, formulated by Milne (1932), which asserts that the universe is homogeneous and isotropic on sufficiently large scales. Today, this principle has substantial empirical support, but in the 1930s it was adopted in part for philosophical reasons. Combining the cosmological principle with general relativity led to relativistic cosmology, which by the 1960s had become the predominant framework, drawing on theoretical work by Alexander Friedmann and Georges Lemaître, and observational work by Edwin Hubble (de Swart, 2020).[5] By the 1950s and 1960s, two non-empirical commitments in particular structured cosmological background assumptions: the *cosmological principle* and a *Machian inheritance*, according to which local inertia is determined globally by the total mass of the universe. In relativistic cosmology, the latter was often taken to favor a closed, finite universe.

These commitments encouraged some cosmologists to interpret empirical results relative to models partly favored for non-empirical reasons. These preferences included both spatially closed models ($k = +1$) and models at critical density. In matter-only cosmologies with zero cosmological constant, $\Omega = 1$ corresponds to the spatially flat Einstein–de Sitter model, at the boundary sometimes described as "just closed" in historical discussions. Keeping this historical usage distinct from positive spatial curvature clarifies the related preferences documented by de Swart (2020):

> Nearing the end of the 1960s, a closed universe was still a much-preferred model for many cosmologists and relativists. "*Philosophically*, there might be a preference", Wolfgang Rindler wrote in 1967, "the choice k = 1 [a positively curved universe] might appear

[5] See Bondi and Gold (1948), who argued that one cannot assume *a priori* that general relativity applies globally without empirical justification, an objection to the early philosophically motivated cosmological assumptions.

> desirable. It implies closed space sections that would, in some sense, validate Mach's principle according to which the totality of matter in the universe and nothing else determines the local inertial frames" (Rindler 1967: 29–30, emphasis in original). [...] The implicit preference for a closed universe was also expressed in observational studies. "One would particularly like to know whether there is enough mass to close the universe", Princeton physicists Peebles and Partridge wrote in 1967, in a piece on estimating the mass density of the universe (Peebles and Partridge 1967: 713). (de Swart, 2020, pp. 274–275)

These non-empirical commitments mattered as *constraints on theoretical modeling*: they shaped which questions and problems seemed pressing and which theoretical ideas were worth pursuing to resolve them. Weinberg captures this reasoning:

> If one tentatively accepts the result that q_0 is of order unity [$\Omega \gtrsim 1$], then one is forced to the conclusion that a mass density of about 2×10^{-29} g/cm^3 must be found somewhere outside the normal galaxies. But where? (Weinberg, 1972, p. 478)

These commitments were revisable in principle, but taking them seriously made certain problems appear more consequential, as Peebles and Partridge's "*whether*" and Weinberg's "*where*" illustrate. Peebles and Partridge (1967) ask whether there is enough mass to close the universe, while Weinberg (1972) presses the follow-up question of where that mass must be found. By the early 1970s, near-critical cosmological models functioned for some authors as a defensible target and for others as a background commitment with independent motivations. Either way, these commitments helped create a context in which missing mass could be pursued as a candidate astronomical solution to a cosmological problem; a problem which was characterized as such entirely without reference to the arrival of new empirical discrepancies or observations.

Against this background, two research groups independently arrived at similar proposals in 1974. Ostriker et al. (1974) and Einasto et al. (1974) looked to galactic astronomy for a common dynamical account of missing mass on galactic and cluster scales, with implications for the cosmic mass budget. The emphasis of the two papers, however, was not identical: Ostriker et al. (1974) foreground questions about cosmological density and closure considerations from the outset, while Einasto et al. (1974) foreground the cluster virial mass discrepancy and treat massive coronas as a way to reduce it (Einasto et al., 1974, 310). Both groups drew on two kinds of phenomena that were already well known but typically treated as separate research problems in different disciplines. The methods and uncertainties differed, and there was no established view in which the phenomena could be considered naturally connected, let alone bear on the same hypothesis in tandem.

1. *Galaxy dynamics.*
   Radio observations of the 21-cm hydrogen line allowed astronomers to measure rotation speeds of spiral galaxies at radii well beyond the visible stellar disk. By the early 1970s, abundant rotation-curve data were available. Keplerian dynamics suggested that rotation speeds should decrease with radius beyond most of the galaxy's luminous matter (see Figure 1, reconstructed from the original plot in Roberts and Rots (1973)). Empirically, however, many spiral galaxies showed approximately flat rotation curves. This behavior is *consistent* with additional gravitating mass at large radii, but did not by itself establish any single ontological conclusion. In practice, these results were often treated as local anomalies, whose significance depended

on auxiliary assumptions in astronomy about tracer populations, mass-to-light ratios, gas distributions, and dynamical modeling.

2. *Galaxy cluster dynamics.*
The virial mass discrepancy had long been known. Zwicky (1933, 1937) inferred a discrepancy between the mass needed to gravitationally bind the Coma Cluster and the mass estimated from its light. Such inferences depended on equilibrium, virialization and the representativeness of observed velocities. Moving from these dynamical estimates to a cosmological density required further mass accounting based on luminosity and population assumptions, as illustrated by Shapiro (1971). I therefore distinguish the measured cluster velocities from the successive inferences to mass and cosmic density.

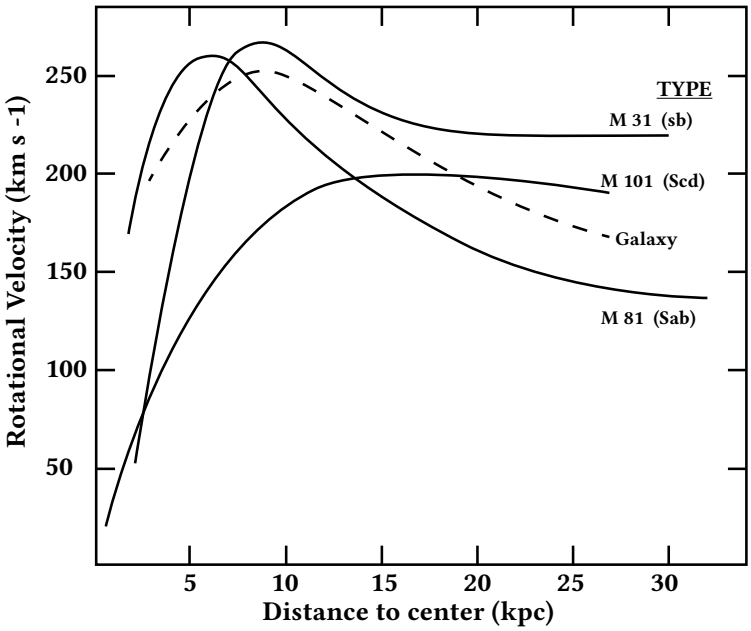


Figure 1: Rotation curves of three spiral galaxies (solid lines) as a function of distance from the center. (The Milky Way's expected Keplerian decline is shown for reference as the dashed line.) The flatness of the curves at large radii indicates the presence of unseen mass.

Ostriker et al. (1974) and Einasto et al. (1974) brought these known observations together under a halo hypothesis. They argued that if each galaxy is embedded in a massive halo extending beyond its visible edge, then:

(i) galaxy rotation curves can remain approximately flat;

(ii) the added mass associated with cluster galaxies can reduce the virial mass discrepancy; and

(iii) the collective mass in galaxy halos could contribute substantially to the cosmic mass budget, potentially bringing Ω closer to unity.

At the start of their paper, Ostriker et al. (1974) write that *"observations may be consistent with a Universe which is 'just closed' ($\Omega = 1$)"*, adding that some held this conclusion for essentially non-experimental reasons (Ostriker et al., 1974, L1). Their mass estimate nonetheless gives $\Omega \approx 0.2$, with an uncertainty of at least a factor of three. Their concluding discussion weighs both low- and high-density arguments, including structure-formation considerations favoring $\Omega \approx 1$ (Ostriker et al., 1974, L4). Einasto et al. (1974) place greater emphasis on reducing the cluster virial mass discrepancy and likewise infer a galactic matter density around 20% of critical density (Einasto et al., 1974, 310). The studies thus connect mass estimates across scales while leaving closure unresolved. A theoretical preference for closure helps explain *why* the mass-budget question was worth pursuing for some researchers, but the case for massive halos still required the astronomical reasoning these papers supplied.

The shared halo framework and its associated parameters made two phenomena previously treated as separate observational anomalies informative about one another. Rotation curves could constrain expectations about cluster dynamics, and cluster dynamics could constrain expectations about rotation curves, through a connection that was not justified when the phenomena were treated in isolation. This changed what could be inferred from the already available observations, marking a transition in 1974 from well-established anomalies to empirical evidence for dark matter. The resulting evidential picture bears directly on the total mass budget of the universe, a question many cosmologists treated as urgent in light of background constraints such as $\Omega \approx 1$.

Two philosophical implications follow from this reconfiguration. First, the shared halo framework allows us to assess how combining rotation curves and cluster dynamics affects confirmation. If the relation between these results is more expected under the halo account than under its alternatives, they can supply comparative support through the informational connection described by Myrvold. The connection and the constraints it imposes therefore matter to the explanation of the 1974 transition.

Second, the same reconfiguration helps explain differences in theory assessment. For researchers concerned with the cosmic mass budget, a cross-scale halo account addressed an especially salient question. Others placed greater weight on unresolved dynamical and observational auxiliaries. The next sections examine how these differences can be represented within a Bayesian assessment, and when the proposed connection contributes comparative support in Myrvold's sense.

4 Permissible community disagreement

Why did the new synthesis persuade some researchers more than others, when they were assessing substantially the same observational results? Their different scientific concerns provide part of the answer: the mass-budget problem gave some cosmologists a reason to pursue connections that researchers working on local dynamics regarded as too dependent on auxiliary assumptions. To assess whether the disagreement was *rationally permissible*, we need to examine the grounds for the background constraints and their role in the inference to missing mass.

4.1 Evidence and background

Rubin's retrospective makes clear that many astronomers initially resisted treating extensive halos as the default conclusion from rotation-curve data:

> In 1977, many astronomers hoped that dark matter might be avoided [...] there were still non-believers. One eminent astronomer said to me,'When you observe low luminosity galaxies, you'll find Keplerian falling rotation curves.' (Rubin, 2004, 6)

She also recalls:

> Kalnajs' (1983) insistence that dark matter is not required, at least for a few galaxies with spatially limited data, convinced a few astronomers that dark matter could be avoided. In retrospect, *we think it is fair to say that many astronomers hoped that Kalnajs was right; dark matter was to be avoided, if at all possible.* (Rubin, 2004, 7) [My emphasis]

Rubin's recollections identify a difference in assessment within an overlapping and changing scientific community. Alongside any sociological explanation, we can ask which background assumptions and modeling commitments made the inference to halos more compelling for some researchers than for others.

For researchers concentrating on galactic dynamics, the salient question was whether the data, together with accepted dynamical and astrophysical auxiliaries, justified a robust inference to a new mass component. Given uncertainties about systematics, tracer populations, and particular dynamical models, it was appropriate to explore whether revised auxiliaries could accommodate the anomalies without pervasive unseen matter. Researchers concerned with global cosmological models also asked what the inferred masses implied for the universe's total density. For some, a near-critical density or closure preference made that question especially pressing and encouraged cross-domain arguments. The rational permissibility of these different assessments depends on how well the relevant constraints and auxiliary judgments can be defended.

4.2 Cosmologists' priors

A Bayesian reconstruction makes the dependence on background assumptions explicit. Let h denote a halo-based dark matter hypothesis (for example, that galaxies are embedded in massive halos), e the relevant empirical results (rotation curves and cluster dynamics), and b the background assumptions fixing modeling commitments and auxiliaries. For each assessment, we specify a probability function and background, giving $\Pr(h \mid e, b)$. Bayes' theorem gives:

$$\begin{aligned} \Pr(h \mid e, b) &= \frac{\Pr(e \mid h, b)\ \Pr(h \mid b)}{\Pr(e \mid b)} \\ &= \frac{\Pr(e \mid h, b)\ \Pr(h \mid b)}{\Pr(h \mid b)\ \Pr(e \mid h, b) + \Pr(\neg h \mid b)\ \Pr(e \mid \neg h, b)} \end{aligned} \tag{1}$$

This representation distinguishes two ways researchers can differ in their assessments while remaining epistemically responsible.

1. *Differences in priors, relative to background assumptions.*
   Communities can assign different $\Pr(h \mid b)$ because their background constraints make different hypotheses antecedently plausible. In the present case, closure-related considerations and associated cosmological principles could raise the prior probability of additional gravitating

mass. These considerations enter through b, shaping which hypotheses are treated as serious candidates within the relevant modeling context.

2. *Differences in likelihoods, due to auxiliaries.*
   Even when researchers agree on the observational data, they may reasonably differ about $\Pr(e \mid h, b)$ and $\Pr(e \mid \neg h, b)$, since these likelihoods depend on auxiliary assumptions about, for example, equilibrium, tracer populations, mass-to-light ratios, or the reliability of particular measurements. In the 1970s, disagreement about such auxiliaries could make it reasonable for some astronomers to treat the support for halos as weaker or more conditional than cosmologists did when working within a mass-budget framework.

For a simplified comparison, suppose that, relative to a broadly shared background b concerning mechanics and measurement practices, both communities agree that h makes e comparatively expected. That is, $\Pr(e \mid h, b)$ is high while $\Pr(e \mid \neg h, b)$ is comparatively low. The evidential force of e for h is captured by the Bayes factor:

$$\mathrm{B} = \frac{\Pr(e \mid h, b)}{\Pr(e \mid \neg h, b)} \tag{2}$$

When B is large, e strongly favors h over its negation. In that setting, if the communities' likelihood assessments are broadly aligned, divergence in $\Pr(h \mid e, b)$ will primarily reflect differences in the prior term $\Pr(h \mid b)$. Historically, closure considerations plausibly affected some prior assessments by making additional gravitating mass antecedently plausible and cross-scale mass accounting worth pursuing (de Swart, 2020). Those considerations operated alongside observational arguments about cosmic density, as the discussion in Ostriker et al. (1974) illustrates. Their relative influence must be understood in the context of each assessment.

How far does this justify permissiveness? Belonging to a successful relativistic framework gives a cosmological commitment a context for assessment, but its justification still requires independent motivation, an empirical basis for integration, and openness to revision, as specified in the overview. A commitment meeting these criteria can reasonably shape a prior assessment. The specific evidence for halos must come from comparing the dynamical results under the halo account and its alternatives. We can now ask how much the proposed unification contributes to this comparative support.

5 Unification

The halo synthesis offered more than a reason to pursue the missing-mass question. It connected the results that would bear on the answer. This is the unificatory contribution attributed to the papers by Ostriker et al. (1974) and Einasto et al. (1974):

> The 1974 papers synthesized the two instances of curious galaxy behaviour into a single framework, thereby coalescing the problems into a single anomaly of missing mass. (de Swart et al., 2017, 6)

The epistemic force attributed to this connection is evident in the description of the studies as "groundbreaking papers, each presenting a strong case for the existence of large amounts of mass in

the outer parts of galaxies" (Bertone & Hooper, 2018, 21).[6] Rubin's retrospective remark identifies what made the connection valuable:

> Science often advances when ideas, formerly very disparate, are united. In retrospect, it took a long time for astronomers to relate Zwicky's dark matter to the flat rotation curves for some galaxies that were beginning to attract attention. (Rubin, 2004, 3)

Tremaine similarly characterized the unification as a watershed in galaxy dynamics and cosmology (Tremaine, 1999, 1223). If the unification of flat rotation curves and the cluster mass discrepancy by dark matter in 1974 was epistemically significant, we need to explain what the connection contributed.

Rubin's remark locates the epistemic value of unification in a connection between known phenomena. An explanationist can appeal to the virtue of explaining those phenomena through a common origin. For a Bayesian, the difficulty is to identify how the connection affects confirmation when the observations are already known and even helped motivate the hypothesis. Myrvold's account of mutual-information unification allows us to examine the probabilistic contribution first, before returning to the historical novelty of the connection.

5.1 Unification, explanation, and the Bayesian challenge

Much of the philosophical literature treats unification as an *explanatory virtue*, linking it to accounts of scientific inference such as inference to the best explanation (IBE). The guiding thought is that if two theories accommodate the same total evidence, but one provides a more integrated explanation, for example by positing a single cause for otherwise disparate phenomena, we are justified in believing that it is true. This explanatory view is defended as epistemically significant by, among others, Kitcher (1989), Lipton (2004), and Psillos (1999, 2009). On this construal, the virtue is non-empirical. It concerns the organization and economy of explanation rather than the introduction of new empirical content.

Fitting this idea into Bayesian confirmation theory requires specifying its probabilistic consequences. In Bayesianism, confirmational relations depend on, and are exhausted by, priors and likelihoods relative to background assumptions.[7] If two hypotheses assign the same joint likelihood to the evidence under the same background assumptions, unification supplies no additional likelihood-based advantage. This is the challenge for a Bayesian account (Hartmann & Sprenger, 2011; Howson & Urbach, 2006).

In the dark matter case, the question is which constraints a common halo explanation imposes across the two domains. By examining how those constraints affect the joint likelihood of the established results, we can assess the unification's comparative contribution to confirmation.

[6] Also, in Rubin, Ford, & Thonnard, 1978, L111: "The major result of this work is the observation that rotation curves of high-luminosity spiral galaxies are flat, at nuclear distances as great as $r = 50kpc$. [...] These results take on added importance in conjunction with the suggestion of Einasto, Kaasik, and Saar (1974), and Ostriker, Peebles, and Yahil (1974) that galaxies contain massive halos extending to large r. Such models imply that the galaxy mass increases significantly with increasing r which in turn requires that rotational velocities remain high for large r. The observations presented here are thus a necessary but not sufficient condition for massive halos."

[7] For treatments of probabilistic reasoning and explanation, see Henderson (2014), Weisberg (2009), Iranzo García (2008), and Douven and Schupbach (2015).

5.2 Mutual-information unification

A hypothesis unifies in Myrvold's sense when it makes distinct evidential claims informative about one another (Myrvold, 2003, 2017). A common-origin explanation can do this if its consequences for one result constrain what we should expect of another. The relevant probabilistic property is the dependence introduced between the results by the hypothesis. Let $\mathbf{e} = \{p_1, p_2, \ldots, p_n\}$ be a body of evidential propositions. Following Myrvold (2003, eq. (7)), we define their informational relevance by

$$I(p_1; p_2 \mid b) = \log \frac{\Pr(p_1 \wedge p_2 \mid b)}{\Pr(p_1 \mid b)\Pr(p_2 \mid b)} \tag{3}$$

The quantity measures informational relevance point-wise, for the particular propositions under assessment. Unlike expected Shannon mutual information, it can be negative: its sign tells us whether one proposition makes the other more probable, leaves its probability unchanged, or makes it less probable. All ratios below require positive probabilities. If $I(p_1; p_2 \mid b) \approx 0$ but $I(p_1, p_2 \mid h, b) > 0$, the hypothesis makes the two results positively relevant to one another. Myrvold captures the unificatory contribution by the following definition:

$$U(p_1, p_2; h \mid b) \;=\; I(p_1; p_2 \mid h \wedge b) \;-\; I(p_1; p_2 \mid b) \tag{4}$$

In the idealized case where $I(p_1; p_2 \mid b) = 0$, U equals $I(p_1; p_2 \mid h \wedge b)$. A positive value identifies an evidential connection absent from the background assessment. To connect this to confirmation, it is useful to work with an information-theoretic measure of evidential support, equivalently a log Bayes factor:

$$C(h; e \mid b) = \log \frac{\Pr(e \mid h, b)}{\Pr(e \mid \neg h, b)} \tag{5}$$

For this log-Bayes-factor measure, Myrvold's decomposition compares the unification provided by h with that provided by its alternative (Myrvold, 2003, eq. (13)):

$$C(h; e_1 \wedge e_2 \mid b) \;=\; C(h; e_1 \mid b) \;+\; C(h; e_2 \mid b) \;+\; U(e_1, e_2; h \mid b) - U(e_1, e_2; \neg h \mid b) \tag{6}$$

Here the difference between the two unification terms determines the additional comparative support. Even a positive U for h supplies no additional unificatory advantage if the alternative generates an equally strong connection. Myrvold's simpler decomposition with just $U(h)$ applies to the different measure based on the log posterior-to-prior ratio (Myrvold, 2003, eq. (12)). For the Bayes-factor comparison used here, we must specify the alternatives and their weights to evaluate $\Pr(e \mid \neg h, b)$. The question is then whether the halo framework's particular constraints make the established combination more expected than the competing accounts do.

There are reasons to look beyond informational dependence when assessing unification's epistemic value. Castellani, Dardashti, and Dawid (2025) identify a further possible source of confirmation in the surprising discovery that a unified account exists at all. Evaluating that contribution requires expectations about theory space and the availability of alternatives, extending the assessment beyond predictions within a theory's intended domain. Applying their argument to 1974

would require contemporary grounds for the relevant expectations and surprise. My concern here is the contribution we can investigate through the halo proposal's cross-domain constraints. Even within that assessment, a likelihood calculation leaves us with the question of whether the connecting auxiliaries are scientifically justified.

5.3 Evidential reconfiguration in 1974

The question now is whether the 1974 synthesis fits this MIU structure. Let h be the halo-based dark matter hypothesis, and let p_1 and p_2 denote the two relevant phenomena:

h: galaxies are embedded in extended halos with substantial unseen mass,
p_1: the galaxy cluster mass discrepancy (virial mass deficit),
p_2: flat galaxy rotation curves at large radii,
b: background assumptions (GR, ordinary matter, measurement conventions),
θ: halo-to-luminous mass normalization parameter.

In a concrete halo model, the parameter space Θ includes profile shapes, scale radii, truncations and population differences. Here θ denotes a schematic halo mass-to-light normalization. The historical estimates indicate the scale and kind of mass accounting involved: Ostriker et al. (1974) estimate a mass-to-light ratio of order 200 in solar units for giant spirals, while Einasto et al. (1974) report values around 100 for spirals and above 120 for ellipticals (Einasto et al., 1974, 310).[8]

For the shared halo framework to make the results informative about one another, conditioning on one result must revise the distribution over halo parameters in a way that *changes expectations about the other*. The evidential connection therefore depends on how this revision affects the model's predictions.[9] The historical contrast begins with the local contexts in which the two phenomena were commonly assessed before the synthesis:

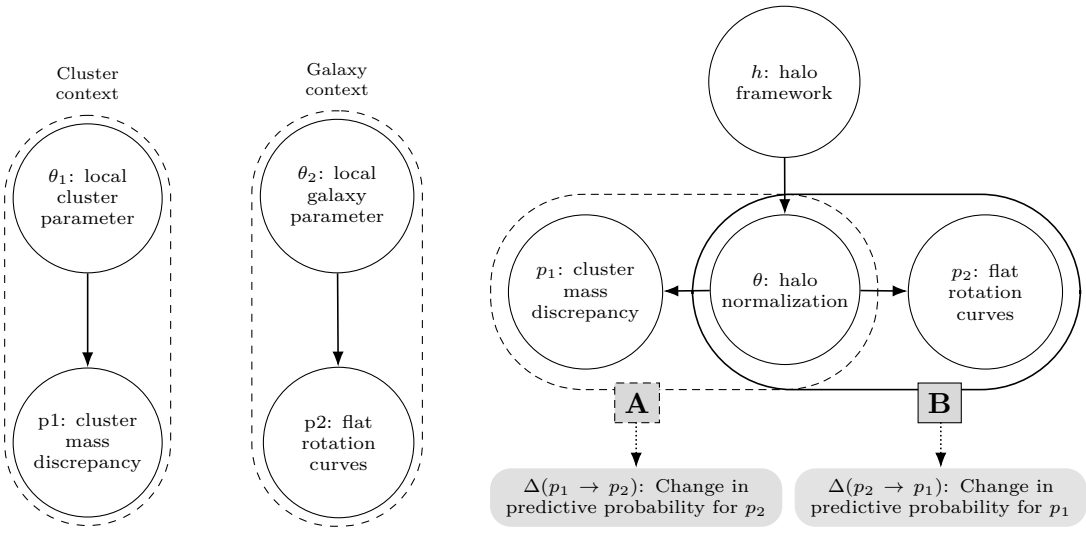
> Clearly, in neither branch of astronomy were these observations at this point cited as positive evidence for the presence of extra matter, or falsifying evidence for any alternative hypothesis. Furthermore [...] the two problems were studied separately. (de Swart et al., 2017, 3)

Figure 2 distinguishes between those local assessments (a) and an integrated halo model (b). In the latter, probabilistic relevance is reciprocal: whenever conditioning on one result changes the probability of the other, the reverse assessment also produces a change. To calculate the contribution of the connection, I introduce an uncoupled control in which the two probabilities factorize. Probabilistic independence is an assumption of this control, requiring more than the historical fact that the investigations were separate. In the shared halo model, both results depend on θ. If conditioning on one result reweights the parameter values so that the other result becomes more probable, the results are positively probabilistically relevant under the halo hypothesis. Relative to a background in which their informational relevance is zero, this gives:

[8] Each estimate belongs to its analysis, with cross-domain implications that depend on the profile, radial range, and population assumptions retained there. Treating these values as estimates of a single universal parameter would conceal those differences.

[9] Reweighting the parameter values can be sufficient, even if all remain in the posterior support.

$$I(p_1; p_2 \mid h \wedge b) > 0 \quad \text{and therefore} \quad U(p_1, p_2; h \mid b) > 0 \tag{7}$$



(a) Separate local assessments. **(b)** Reciprocal inference through a shared halo normalization.

Figure 2: Separate and integrated assessments of the same empirical results. Solid arrows represent modeling dependencies. In (b), A and B show the two paths where conditioning on one result constrains expectations for the other, mediated by tuning the distribution over the shared halo normalization θ. Here $\Delta_{i \to j} = \Pr(p_j \mid p_i, h, b) - \Pr(p_j \mid h, b)$, for $(i, j) = (1, 2)$ or $(2, 1)$. The sign and magnitude of these changes depend on the likelihoods and parameter distribution.

Under a halo framework, flat rotation curves constrain the magnitude and distribution of additional gravitating mass associated with galaxies. Because those same halo parameters determine how much mass galaxies contribute to clusters, the rotation-curve result also alters predictions about cluster dynamics. The inference can run in the other direction as well: cluster mass discrepancies constrain the typical magnitude of galactic halos and thereby predictions about rotation curves. Whether this two-way connection increases support depends on how the observed combination compares with the model's expectations. We can display the mechanism through the posterior predictive distribution:

$$\Pr(p_2 \mid p_1, h, b) \;=\; \int_\Theta \Pr(p_2 \mid \theta, h, b)\, \pi(\theta \mid p_1, h, b)\, d\theta \tag{8}$$

Here π is a proper parameter density, and the equation assumes conditional independence of the two results given θ, h, b. The posterior weights the predictions for p_2 at each parameter value. For p_1 to change the prediction for p_2, it must reweight parameter values at which those predictions differ. The following example shows how parameter learning can produce this predictive change.

5.4 A toy model comparison

Suppose a deliberately simplified halo model M_1 admits two mass-to-light normalizations, θ_L and θ_H, with equal prior probabilities. The events p_1 and p_2 represent the cluster mass discrepancy and

the outer rotation-curve result, respectively. Conditional on the normalization, their probabilities are:

	θ_L	θ_H
$\Pr(p_1 \mid \theta, M_1, b)$	1/5	4/5
$\Pr(p_2 \mid \theta, M_1, b)$	1/5	4/5

The two events are assumed to be independent, conditional on θ, and each event's complement receives the probability left by its table entry. The numbers define a normalized toy model in which a more massive halo population makes both dynamical results more expected under fixed transfer and measurement assumptions.[10] Each result has marginal probability 1/2. Conditioning on p_1 changes $\Pr(\theta_H \mid M_1, b) = 1/2$ to $\Pr(\theta_H \mid p_1, M_1, b) = 4/5$. Consequently,

$$\Pr(p_2 \mid p_1, M_1, b) = \frac{4}{5} \times \frac{4}{5} + \frac{1}{5} \times \frac{1}{5} = \frac{17}{25}, \tag{9}$$

which exceeds its marginal probability 1/2. Conditioning on the first result raises the probability of the second because p_1 favors the higher normalization, where p_2 is also more probable. The joint likelihood is 17/50. For comparison, let M_0 permit independent local normalizations, each with the same prior and response probabilities. It can describe two local mass discrepancies without imposing the cross-domain relation. Both single-result likelihoods remain 1/2, but the joint likelihood is 1/4. The combination therefore gives:

$$B_{M_1:M_0}(p_1, p_2 \mid b) = \frac{17/50}{1/4} = \frac{34}{25} > 1 \tag{10}$$

Although each result has the same marginal likelihood under both models, their combination favors M_1 because of the relation its shared parameter imposes. We can see how much depends on that particular relation by changing the response probabilities for p_2. Setting both to 1/2 gives $B = 1$, since reweighting the normalizations now leaves the expectation for p_2 unchanged. Reversing them gives $B = 16/25 < 1$: the normalization favored by p_1 reduces the probability of p_2. The shared parameter permits all three outcomes.

How far does this comparison take us in assessing the historical episode? M_0 is a specified uncoupled control, whereas the alternatives scientists actually considered involved substantive judgments about visible mass distributions and mass-to-light ratios, tracer and projection assumptions, and whether clusters were bound systems in equilibrium. Those judgments could generate their own relations between results. Comparing these alternatives with a halo account requires specifying their likelihoods and proper parameter priors, as equation (6) makes explicit. Disk-stability arguments also supplied a separate reason to consider halos (Ostriker et al., 1974, L3). These contemporary considerations define the relevant historical comparison. Later modified-gravity proposals illustrate further alternatives to halo accounts, but were unavailable as competitors in 1974.

The toy model establishes how the proposed mechanism works and when it fails. Applying its numerical result to 1974 would require *actual*, justified historical likelihoods and priors, whose estimation is beyond the scope of this paper. The historical argument is that the halo synthesis made a restrictive cross-domain assessment possible, opening the question of how much support its

[10] Note that these probabilities are *stipulated* for the illustration, not estimates from the 1974 papers.

particular constraints supplied relative to defensible alternatives. The calculation reconstructs that contribution through a static model comparison. Conditioning first on p_1 displays its informational relevance to p_2, while both results were already known to the scientists whose reasoning we are examining. The historical change lay in constructing and justifying the connection that made this assessment available. We can now ask what kind of novelty that amounts to in the context of old evidence.

6 Old evidence and structural novelty

The old-evidence problem is usually framed as follows (Glymour, 1980). If an agent assigns probability 1 to an evidential proposition e, then conditionalizing on e cannot change their credences. How, then, can established facts confirm a hypothesis, especially when they helped motivate it? Bayes' theorem makes the difficulty explicit:

$$\Pr(h \mid e, b) = \frac{\Pr(e \mid h, b)\ \Pr(h \mid b)}{\Pr(e \mid b)} \tag{11}$$

If $\Pr(e \mid b) = 1$, coherence also requires $\Pr(e \mid h, b) = 1$ whenever the conditional is defined, so the posterior equals the prior. An agent who already treats e as certain therefore needs something further to learn or revise if their credence in h is to change. In the preceding model comparison, by contrast, b specifies the auxiliaries used to assess the results, leaving their likelihoods open to evaluation under each model. We can therefore calculate the support supplied by known results without yet explaining how scientists came to reassess them.

Eva and Hartmann (2020) examine such reassessment through the alternatives to a hypothesis. Scientists can become more confident that a hypothesis supplies the only available adequate explanation of known evidence, and Jeffrey conditionalization on that comparative explanatory claim can then increase confidence in the hypothesis. Their account makes the role of alternatives especially clear: finding an explanation has limited force if its competitors explain the same results equally well. The halo comparison addresses a related question by identifying how a cross-domain connection can make known results comparatively informative. The calculation leaves open the stronger claim that the halo proposal was the only available adequate explanation. Establishing that claim would require the further assessment of alternatives emphasized by Eva and Hartmann.

The historical question is how the accepted results acquired their joint evidential role. The halo synthesis changed which propositions could bear on the same hypothesis and which relations among them scientists had reason to accept. This is the change in *evidential structure* identified by de Swart et al. (2017):

> [...] representations of the establishment of dark matter in terms of an accumulation of evidence miss an essential part of this history: they overlook the necessary conditions that made this very accumulation possible, an accumulation that, at face value, was substantially, even if partly, a reinterpretation of existing observations. Here, theory, along with an institutional shift and expansion in astronomy, played a substantive role: a role that reminds us that simply asking "what was the evidence for missing matter" misses the point; we need to understand why certain observations were eventually conceived as 'evidence' of anything in the first place. (de Swart et al., 2017, 6)

Counterfactual strategies that remove e from the background can help assess the support it supplies, even though scientists continue to know it. For the historical question raised here, we also need to understand what made the new assessment available. The halo connection provides a candidate answer by changing how the accepted results bear on one another.

6.1 Structural novelty as evidence

What was novel about this connection? Two familiar concepts locate novelty in the history of the evidence relative to a theory. *Temporal novelty* concerns evidence unknown before the theory was formulated, while *use novelty* concerns evidence not used in constructing it. Earman (1992) gives the following confirmation criteria, crediting Worrall (1985, 1989) for the latter:

> **Temporal novelty**: Suppose that $T \vDash E$. If E was already known to be true prior to the articulation of T, then E does not confirm T.
>
> **Use novelty**: Suppose that $T_1 \vDash E.$ and $T_2 \vDash E$. If E was used in constructing T_1, but not in constructing T_2, then T_2 receives more support from E than does T_1. (Earman, 1992, 114)

The rotation-curve and cluster-dynamics results were already available and helped motivate the halo proposal, so neither kind of novelty characterizes them. Earman's use-novelty criterion still allows used evidence to supply support, giving more support to the theory constructed independently of it. Knowing when the results were obtained and how they were used leaves unanswered what the new cross-domain connection contributed.

Possessing a result and assessing its evidential role are different matters. The rotation-curve measurements and cluster velocities required interpretation through auxiliaries that remained open to dispute. In particular, the cluster mass discrepancy follows from the velocities together with dynamical assumptions. Counting it as a further independent measurement would count the same empirical input twice. The propositions p_1, p_2 summarize the interpreted results under the stated auxiliaries. The halo synthesis gave scientists a reason to connect them through a shared parameterization θ, allowing each to constrain the evidential significance of the other. I use *structural novelty* to identify this new connection between established results. The simple model demonstrates the positive-relevance claim required for the proposed explanation:[11]

$$S : \; I(p_1; p_2 \mid h \wedge b) > 0 \quad \text{and} \quad p_1 \perp p_2 \mid \theta, h, b \tag{12}$$

In Equation 12, the shared parameter mediates the connection: the results are independent given θ, yet positively relevant when we average over it. As the controls in section 5.4 show, the latter property depends on the response probabilities and parameter weights. S gives a concise description of this relation within the model. We assess its contribution through the joint likelihood of p_1, p_2 and comparison with the corresponding likelihoods under alternatives. The empirical input remains the established results, with no additional observation of S entering the calculation.

[11] I use $\perp$ for (conditional) probabilistic independence. Formally, $p_1 \perp p_2 \mid X$ iff $\Pr(p_1 \wedge p_2 \mid X) = \Pr(p_1 \mid X)\ \Pr(p_2 \mid X)$, equivalently $\Pr(p_1 \mid p_2 \wedge X) = \Pr(p_1 \mid X)$ (when the relevant conditional probabilities are defined). Informally, conditional on X, learning p_1 does not provide any further information about p_2 (and vice versa).

6.2 When is structural novelty epistemically warranted?

If a parameterization is flexible enough, we may be able to accommodate the observations while learning little from their connection. In the halo case, p_1 might change the parameter weights too little to appreciably alter the expectation for p_2. The framework might also leave profile shapes, radial extents, or population differences unspecified, allowing a fit to p_1 with little consequence for p_2. Even a restrictive halo model must be assessed against defensible alternatives, some of which may produce a comparable dependence through different auxiliaries.

The uncoupled model M_0 defined in section 5.4 provides a control against which we can isolate the contribution of the halo connection. Its joint likelihood factorizes:

$$\Pr(p_1, p_2 \mid b, M_0) = \Pr(p_1 \mid b, M_0)\ \Pr(p_2 \mid b, M_0) \tag{13}$$

Let M_1 be a halo model linking the two domains through a common parameter. Under the stated conditional-independence assumption, its joint likelihood averages over that parameter:

$$\Pr(p_1, p_2 \mid b, M_1) = \int_\Theta \Pr(p_1 \mid \theta, M_1, b)\ \Pr(p_2 \mid \theta, M_1, b)\, \pi(\theta \mid M_1, b)\, d\theta \tag{14}$$

Their comparison is the Bayes factor

$$\mathrm{B}_{M_1:M_0}(p_1, p_2 \mid b) = \frac{\Pr(p_1, p_2 \mid b, M_1)}{\Pr(p_1, p_2 \mid b, M_0)} \tag{15}$$

The finite example in section 5.4 implements this comparison with the observational results and assessment background fixed. Because the models differ in their cross-domain commitments, their joint likelihoods can differ even when their single-result likelihoods agree. As above, b specifies the auxiliaries for evaluating p_1, p_2, rather than a scientist's total belief state in which both are already certain. The calculation assesses the connection's contribution to comparative support. Explaining the historical transition also requires establishing why scientists had reason to construct and accept it.

Dardashti, Hartmann, Thébault, and Winsberg (2019) make the need to justify a mediating connection explicit in their Bayesian analysis of analog experiments. Universality arguments support assumptions linking the empirical adequacy of source and target models, giving scientists grounds to apply a relationship whose mathematical form alone would leave its empirical relevance unsettled. In the dark-matter case, the available astronomical results are connected directly through the halo account, whereas an analog experiment uses a laboratory source to investigate an inaccessible target. In both cases, the inference depends on justified assumptions linking the domains. For the halo synthesis, we must ask why the shared mass accounting was a defensible description of galaxies and clusters.

6.3 MIU and old evidence

We can now locate the contribution to the old-evidence problem more precisely. The halo synthesis made familiar results jointly informative through a cross-domain connection. MIU explains how that connection can contribute to comparative support when it makes the observed combination more expected under the halo account than under its alternatives. This is the probabilistic consequence of the proposed reconfiguration, evaluated once the models and their auxiliaries are specified.

To explain rational credence change between the historical assessments, we would need to identify what scientists learned or revised in accepting the connection and how this affected their beliefs. The present argument supplies one part of that explanation by identifying the new evidential role made available in 1974 and reconstructing the support it could provide. Its answer to the historical puzzle is that the same results became evidence in a new way because scientists could now assess what their relation implied.

7 Conclusion

The 1974 papers by Ostriker et al. (1974) and Einasto et al. (1974) are recognized as pivotal in the evidential consolidation of the dark matter hypothesis, despite relying on observations already available. I have argued that their epistemic contribution can be understood through the new evidential roles those observations acquired within a shared halo framework.

Within this framework, rotation curves and cluster dynamics could constrain one another. The simple model makes the inference explicit: a result changes the assessment of a common halo parameter, which alters expectations for the other result. Comparison with the uncoupled control then shows how the observed combination can favor the connected model, while the modified response probabilities show when it fails to do so. The historical studies motivate this reconstruction. The toy model's numerical values illustrate the mechanism, leaving the strength of the historical support dependent on the relevant likelihoods, priors, and alternatives.

Researchers' scientific concerns and background commitments help explain differences in their reception of the halo connection. Some gave greater weight to the cosmic mass budget, others to unresolved local auxiliaries, with concerns overlapping across astronomy and cosmology. Their assessments could be rationally permissible insofar as the commitments were independently motivated, empirically fruitful, and open to revision. That judgment rests on the reasons available to them at the time.

The observations were familiar, yet their evidential role changed. *Evidential reconfiguration* identifies this historical transformation, and *structural novelty* locates the new connection through which the results became jointly informative. The Bayesian comparison shows how this connection can contribute to confirmation, while a full account of historical belief change would also have to explain its acquisition and acceptance. The 1974 episode shows the work required for observations to function as evidence: scientists needed a defensible relation between the local discrepancies to assess their joint implications for missing mass.

References

Allzén, S. (2021). Scientific realism and empirical confirmation: A puzzle. *Studies in History and Philosophy of Science Part A*, *90*, 153–159.

Allzén, S. (2024). *Dark matter: Explanatory unification and historical continuity.* arXiv. (arXiv:2412.13404) doi: 10.48550/arXiv.2412.13404

Allzén, S. (2026). Dark matter realism: How referential semantics restricts realism in contemporary fundamental physics. *Synthese*, *207*(6), 229. doi: 10.1007/s11229-026-05589-8

Antoniou, A. (2023). Robustness and dark-matter observation. *Philosophy of Science*, *90*(3), 629–647.

Antoniou, A. (2025). Why did the dark matter hypothesis supersede modified gravity in the 1980s? *Studies in History and Philosophy of Science*, *112*, 141–152.

Bertone, G., & Hooper, D. (2018). History of dark matter. *Reviews of Modern Physics*, *90*(4), 045002.

Bondi, H., & Gold, T. (1948). The steady-state theory of the expanding universe. *Monthly Notices of the Royal Astronomical Society*, *108*(3), 252–270.

Bosma, A. (1978). *The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types* (PhD Thesis). Rijksuniversiteit te Groningen.

Castellani, E., Dardashti, R., & Dawid, R. (2025). *Unification and surprise: On the confirmatory reach of unification.* Retrieved from `https://philsci-archive.pitt.edu/26916/` (Preprint)

Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C., & Zaritsky, D. (2006). A direct empirical proof of the existence of dark matter. *The Astrophysical Journal Letters*, *648*(2), L109–L113. doi: 10.1086/508162

Dardashti, R., Hartmann, S., Thébault, K., & Winsberg, E. (2019). Hawking radiation and analogue experiments: A Bayesian analysis. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *67*, 1–11. doi: 10.1016/j.shpsb.2019.04.004

Dawid, R. (2013). *String theory and the scientific method*. Cambridge University Press.

Dawid, R. (2019). The significance of non-empirical confirmation in fundamental physics. In R. Dardashti, R. Dawid, & K. Thébault (Eds.), *Why trust a theory? Epistemology of fundamental physics* (pp. 99–119). Cambridge University Press. doi: 10.1017/9781108671224.009

Dawid, R., Hartmann, S., & Sprenger, J. (2015). The no alternatives argument. *The British Journal for the Philosophy of Science*, *66*(1), 213–234. doi: 10.1093/bjps/axt045

De Baerdemaeker, S. (2021). Method-driven experiments and the search for dark matter. *Philosophy of Science*, *88*(1), 124–144.

De Baerdemaeker, S., & Boyd, N. M. (2020). Jump ship, shift gears, or just keep on chugging: Assessing the responses to tensions between theory and evidence in contemporary cosmology. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *72*, 205–216.

De Baerdemaeker, S., & Dawid, R. (2022). MOND and meta-empirical theory assessment. *Synthese*, *200*(5), 1–28.

de Swart, J. (2020). Closing in on the cosmos: Cosmology's rebirth and the rise of the dark matter problem. In A. S. Blum, R. Lalli, & J. Renn (Eds.), *The renaissance of general relativity in context* (pp. 257–284). Cham: Springer International Publishing. doi: 10.1007/978-3-030-50754-1_8

de Swart, J., Bertone, G., & van Dongen, J. (2017). How dark matter came to matter. *Nature Astronomy*, *1*(3), 0059. doi: 10.1038/s41550-017-0059

Douven, I., & Schupbach, J. N. (2015). Probabilistic alternatives to Bayesianism: The case of explanationism. *Frontiers in Psychology*, *6*, 459.

Duerr, P. M., & Wolf, W. J. (2023). Methodological reflections on the MOND/dark matter debate. *Studies in History and Philosophy of Science*, *101*, 1–23. doi: 10.1016/j.shpsa.2023.07.001

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Einasto, J., Kaasik, A., & Saar, E. (1974). Dynamic evidence on massive coronas of galaxies. *Nature*, *250*(5464), 309–310.

Eva, B., & Hartmann, S. (2020). On the origins of old evidence. *Australasian Journal of Philosophy*, *98*(3), 481–494. doi: 10.1080/00048402.2019.1658210

Glymour, C. (1980). *Theory and evidence*. Princeton University Press.

Hartmann, S., & Sprenger, J. (2011). Bayesian epistemology. In S. Bernecker & D. Pritchard (Eds.), *The Routledge companion to epistemology* (pp. 609–619). Routledge.

Henderson, L. (2014). Bayesianism and inference to the best explanation. *The British Journal for the Philosophy of Science*, *65*(4), 687–715. doi: 10.1093/bjps/axt020

Howson, C., & Urbach, P. (2006). *Scientific reasoning: the Bayesian approach*. Open Court Publishing.

Iranzo García, V. (2008). Bayesianism and inference to the best explanation. *Theoria*, *23*(1), 89–106. doi: 10.1387/theoria.11

Jacquart, M. (2021). ΛCDM and MOND: A debate about models or theory? *Studies in History and Philosophy of Science Part A*, *89*, 226–234.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (pp. 410–505). Minneapolis: University of Minnesota Press.

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). Routledge.

Martens, N. C. M. (2022). Dark matter realism. *Foundations of Physics*, *52*(1), 16.

Martens, N. C. M., Carretero Sahuquillo, M. Á., Scholz, E., Lehmkuhl, D., & Krämer, M. (2022). Integrating dark matter, modified gravity, and the humanities. *Studies in History and Philosophy of Science*, *91*, A1–A5.

Merritt, D. (2021a). Cosmological realism. *Studies in History and Philosophy of Science Part A*, *88*, 193–208.

Merritt, D. (2021b). MOND and methodology. In Z. Parusniková & D. Merritt (Eds.), *Karl Popper's science and philosophy* (pp. 69–96). Springer.

Milne, E. A. (1932). World structure and the expansion of the universe. *Nature*, *130*(3270), 9–10.

Myrvold, W. C. (2003). A Bayesian account of the virtue of unification. *Philosophy of Science*, *70*(2), 399–423.

Myrvold, W. C. (2017). On the evidential import of unification. *Philosophy of Science*, *84*(1), 92–114.

Ostriker, J. P., Peebles, P. J. E., & Yahil, A. (1974). The size and mass of galaxies, and the mass of the universe. *The Astrophysical Journal*, *193*, L1–L4. doi: 10.1086/181617

Peebles, P. J. E., & Partridge, R. B. (1967). Upper limit on the mean mass density due to galaxies. *The Astrophysical Journal*, *148*, 713. doi: 10.1086/149196

Psillos, S. (1999). *Scientific realism: How science tracks truth*. Routledge.

Psillos, S. (2009). *Knowing the structure of nature: Essays on realism and explanation*. Basingstoke: Palgrave Macmillan.

Roberts, M. S., & Rots, A. H. (1973). Comparison of rotation curves of different galaxy types. *Astronomy and Astrophysics*, *26*, 483–485.

Rubin, V. C. (2004). A brief history of dark matter. In M. Livio (Ed.), *The dark universe: Matter, energy and gravity* (pp. 1–13). Cambridge University Press. doi: 10.1017/CBO9780511536298.003

Rubin, V. C., & Ford, W. K., Jr. (1970). Rotation of the Andromeda nebula from a spectroscopic survey of emission regions. *The Astrophysical Journal*, *159*, 379–403. doi: 10.1086/150317

Rubin, V. C., Ford, W. K., Jr., & Thonnard, N. (1978). Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa → Sc. *The Astrophysical Journal*, *225*, L107–L111. doi: 10.1086/182804

Shapiro, S. L. (1971). The density of matter in the form of galaxies. *The Astronomical Journal*, *76*, 291–293.

Sus, A. (2014). Dark matter, the equivalence principle and modified gravity. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *45*, 66–71.

Tremaine, S. (1999). Comments on Ostriker, Peebles, & Yahil (1974), 'the size and mass of galaxies, and the mass of the universe'. *The Astrophysical Journal*, *525*, 1223–1224.

Vanderburgh, W. L. (2003). The dark matter double bind: Astrophysical aspects of the evidential warrant for general relativity. *Philosophy of Science*, *70*(4), 812–832.

Vanderburgh, W. L. (2005). The methodological value of coincidences: Further remarks on dark matter and the astrophysical warrant for general relativity. *Philosophy of Science*, *72*(5), 1324–1335.

Vanderburgh, W. L. (2014a). On the interpretive role of theories of gravity and 'ugly' solutions to the total evidence for dark matter. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *47*, 62–67.

Vanderburgh, W. L. (2014b). Quantitative parsimony, explanatory power and dark matter. *Journal for General Philosophy of Science / Zeitschrift für Allgemeine Wissenschaftstheorie*, *45*(2), 317–327. doi: 10.1007/s10838-014-9261-9

Vaynberg, E. (2024). Realism and the detection of dark matter. *Synthese*, *204*(3), 82.

Weinberg, S. (1972). *Gravitation and cosmology: Principles and applications of the general theory of relativity*. New York: John Wiley & Sons.

Weisberg, J. (2009). Locating IBE in the Bayesian framework. *Synthese*, *167*(1), 125–143.

Wolf, W. J., & Read, J. (2025). *Navigating permanent underdetermination in dark energy and inflationary cosmology.* Retrieved from `https://arxiv.org/abs/2501.13521` (arXiv:2501.13521)

Worrall, J. (1985). Scientific discovery and theory-confirmation. In J. C. Pitt (Ed.), *Change and progress in modern science: Papers related to and arising from the Fourth International Conference on History and Philosophy of Science, Blacksburg, Virginia, November 1982* (pp. 301–331). Dordrecht: D. Reidel. doi: 10.1007/978-94-009-6525-6_11

Worrall, J. (1989). Fresnel, Poisson and the white spot: The role of successful predictions in the acceptance of scientific theories. In D. Gooding, T. Pinch, & S. Schaffer (Eds.), *The uses of experiment: Studies in the natural sciences* (pp. 135–157). Cambridge University Press.

Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, *6*, 110–127. doi: 10.5169/seals-110267

Zwicky, F. (1937). On the masses of nebulae and of clusters of nebulae. *The Astrophysical Journal*, *86*, 217–246. doi: 10.1086/143864