Chapter 8

To Save the CMB

8.1 Extending A Keplerian Defense

At the end of the sixteenth century, astronomers had to reconsider the nature of the solar system in the wake of Copernicus's *De Revolutionibus*. One thread of the ensuing debates concerned how and whether the choice among competing world systems could be resolved. Competing hypotheses arguably "saved the phenomena" – planetary positions – as well as the Copernican system, despite striking differences in how they represented the solar system. The existence of apparently empirically equivalent yet distinct rival systems seemed to vindicate skepticism: how could astronomers determine the true nature of the planets and their motions through the extremely limited means available to them — essentially, observations of angular positions of the planets as a function of time? This is a striking challenge, perhaps too easily dismissed with the benefit of hindsight.

Duhem's *To Save the Phenomena* (1908) recovered a line of thought in this debate that treated astronomical hypotheses as mathematical accommodations of observations, "mere mathematical contrivances," rather than descriptions of physical reality. Duhem himself treated this view with considerable sympathy, taking Kepler's and Galileo's confidence in the physical reality of celestial mechanisms as a departure from its prudent restraint. In *The Aim and Structure of Physical Theory* he acknowledged that successful theories may progress toward what he called a "natural classification" of phenomena, which reflect something real about the world. But Duhem consistently resisted the inference from theoretical success to physical understanding of underlying causes, and his treatment of underdetermination has shaped subsequent philosophical reflection on the limits of what observation can establish. The question this tradition raises is whether saving the phenomena — even saving them very well, over decades of increasingly precise data — is all that scientific success ever amounts to. Contemporary early universe cosmology faces this question in a particularly sharp form: theories of the early universe are evaluated primarily by their compatibility with features of the CMB, prompting the challenge that proposals which save these features successfully may yet fail to identify the physics that actually generated them.

Kepler countered just such a pessimistic position in his neglected early work, *A Defense of Tycho against Ursus* (1600). To preserve his chance of gaining access to Tycho Brahe's data, Kepler had to pen a response to a screed by Tycho's would-be rival Nicholas Ursus, a sustained skeptical critique of astronomy and hypothetical reasoning leading to a position

similar to Duhem's. Two of Kepler's arguments are particularly insightful and salient for our discussion. First, Kepler showed that in several specific cases the alleged conflict between astronomical hypotheses was illusory. In these cases, the hypotheses shared a common structure sufficient to account for observations. This suggests a general strategy: for any set of (allegedly) empirically equivalent hypotheses, seek such a common structure; if it can be found, then the differences can be dismissed as irrelevant to the evidential reasoning at hand, and there is in fact no further choice to be made. Second, Kepler faults astronomers for considering "only the numbers" (that is, planetary positions) as the basis for comparing hypotheses. We should not take success to be so narrowly circumscribed; physical considerations, of various types, should also be taken into account. Kepler faulted his contemporaries for not taking astronomical hypotheses seriously enough, and, like Galileo, demanded that the treatment of the planets should be consistent with a general account of motion and its causes. Kepler's arguments lead to a refinement of what counts as conflicting hypotheses along with a broadening of the kinds of considerations that should be brought to bear in assessing the competitors.

Both arguments require some elaboration to be applied beyond their original context. Kepler invoked the first strategy in cases where common structure was actually found, and where the structure turned out to be physically significant: the apparent rivalry of the hypotheses dissolved, and the way forward was to pursue the implications of the shared structure. Generalized as a procedure for evaluating any set of allegedly empirically equivalent hypotheses, however, the strategy is best understood as diagnostic — a test whose outcome is informative either way. The verdict in Kepler's own case is clear: there is common physical structure, the apparent rivalry dissolves, and the way forward is to pursue the implications of the shared structure. But it can also be the case that there is no common structure of any kind among the rival hypotheses, or the common structure may be merely phenomenological and not tied to anything deeper. As we will argue, both these possibilities are relevant for contemporary cosmology.

Duhem's natural-classification view allowed for theoretical progress, but stopped short of treating such progress as license for inferences about the physical sources of the phenomena — exactly the kind of inference that the Keplerian strategies, carried through to Newton, eventually vindicated. Pursuing Kepler's proposal to develop a physical astronomy successfully would take nearly a century. Kepler was right to insist on the need to understand how astronomical hypotheses related to possible causes of planetary motion, and of course he made a significant improvement in accuracy by introducing elliptical orbits in his *Astronomia Nova* (1609). But his own speculative physics failed to support an illuminating line of further inquiry, because it lacked the specificity to answer further questions about planetary motions. It was not until Newton that the properties of orbital motion could be used to infer quantitative features of the underlying cause, gravity, and to then pursue further implications of this force. As we noted in Chapter 2, Newtonian physics both clarified the physical significance of Keplerian orbits – by identifying the idealized, counterfactual circumstances in which they hold – and showed how departures from them could reveal further facts about the solar system. Pursuing the closing the loop strategy, as we argued above, provides a compelling response to Duhem. Discoveries of new features of the system that can be independently assessed make it hard to dismiss the underlying physics as a mere fiction or mathematical contrivance. Any rival theory, to be taken seriously, should

not only "save the phenomena" equally well, but account for why this reasoning succeeded. This methodology also exposes the underlying theory to ongoing scrutiny, as there is always a risk that a given discrepancy will reveal errors in the proposed physics rather than the need to add another robust feature.

In this chapter, we will make the case that early universe cosmology is at a stage comparable to Kepler's achievement in *Astronomia Nova*. Cosmologists generally agree that one theory of the early universe, inflationary cosmology, "saves the phenomena." Inflation has remained compatible with observations of the CMB through several decades of observational work that have generated an increasingly detailed picture of the early universe. We hold that there are viable alternative hypotheses as well, a more controversial point we will return to below. Successfully saving the phenomena is a significant achievement. Yet throughout the four decades since inflation was introduced, a vocal minority has repeatedly argued that it should be regarded as nothing more than a way of redescribing the phenomena, in effect as a "merely mathematical contrivance." Extending a Keplerian defense of inflation, or an alternative early universe theory, requires posing questions that go beyond compatibility with the observations: to what extent has the theory succeeded in identifying physically significant features of the early universe, established the relevant physics, and then supported the pursuit of further consequences based on it? These theories treat properties of the CMB as the consequences of novel high-energy physics in the very early universe. Inflation, for example, treats the temperature anisotropies in the CMB as the result of dynamical evolution of fluctuation modes of a scalar field ϕ (called the "inflaton") through a phase of exponential expansion. How much do cosmologists gain by doing so, and what are the prospects for a risky and constrained line of inquiry based on accepting this proposal? This is one line of argument we will pursue in this chapter. Although we lack the advantages of hindsight that make a long-term assessment of celestial mechanics possible, we will offer a prospective assessment that clarifies obstacles to closing the loop in early universe cosmology due to the flexibility of physics at the relevant scales, exemplified by the case of inflation.

Furthermore, the Keplerian strategies identify the preconditions for pursuing iterative refinements along the lines we have discussed in earlier chapters. Closing the loop requires identifying the core commitments that have been relevant in evidential reasoning, and this core must also be sufficiently anchored in independently constrained physics to serve as a stable framework for iterative refinements. Without satisfying these two requirements, the relationships between observables and underlying physics are too uncertain and flexible for discrepancies to be used to identify further physical sources, rather than merely providing occasions for parameter adjustment. As we have argued above, Newton's precise characterization of the gravitational force law met these requirements, supporting the iterative inquiry pursued in celestial mechanics in the following centuries. But when there is merely phenomenological common structure shared by competing hypotheses — reflecting generic features that any adequate theory must reproduce, rather than identifying a specific physical source — the first strategy highlights the scope of underdetermination rather than dissolving it. As we will argue, this is unfortunately the situation for inflationary cosmology: the features of the CMB that inflation successfully saves are shared by physically distinct alternatives, and the route to anchoring inflation in independently motivated physics has not been realized.

8.2 The Very Early Universe

Developments in particle physics during the 1970s and 1980s opened up new possibilities for dynamical accounts of the early universe. As the Standard Model of particle physics consolidated and theorists turned to Grand Unified Theories extending it to higher energies, the early universe provided a promising regime where the novel effects of these theories could be explored. Extrapolating field theory to the high temperatures of the early universe implied that the universe had passed through a series of symmetry-breaking phase transitions as it cooled. These phase transitions provided new resources for cosmologists: rather than taking observations to reveal properties of the initial state, one could see features of the early universe as the traces left after a sequence of phase transitions. The emerging connections between particle physics and cosmology thus raised the prospect of replacing the mysteries of initial state discussed in the previous chapter with the outcome of physical processes governed by the same field theories being developed to unify the fundamental interactions.

At the same time, advances in observational cosmology held out the real possibility of empirically evaluating competing proposals. The CMB, aptly called the "cosmic Rosetta stone," provided a rich source of further evidence: its features could be interpreted based on well-established physics, linearized perturbation theory in general relativity and plasma physics to describe the coupled baryon-photon fluid, and the quality and variety of observations improved steadily from the first detection of anisotropies by COBE in 1992 through the precision measurements of WMAP and Planck. Tests of early universe theories initially focused on the angular power spectrum of temperature fluctuations, encoding information about the amplitude, scale-dependence, and statistical properties of primordial perturbations, with subsequent work establishing cosmological parameter constraints based on weak gravitational lensing of the CMB. Cosmologists still aim to extract further insights regarding early universe dynamics, for example based on polarization of CMB photons.

By the late 1990s, these tests aimed to decide between two broad approaches to the origin of primordial perturbations that had emerged as the leading contenders: inflationary cosmology and structure formation via topological defects. Both drew on the new connections between particle physics and cosmology. Inflationary models proposed that the early universe passed through a phase of exponential expansion driven by a scalar field in a false vacuum state, during which quantum fluctuations were stretched to cosmological scales and "frozen in" as classical density perturbations. Topological defect theories, by contrast, proposed that the symmetry-breaking phase transitions themselves produced stable configurations of field energy (cosmic strings, domain walls, monopoles, or textures) that persisted through subsequent evolution and gravitationally seeded the formation of structure.

8.2.1 Contrasting Mechanisms

The sharp contrast between these two accounts of structure formation illustrates how proposals for early universe dynamics can in fact be evaluated observationally. According to inflation, the process of structure formation in the early universe is "passive": the primordial perturbations are imprinted at early times, and thereafter evolve according to linear perturbation theory with no persisting source. The perturbations evolve "on their own" once

the inflationary phase ends. In addition, inflation produces *phase-coherent* perturbations: the dynamics of horizon exit and re-entry leads to synchronization of the Fourier modes, so that fluctuations at all wavelengths are in phase. This coherence produces a characteristic oscillatory pattern in the angular power spectrum of CMB temperature fluctuations—the acoustic peaks that reflect coherent standing-wave oscillations in the baryon-photon fluid prior to recombination.

Structure formation via topological defects, by contrast, is "active": the network of defects persists through cosmic history and continues to interact gravitationally with the other constituents of the universe. In the evolution equation for perturbations, there is a source term representing the stress-energy of the defect network, and determining the evolution of perturbations requires calculating the nonlinear dynamics of the defects themselves. This ongoing, non-linear sourcing leads to *decoherence*: fluctuations at different wavelengths are not in phase, because the source term mixes perturbations across different modes. The resulting angular power spectrum lacks the sharp secondary peaks characteristic of inflation; these features are "washed out" by the decoherence. In addition, defect theories generically predict a primary peak at larger multipole moment ($\ell \geq 300$) than inflation ($\ell \approx 200$), and they produce scalar, vector, and tensor perturbations of roughly equal magnitude.

These contrasting observational signatures made decisive tests through the CMB possible. Observational results beginning in the late 1990s and culminating in the WMAP measurements provided strong support for the inflationary account of structure formation: the observed power spectrum displays the pattern of acoustic peaks at the angular scales expected based on inflation. Structure formation via topological defects was by contrast effectively ruled out, at least as the primary mechanism of structure formation. As is often the case in physics, the sharp contrast reflects a significant simplifying assumption: that only one mechanism makes the dominant contribution to structure formation. Allowing for the possibility of hybrid models combining inflation with a subsequent phase of defect formation muddies the waters considerably, with regard to observational evaluation, although these models still have some appeal.

8.2.2 Contrasting Methodologies

There are also striking methodological contrasts between the two approaches. Despite uncertainty regarding the detailed physics of the phase transitions, the account of structure formation via defects is sufficiently constrained by general theoretical principles to produce specific observational signatures. The observational signatures follow from the topology of the vacuum manifold and general features of defect evolution; they do not depend sensitively on free parameters that can be adjusted to accommodate observations. Physicists working on defects often highlighted this rigidity as a virtue, characterizing their theories as "falsifiable" in a (roughly) Popperian sense. Inflation, by contrast, is implemented through a wide variety of models differing in the nature of the inflaton field and the form of its effective potential $V(\phi)$. They share a common account of the creation and evolution of the initial perturbations, leading to phase-coherent oscillations, but other features of the perturbations depend on these details, which vary considerably over the space of models.

Accounts based on topological defects set aside several of the mysteries of the initial state discussed earlier, focusing on the dynamical production of seed perturbations rather

than directly addressing the flatness and uniformity of the early universe. This contrast, however, did not play a significant role in the detailed observational evaluation of the two approaches. The decisive evidence came from the predicted features of the perturbation spectrum – coherence versus decoherence, the position and structure of acoustic peaks – rather than from considerations of fine-tuning or naturalness of initial conditions.

This observation reflects a broader debate about what qualifies as a successful theory of the early universe. Guth made a case for inflation based on its solutions to the horizon and flatness problems, and these arguments continue to dominate textbook presentations and popular accounts. For many cosmologists, the problems themselves then shifted from mere puzzles or enigmas to a benchmark for evaluating accounts of the early universe. Yet the actual empirical comparison between inflation and its principal rival in the 90s turned on quite different considerations: the physical mechanism for generating perturbations and the distinctive signatures each mechanism imprints on the CMB.

Inflation's victory over topological defects was achieved by a large class of models rather than a single, tightly specified theory. The features that proved to be decisive in the comparison follow from the general mechanism of horizon-crossing amplification and are shared across a wide range of inflationary potentials $V(\phi)$. This is a virtue in one respect: the successful predictions are robust, in the sense that they do not depend on fine details of inflationary models. But in another respect this is a liability: the same robustness means that this empirical success tells us relatively little about the specific physics underlying inflation. The data have selected a broad class of models without pinning down a specific theory within that class, a situation in which the first Keplerian strategy (identifying common structure) succeeds, but at the cost of leaving the second strategy (connecting to underlying physics) with very little to work with.

8.3 Inflationary Cosmology

Inflation's potential for launching a productive line of inquiry in early universe cosmology rested on two main features: the possibility of embedding inflation within a specific model of high energy physics, and its ability to provide an account of the formation of the initial perturbations needed to seed structure formation. Resolution of the fine-tuning problems discussed in the previous chapter, even for those who regard it as a legitimate goal, reveals very little regarding the underlying dynamics — it only places a lower bound on the duration of the inflationary phase. By contrast, the account of structure formation promised to be far more revealing. It became clear with the release of the COBE data in 1992 that observations of the CMB would continue to be a rich source of information regarding these initial perturbations, making it feasible to not only compare inflation to competing accounts of structure formation, but also to discriminate among competing inflationary models — and perhaps even reveal the underlying dynamical details of how inflation works. Cosmologists often wrote of a promising unification of strikingly different physical domains: the new physics introduced in extensions of the Standard Model of particle physics would also provide the key to understanding the earliest moments of the universe's history.

Ironically, pursuing inflation's account of structure formation required abandoning the initial proposals for the physical source of the inflationary phase, making the optimistic vision of unification more remote. The amplitude of the seed perturbations constrains the

self-coupling in a simple scalar field model to be extremely small, ruling out the candidates for the inflaton field drawn from then-current particle physics (such as the Higgs field in grand unified theories, as in Guth's original proposal). In response, cosmologists began to speak of the "inflaton" field — a new fundamental scalar field with properties tailored to drive inflation, postponing, if not setting aside, the question of how it relates to high energy physics.

In terms of the Keplerian strategies introduced above, the subsequent history of inflationary model-building reveals a tension between two desiderata that one would hope to pursue in tandem. Kepler's first strategy directs us to identify the parts of a theory that are actually doing the evidential work, and to recognize that the remaining features are free wheels that are not engaged in the confrontation between theory and observation. Applied to inflation, this strategy isolates the general mechanism of horizon-crossing amplification and the resulting predictions for the perturbation spectrum as the evidential core. These features are shared across the broad class of slow-roll models and do not depend on the specific choice of $V(\phi)$. Kepler's second strategy, however, demands precisely what the first strategy sets aside: a connection between the inflationary hypothesis and independently motivated physics, which would anchor the theory and guide further refinements. The original GUT-based proposals for inflation promised to fulfill both strategies simultaneously, by identifying the inflaton with a field whose properties could be constrained from multiple directions.[1] The failure to realize this promise — and the resulting proliferation of phenomenological models — means that the two strategies now pull in different directions. The evidential core identified by the first strategy is too thin to support the second, whereas the second strategy's demand for a connection to fundamental physics has led to a vast proliferation of speculative models.

8.3.1 Inflation 101

Extrapolations back to approximately 10^{-6} s rely on well-understood physics. The normal matter and radiation that dominates the evolution for most of cosmic history decelerates the expansion, with $\ddot{R} < 0$. But if a type of matter with sufficiently negative pressure dominates the dynamics at earlier times, the expansion accelerates, $\ddot{R} > 0$. Inflation postulates that the universe went through such a transient phase of accelerated expansion at $t \approx 10^{-35}$ s, with $R(t) \propto e^{\xi t}$ ($\xi > 0$), producing enormously more expansion over a given interval than the radiation-dominated rate $R(t) \propto t^{1/2}$.[2] This enormous growth phase must then transition smoothly to the slower expansion of the standard big bang model, through a process known as "reheating" in which the energy stored in the inflaton field is transferred to other particle species. (See, e.g., Chapter 4 of Baumann (2022) for a recent introductory treatment.)

An inflationary stage resolves the fine-tuning problems described above by modifying the causal structure and dynamics of the early universe, effectively setting the stage for the

[1] The identification of the inflaton with the Standard Model Higgs field, non-minimally coupled to gravity, remains a possibility that would partially restore the connection to independently constrained physics. But such models require a large non-minimal coupling whose origin is not well understood, and they have not yet provided the kind of tight, independently motivated constraints on inflationary dynamics that the original GUT-based program envisioned.

[2] The exact expansion rate varies among inflationary models, but nearly all feature a period of nearly exponential growth of the scale factor.

ΛCDM model. Through a sufficient period of inflation (≥ 60 e-foldings), Ω is driven rapidly towards 1, and the entire observed universe can be traced back to a single pre-inflationary region small enough to be causally connected.[3] The uniformity of conditions across widely separated regions of the surface of last scattering then traces not to thermal equilibration — in fact, as **Carroll2014** points out, allowing more time for distant regions to interact in the presence of gravity would likely enhance inhomogeneities rather than erase them — but to the synchronized evolution and decay of the inflaton field throughout the inflated region.

The further consequences of inflation for the formation of structure follow from the behaviour of the Hubble radius compared to comoving length scales. The Hubble parameter $H = \dot{R}/R$ is approximately constant during exponential expansion, so the Hubble radius H^{-1} also remains nearly fixed. But the physical wavelength of a comoving fluctuation mode grows exponentially with the scale factor. A mode that starts at sub-Hubble scales will therefore be stretched past the Hubble radius during inflation. After inflation ends, as the expansion decelerates in the subsequent radiation- and matter-dominated eras, the Hubble radius grows faster than comoving scales, and modes that had previously "exited" the Hubble radius eventually "re-enter" it. This mechanism makes it possible for quantum fluctuations generated through local physics at sub-Hubble scales to be stretched, frozen in as they cross the Hubble radius, and later re-enter as coherent perturbations that seed the phase-coherent acoustic oscillations observed in the CMB — one of inflation's most distinctive observational signatures.

The simplest class of models introduces the inflaton as a scalar field ϕ with an effective potential $V(\phi)$:

$$\mathcal{L} = -\frac{1}{2} g^{ab} \partial_a \phi \partial_b \phi - V(\phi) + \mathcal{L}_I(\phi, A_a, \psi, \ldots), \tag{8.1}$$

where $\mathcal{L}_I$ specifies interactions with other fields.[4] If the scalar field is sufficiently uniform in a spatial region, its stress-energy tensor reduces to $T_{ab} \approx -g_{ab} V(\phi)$, which for $V > 0$ drives exponential expansion, approaching the de Sitter solution $R(t) \propto e^{\xi t}$ with $\xi^2 = \frac{8\pi}{3} V$.[5] Different inflationary models correspond to different choices of $V(\phi)$ and $\mathcal{L}_I$.

Cosmologists have focused in particular on "slow-roll" models, in which the effective potential is constrained to be extremely flat, so that ϕ rolls gradually toward its true vacuum state, ensuring the prolonged inflationary phase needed to resolve the horizon problem. The features of the primordial perturbation spectrum depend on the shape of $V(\phi)$ when the relevant fluctuation modes exited the Hubble radius. A flat potential yields a nearly scale-invariant spectrum, compatible with observations, with a spectral index slightly less than 1 ("red tilt") reflecting the fact that the Hubble parameter is not exactly constant throughout inflation. The inflationary mechanism also generates tensor perturbations — primordial gravitational waves — and predicts a consistency relation linking the tensor-to-scalar amplitude ratio r to the tensor spectral index n_t. As we discuss below, this consistency

[3] Check: Confirm this claim against criticisms by Trodden and Vachaspati regarding whether a single Hubble patch suffices, or whether a somewhat larger causally connected pre-inflationary region is required.

[4] The space of inflationary models includes many further options, such as models with multiple scalar fields, non-minimal coupling between the scalar field and the Ricci scalar, or non-standard kinetic terms.

[5] Cosmic "no-hair" theorems provide a more precise characterization of the approach to a de Sitter solution, although gaps remain between what can be established rigorously and what is usually taken to follow from heuristic arguments about the rapid dilution of pre-existing matter and radiation.

relation is among the most promising targets for establishing the physical significance of the inflationary mechanism.[6]

Finally, the transition back to the standard cosmological model requires "reheating": all pre-existing matter is diluted rapidly during inflation, leaving just the inflaton field alone. The interaction terms $\mathcal{L}_I$ and the shape of $V(\phi)$ near its minimum must be chosen so that ϕ decays into the appropriate particle species, repopulating the universe with all of the appropriate constituents needed for the ΛCDM model. Reheating is essential for the consistency of inflation as a whole, but observational constraints on its details remain substantially weaker than those bearing on the generation of primordial perturbations (see, e.g., Martin, Ringeval, and Vennin, 2015). This asymmetry matters for the assessment of inflation, because reheating constitutes one of the two dynamical transitions — along with the onset of inflation — that lie beyond the reach of inflation's best-developed theoretical framework, as we discuss below.

8.3.2 Physical Significance of Features of the Early Universe

If inflation occurred, observations of the CMB acquire physical significance beyond their role in constraining the ΛCDM model: they reflect specific features of inflationary dynamics, and can in principle be used to probe physics at energy scales far beyond those accessible in terrestrial experiments. The relevant observational constraints bear on two distinct dynamical regimes of the inflaton's evolution. The first concerns the amplification of quantum fluctuations at the time of horizon crossing, approximately 60 e-folds before the end of inflation. Features of the scalar and tensor perturbation spectra — the amplitude, spectral index, and running — can be related, in slow-roll models, to $V(\phi)$ and its derivatives evaluated at the time the relevant modes crossed the Hubble radius.[7] Detection of CMB B-mode polarization would directly measure the tensor-to-scalar ratio r, and, in combination with a measurement of n_t, would test the consistency relation and provide a direct constraint on the energy scale of inflation. The second regime concerns the end of inflation and reheating, which depends on $V(\phi)$ near its minimum together with the interaction terms $\mathcal{L}_I$. Observational constraints on reheating are much weaker than those derived from the perturbations, and more heavily model-dependent.

The most revealing possibility is the direct reconstruction of the effective potential from observations. As noted above, a scale-invariant spectrum of scalar perturbations was proposed well before inflation, but there is no similar theory-independent reason to expect a scale-invariant tensor spectrum, or for the scalar and tensor spectra to satisfy the consistency relation predicted by inflation. Measurements of the tensor perturbation spectrum

[6] [CHECK: Some inflationary models apparently violate the standard single-field consistency relation. Clarify which model classes are involved and what this implies for the argument.]

[7] The scalar perturbation spectrum is conventionally parametrized as $P_s(k) = A_s(k/k_*)^{n_s-1+\frac{1}{2}\alpha_s \ln(k/k_*)}$, where A_s is the amplitude at a pivot scale k_*, n_s is the scalar spectral index (with $n_s = 1$ corresponding to exact scale invariance), and $\alpha_s \equiv dn_s/d\ln k$ is its "running." The tensor spectrum $P_t(k)$ is parametrized analogously, with amplitude A_t and tensor spectral index n_t ($n_t = 0$ corresponding to scale invariance, by convention); the tensor-to-scalar ratio is $r \equiv A_t/A_s$. In single-field slow-roll models these quantities are determined at leading order by the slow-roll parameters $\epsilon \equiv (M_P^2/2)(V'/V)^2$ and $\eta \equiv M_P^2 V''/V$, evaluated when the relevant modes crossed the Hubble radius: $n_s - 1 \approx 2\eta - 6\epsilon$, $r \approx 16\epsilon$, and $n_t \approx -2\epsilon$ (so $r \approx -8n_t$, the consistency relation), with α_s depending on higher derivatives of $V(\phi)$. See, e.g., Chapter 2 of Baumann (2022) for more.

at different length scales could, in principle, be used to reconstruct $V(\phi)$ directly. This prospect is one of inflation's most appealing features, because it would provide precisely the kind of independent, converging constraints needed to make a more convincing case that an inflationary stage occurred.

To see what is at stake, it is useful to draw a contrast with other cases in the history of physics where observations acquired increasing physical significance in the course of an extended line of inquiry. The case of the molecular hypothesis between roughly 1900 and 1913 provides an illuminating comparison. Prior to Perrin's work, discussed earlier in Chapter 4, attempts to constrain physical properties of atoms and molecules relied on a variety of different theoretical models, each involving speculative and highly model-dependent assumptions. Although several of these approaches yielded estimates of Avogadro's number, the derivations depended so heavily on specific modeling choices that the resulting agreement carried limited evidential weight. What changed with Perrin was not merely the addition of new measurements, but a qualitative shift in the character of the evidential reasoning: his determinations of N drew on diverse phenomena — Brownian motion, sedimentation equilibrium, the blue of the sky, radioactive decay — through theoretical routes whose linking assumptions were far less speculative than those of earlier approaches, and which constrained the same quantity through substantially independent lines of reasoning. The agreement among the resulting determinations was not something that could easily be reproduced by competing hypotheses.

We can briefly note two features of Perrin's case, discussed further in §8.4.2, where the structure of its evidential reasoning and the contrast with inflation are drawn out in detail. First, the theoretical relationships connecting observations to N were *stable*: they did not depend sensitively on assumptions that might shift as the underlying physics developed. Second, the framework was *specific*: it yielded determinate, quantitatively precise predictions for N that could serve as premises for further inquiry.

The question for inflation, then, is whether its account of structure formation has attained — or can attain — a comparable status. Are the theoretical relationships between inflationary dynamics and observable features of the CMB sufficiently stable and specific to support the kind of iterative, self-correcting inquiry that characterizes a mature physical theory? Current observations, principally from Planck, have established that the scalar perturbation spectrum is very nearly scale-invariant with a slight red tilt ($n_s \approx 0.965$), Gaussian to high precision, and predominantly adiabatic.[8] These results are compatible with the simplest slow-roll models. No primordial tensor perturbations have been detected; current upper bounds place $r < 0.036$ at 95% confidence (**bicepkeck2021**). The power spectrum is consistent with vanishing non-Gaussianity, meaning that higher-order corrections to the simplest inflationary picture remain unconstrained.

These observations provide significant constraints on inflationary phenomenology, but their bearing on the underlying physics is less clear-cut. Compatibility with slow-roll inflation is also compatibility with a very large class of specific models. The free function $V(\phi)$, even when constrained by current data, remains dramatically underdetermined: several distinct classes of models remain consistent with the Planck dataset (**martin2024encyclopaedia**). The observations constrain the shape of $V(\phi)$ over a limited range corresponding to the ≈ 8

[8] See **planck2018x** for the most recent Planck constraints on inflationary models.

e-folds during which observable modes crossed the Hubble radius, but tell us very little about the potential during other phases of evolution. In the absence of observations of primordial tensor modes, inflation's most distinctive signature, the consistency relation, remains untested.

8.3.3 The Persistence of Fine-Tuning

It will come as no surprise that fine-tuning worries of the sort that motivated inflation resurface in a new guise. The first type concerns the initial conditions required for inflation to begin. An inflationary phase requires the inflaton field to be in a suitable state — sufficiently uniform over a region slightly larger than the Hubble radius — and it is unclear whether such a state is "generic" or itself requires explanation. General thermodynamic considerations, along the lines discussed in Chapter 7, suggest that the pre-inflationary state should have lower entropy than the post-inflationary state, since inflation is an irreversible dynamical process.[9] That is, inflation resolves certain fine-tuning problems only at the cost of introducing a different set of special initial conditions whose plausibility depends on unknown pre-inflationary physics. To be clear, this criticism targets one rationale for inflation rather than the viability of inflation in its own right. Inflation certainly could occur as a distinct dynamical phase of early universe evolution, regardless of whether some advocates of the proposal have correctly characterized its impact or what is required for a viable model.

The second type of fine-tuning concerns the inflaton Lagrangian itself. Slow-roll models require an extremely flat effective potential, which is difficult to maintain in the face of radiative corrections. This is the "η problem": successful slow-roll inflation requires the dimensionless parameter $\eta = m_\phi^2/3H^2 \ll 1$, but generic quantum corrections to a scalar field potential yield $\Delta\eta \sim \mathcal{O}(1)$, spoiling the slow-roll condition. More generally, the observationally preferred plateau-type potentials appear unnatural within the framework of effective field theory, a point to which we return below.

Both types of fine-tuning bear on the broader question of how inflation relates to fundamental physics. The initial-conditions problem raises questions about the onset of inflation that cannot be answered within the inflationary framework itself. The η problem (along with other aspects of inflation) signals a sensitivity of inflationary model-building to physics at energy scales far above those directly constrained by observations. Taken together, they suggest that inflation, despite its phenomenological successes, does not yet rest on a foundation that is either physically well-motivated or self-contained enough to support the kind of sustained empirical inquiry that would allow cosmologists to close the loop.

8.3.4 The Modeling Space

Inflationary model-building has undergone a significant transformation since the earliest proposals. Guth's original model was embedded within a grand unified theory, and the earlier **Starobinsky** model arose from a quantum-gravitational modification of the Einstein-

[9] Penrose has pressed this point; see Chapter 3 of **Penrose2016** for a recent discussion. Assessments of the initial conditions required for inflation have been a recurring research theme, but the assessment is model-dependent and no clear consensus has emerged. See also, e.g., Gibbons and Turok 2006; Carroll and Chen 2004; Ijjas, Steinhardt, and Loeb 2013; Schiffrin and Wald 2012.

Hilbert action. Both proposals derived their appeal partly from connections with independently motivated physics: the physical source of an inflationary stage was not introduced *ad hoc* but emerged from theoretical frameworks with independent physical motivations. Neither model (as originally understood) has been vindicated, but the initial program of inflationary model-building was guided by the aspiration to identify the inflaton with a field already present in, or strongly motivated by, high energy physics.[10]

This aspiration has not been realized. Instead, the inflaton has become a phenomenological placeholder: a scalar field whose properties are chosen to fit cosmological data, with the question of its microphysical identity largely set aside. The result has been an extraordinary proliferation of models. **martin2024encyclopaedia** catalogue a few hundred distinct inflationary models, and analyze the classes of models that remain compatible with current observations. What has been achieved is not the hoped-for substantive unification of early universe cosmology with particle physics, but rather a *methodological* unification in which early universe theorists employ the tools and techniques of quantum field theory to perform calculations and analyze the implications of early universe observations — without, however, the independent physical constraints that originally motivated the enterprise.

The EFT Approach

The most systematic attempt to manage this proliferation treats inflation as an effective field theory. If inflation acts as a dynamical attractor — so that diverse pre-inflationary states converge on the same post-inflationary outcome — and if consistent inflationary models can be constructed in quantum field theory through minimal extensions of the Standard Model that are insensitive to higher-energy physics, then inflation can be treated as an EFT organized by the symmetries of the quasi-de Sitter background.

The influential framework of Cheung et al. (2008) realizes this program by working in the unitary gauge, where scalar field fluctuations are absorbed into the metric. The background spacetime breaks time-translation invariance (inflation must, after all, begin and end), and the most general EFT compatible with this symmetry breaking can be organized in terms of operators involving the metric perturbation δg^{00} and the extrinsic curvature $K_{\mu\nu}$. At zeroth order, the action takes the form

$$S_0 = \int d^4x\sqrt{-g}\left[\frac{M_P^2}{2}R - c(t)\, g^{00} - \Lambda(t)\right], \tag{8.2}$$

where $c(t) = -M_P^2\dot{H}$ and $\Lambda(t) = M_P^2(3H^2 + \dot{H})$ are fixed by the background expansion history. All single-field slow-roll models can be cast in this form. Higher-order terms in the EFT expansion are connected to observable departures from the simplest predictions: non-Gaussianities in the CMB, characterized by the bispectrum $B_\zeta(k_1, k_2, k_3)$, serve as the principal observable linking the CMB to the space of EFT operators. Different shapes of the bispectrum implicate specific higher-order terms, offering a direct connection, in principle, between observed non-Gaussianities and new physics beyond the zeroth-order inflationary picture.

[10] Add: Starobinsky's model is still a viable model, although the original physical motivation is no longer accepted.

The EFT framework provides a model-neutral common language for inflation, and it clarifies what the symmetries of an inflationary spacetime imply for the structure of the theory. But it also makes the limitations of the inflationary program more transparent, in the following respects.

The EFT describes only the *middle* of inflation: small fluctuations around a quasi-de Sitter background whose existence is assumed from the outset. The onset of inflation and the transition to reheating both involve dynamics that break the symmetries of the background, and they fall outside the EFT's scope. This is significant because these transitions are precisely the aspects of inflation that connect it to the broader questions of initial conditions and the matter content of the post-inflationary universe.

The standard response to the EFT's silence about the onset of inflation appeals to cosmic no-hair conjectures, which suggest that de Sitter spacetime is a dynamical attractor for rapidly expanding spacetimes (**Wald1983**; **Kaloper2019**). If correct, these conjectures justify the quasi-de Sitter starting assumption regardless of pre-inflationary details: whatever the initial state, the dynamics drive the spacetime towards the attractor before structure-forming modes exit the Hubble radius. The EFT does not itself describe the onset of inflation, and it also does not describe the exit from inflation: the reheating process. This distinguishes an inflationary EFT from familiar limitations of EFT reasoning. Standard EFTs are agnostic about UV physics by design and their cutoffs delimit the energy scales above which the framework is not expected to apply. But effective field theories are normally expected to be representationally adequate at and below the cutoff scale. The inflationary EFT fails this expectation in the deep IR: the end of inflation and the transition to reheating occur at energies well within its expected domain, yet the framework lacks the resources to describe them. The framework's limits at both endpoints are not a matter of computational convenience but a structural feature: the EFT depends on a quasi-de Sitter background whose existence it cannot itself account for, and which it also cannot relinquish.

Second, several features of inflationary spacetimes strain the assumptions of the EFT framework itself. Standard EFT constructions rely on a stable separation between high- and low-energy modes. In a rapidly expanding spacetime, this separation breaks down: physical modes are continually stretched to lower energies, and modes that seed structure formation may have originated at super-Planckian energies — the trans-Planckian problem (Martin and Brandenberger, 2001). The standard response holds that predictions are insensitive to trans-Planckian physics provided that modes cross the EFT cutoff adiabatically, in the Bunch-Davies vacuum state. But justifying the Bunch-Davies condition from within the EFT remains controversial, and the problem highlights that inflation's predictions can in principle depend on unknown physics at the highest energies.

Finally, the empirically favoured classes of inflationary EFTs suffer from naturalness problems. The η problem, mentioned above, is a manifestation of a broader difficulty: the flat effective potentials required by slow-roll inflation are generically destabilized by radiative corrections, and protecting them requires either fine-tuning or the imposition of symmetries whose ultraviolet origin is unclear. These naturalness concerns suggest that inflation couples sensitively to high-energy physics in ways that undermine the central promise of the EFT approach — namely, systematic insensitivity to unknown ultraviolet details.

These limitations come into sharper focus when viewed in terms of the Keplerian strategies introduced above. The EFT framework can be understood as a formalization of the

first Keplerian strategy, applied within the space of inflationary models. It establishes the common structure, shared across the hundreds of specific models that have been proposed, that actually does the evidential work. The answer is the zeroth-order action S_0, fixed by the symmetries of the quasi-de Sitter background and by the background expansion history $H(t)$. All single-field slow-roll models reduce to this form; the model-specific details are absorbed into higher-order corrections. The EFT strips away what is idiosyncratic to each model and isolates what is shared.

But what is the status of the common structure the EFT identifies? In the case that vindicated Kepler's program, the common structure Newton identified turned out to be physically significant, and provided the basis for further iterative refinement. By contrast, the zeroth-order inflationary EFT is organized by the symmetries of the background spacetime, not by any particular physical mechanism or independently constrained field content. The background quantities $c(t)$ and $\Lambda(t)$ are fixed by $H(t)$ and $\dot{H}$; they encode *what the expansion history was*, not *why it occurred* or what physical degrees of freedom were responsible for it. The common core is, in this sense, phenomenological: it captures the kinematic framework within which all single-field models operate, without constraining the dynamical content that distinguishes them.

This would not, on its own, be a damaging observation. Effective field theories are by their nature agnostic about the ultraviolet completion; that is their methodological rationale. In well-grounded applications of EFT reasoning, this agnosticism is a virtue rather than a limitation, because the leading-order predictions are *robust*: they are insensitive to the details of the UV physics, depending only on symmetries and a small number of low-energy constants. Chiral perturbation theory provides the paradigmatic example. There, the low-energy interactions among pions are dictated by the pattern of chiral symmetry breaking in QCD. The UV theory constrains the EFT — it determines which symmetries are broken, by what mechanism, and in what pattern — but the low-energy predictions do not depend sensitively on the details of the strong dynamics at higher energies. This robustness is what makes the EFT a genuine bridge between energy scales: low-energy measurements can be used to determine EFT coefficients with the confidence that those coefficients reflect stable physical relationships. And the anchoring works in both directions: lattice QCD calculations provide independent constraints on the same coefficients, enabling precisely the kind of mutual checks that characterize overdetermination.

This characterization of the unusual status of inflationary EFTs can be sharpened. The inflationary EFT is not merely a phenomenological framework, where the lack of an underlying physical source would be unobjectionable, since no richer commitment was being claimed. Nor does it deliver what EFTs in particle physics provide: a low-energy framework whose structure is dictated by the symmetries of a known (or at least well-constrained) ultraviolet theory, and whose predictions are correspondingly insensitive to UV details. As Koberinski and Smeenk argue at greater length elsewhere (**KoberinskiSmeenk2024**), the inflationary EFT occupies an uncomfortable middle position. It carries more theoretical commitments than a purely phenomenological parameterization warrants — such as a stable separation of scales, and systematic insensitivity to UV physics — yet lacks structural features shared by paradigmatic EFTs in particle physics. The framework inherits the burdens of being treated as physically principled without delivering the corresponding payoffs.

The η problem (and, more generally, the sensitivity to assumptions about UV physics)

reveals that the inflationary EFT does not share the robustness of the EFTs used in particle physics. The near scale-invariance of the primordial perturbation spectrum, inflation's signature observational success, requires an extremely flat effective potential, with $\eta \ll 1$. But generic radiative corrections from physics above the EFT cutoff yield $\Delta\eta \sim \mathcal{O}(1)$, which would spoil slow-roll entirely. The flatness that the data confirm is therefore not a natural low-energy consequence of whatever the UV physics turns out to be. It is a feature that the UV physics must be specifically arranged to produce, either through fine-tuning or through the imposition of a protective symmetry (such as an approximate shift symmetry for the inflaton) whose ultraviolet origin remains unclear. The common core identified by the first Keplerian strategy is thus not merely thin but *fragile*: its leading-order predictions are unstable under the kinds of corrections that unknown UV physics would generically introduce. Rather than providing a secure low-energy foundation on which further inquiry can be built, the zeroth-order EFT delivers predictions whose very success calls for explanation from the UV theory, an explanation that is not currently available.

A further manifestation of this UV sensitivity is the cosmological constant problem. The cosmological constant problem is the most (in)famous anomaly for EFT reasoning for systems involving gravity coupled to scalar fields. Standard EFT reasoning applied to the vacuum energy gives a contribution of order $M_{\rm UV}^4$, which exceeds the observed value of Λ by some 10^{60} to 10^{120} orders of magnitude depending on the choice of a UV cutoff. The inflaton's effective potential $V(\phi)$ acts during inflation as a time-dependent vacuum energy, of precisely the kind that standard EFT treatment fails to account for. The η problem and the cosmological constant problem both reflect the fact that EFT for scalar fields coupled to gravity does not behave as paradigmatic EFTs in particle physics do, where the relevant operators are protected from large UV contributions by symmetries. The viability of inflationary EFTs is hence sensitive to UV physics. Optimists take this as an indication that inflation may provide further clues about high energy physics; whether one shares this view, there is clearly more foundational work required to clarify the status of inflationary models.

The sensitivity to UV physics poses an obstacle to the second Keplerian strategy. The EFT framework was designed, in part, to provide a route from CMB observations to constraints on underlying physics by organizing inflationary predictions in a systematic expansion. Higher-order terms in the EFT, connected to non-Gaussianities in the CMB, would in principle probe the physics beyond the zeroth-order picture, progressively revealing the dynamical content that the leading-order description leaves out. This is precisely the kind of iterative, increasingly constraining program in line with the second strategy. But the program faces two challenges. Empirically, non-Gaussianities have not been detected: current data are consistent with $f_{NL} = 0$ at $\mathcal{O}(1)$ precision, leaving the higher-order EFT coefficients, which would discriminate among models and connect to new physics, entirely unconstrained.[11] Theoretically, even if non-Gaussianities were detected, the leading-order framework through which they would be interpreted suffers from the UV sensitivity just described: the "linking assumptions" connecting EFT coefficients to the underlying dynamics are themselves uncertain. Discrepancies between observation and the zeroth-order EFT

[11] See **planck2018ix** for the most recent Planck constraints on primordial non-Gaussianity.

could not be unambiguously attributed to specific higher-order physical effects, because they might instead reflect the instability of the leading-order description itself.

There is an irony worth noting here. The higher-order EFT terms connected to non-Gaussianities are in certain respects *less* vulnerable to these concerns than the zeroth-order terms, because they are tied directly to the symmetry-breaking pattern of the quasi-de Sitter background rather than to the flatness of the potential. A detection of non-Gaussianities would provide genuine constraints on the EFT at an order where the framework is on firmer ground. The current situation is thus one in which the EFT is most trustworthy at the orders where the data provide no constraints, and least trustworthy at the order where the data are most informative. This further underscores the distance between the inflationary EFT and the kind of well-grounded effective theory that can serve as a stable platform for iterative refinement.

In sum, the EFT approach to inflation formalizes one application of the first Keplerian strategy — the search for common structure across models within a research program — and the diagnostic verdict is the third of those flagged in the introduction. Common structure is identified: the zeroth-order action S_0 is shared across all single-field slow-roll models, and the EFT framework strips away what is idiosyncratic to specific potentials to isolate what is shared. But the identified structure is, arguably, phenomenological rather than physically significant: it captures the kinematic framework of single-field inflation without anchoring it in independently constrained physics. And it is fragile rather than robust, in the specific sense that its leading-order predictions depend on UV arrangements that the EFT framework itself cannot justify. The strategy thus succeeds at its narrow task while showing that the substantive dissolution of underdetermination it was meant to support is not available: the thin, UV-sensitive common core does not anchor inflation, and provides insufficient ground for pursuing the second strategy and further iterative refinements.

8.4 Closing the Loop?

Inflation has survived four decades of increasingly precise and informative observations. It was not ruled out by the COBE measurements of 1992, nor by the first detection of acoustic peaks in the late 1990s, nor by the precision data from WMAP and Planck. The observed CMB is compatible with the generic predictions of the simplest inflationary models: a nearly scale-invariant, Gaussian, adiabatic spectrum of primordial scalar perturbations. But compatibility with observations, sustained over time, is not by itself sufficient to establish that a theory has achieved the kind of empirical success that warrants treating it as a secure foundation for further inquiry. The question is whether inflation's track record constitutes the first steps in a progressive program of iterative refinement — a process in which observations constrain the theory with increasing precision, discrepancies prompt the identification of new physical effects, and the resulting corrections lead to still closer agreement with observation — or whether it reflects something more modest: the survival of a flexible framework that has not yet been subjected to the kind of demanding tests that would establish that a stage of inflationary expansion actually occurred.

8.4.1 From Compatibility to Iterative Refinement

What would it mean for inflation to achieve the kind of empirical success that warrants treating it as a secure foundation for further inquiry? On the account developed above, empirical success is not merely a matter of predictive accuracy or compatibility with observations. A theory achieves genuine empirical success when it supports the *overdetermination* of its theoretical parameters — when multiple, independent bodies of data can be brought to bear on the same quantities, providing mutual checks — and when it can serve as a starting point for a process of iterative refinement, in which systematic discrepancies between calculation and observation are traced to physical sources that lead to progressively more detailed and accurate models.

This is the standard developed in Smith's account of the methodology of celestial mechanics. Newton did not merely show that an inverse-square force law was compatible with the observed planetary motions. He showed that diverse, independent lines of evidence (the motions of the planets, their satellites, comets, the tides, the precession of the equinoxes, and so on) could be accounted for based on the same force law. The subsequent development of celestial mechanics showed that systematic discrepancies between the idealized theory and observation could be traced to specifiable physical sources, leading to refined models that achieved still closer agreement with the data. Newton's theory supported a progressive research program in which each round of increasingly precise observations led to the identification of new physical effects and the incorporation of further details. Applying this approach to early universe cosmology, and paraphrasing Smith's treatment of celestial mechanics: the challenge is not just whether observed features of the early universe, including the overall uniformity and temperature anisotropies in the CMB, can be fit with a particular choice of the "inflaton" field, but whether robust physical sources can be found for each systematic discrepancy between calculations and observation — with the further demand of achieving closer agreement with observation in a series of successive approximations in which more and more details of the early universe that make a difference become identified, along with the differences they make.

Meeting this standard does not require a comprehensive theory of physics at the energy scales relevant to the early universe, nor does it require that all foundational problems concerning inflation be resolved before the theory can be put to productive use. In the case of celestial mechanics, after all, Newton was able to use the inverse-square law as the basis for an extraordinarily productive line of inquiry while remaining agnostic about the physical mechanism underlying gravitational attraction. What *is* required, however, is that the theoretical framework be sufficiently specific and stable to support the iterative process just described: observations must be able to constrain the theory tightly enough that discrepancies are informative, and the framework must be rigid enough that the response to a discrepancy is to identify a new physical effect rather than to adjust a free parameter.

8.4.2 The Flexibility of Inflationary Modeling

Overdetermination?

The target of observational constraints on inflation is the inflaton Lagrangian — in particular, the effective potential $V(\phi)$ and the interaction terms $\mathcal{L}_I$. (This is for the simplest

single field models, more general models include additional parameters that would need to be constrained.) One seeks multiple, independent ways of constraining the same features of this Lagrangian. Ideally, different classes of data would bear on the same theoretical quantities through distinct chains of reasoning, so that their agreement would be surprising if the underlying theory were wrong.

In practice, the two principal observational windows on inflation — the primordial perturbation spectrum (constraining the shape of $V(\phi)$ at horizon crossing) and the details of reheating (constraining $V(\phi)$ near its minimum together with $\mathcal{L}_I$) — bear on largely independent aspects of the Lagrangian. The shape of the potential over the ≈ 8 e-folds during which observable modes crossed the Hubble radius is essentially unconstrained by whatever happens during reheating, and vice versa. Rather than providing mutual checks, these two sources of evidence constrain disconnected segments of the inflaton's dynamical history.

This situation contrasts sharply with cases of overdetermination that proved decisive in the history of physics, such as early twentieth century measurements of Avogadro's number To see the force of this style of argument, we need to first provide more careful articulation of the concept of a "linking assumption," mentioned above. By a linking assumption we mean the theoretical relationship that connects observations to the parameter(s) being constrained. In Perrin's review of different methods of measuring molecular scale parameters, each route to N relied on a distinctive linking assumption. For Brownian motion, for example, Perrin's inferences based on the displacement of granules assumed that the mean translational kinetic energy of granules undergoing Brownian motion equals, at a given temperature, the mean translational kinetic energy of the molecules of the substrate in which the granules are suspended. The route through emulsion sedimentation invoked the Boltzmann distribution applied to suspended particles in a gravitational field. The route through radioactive decay assumed the counting of alpha particles emitted per gram of radium per second, combined with measurements of the volume of helium produced.[12] The linking assumptions in each case are largely disjoint: a failure of the kinetic theory underlying Brownian motion would not directly undermine the measurements based on radioactive decay. And the agreement among the resulting determinations obtained by the early 1920s is often taken to exemplify a successful overdetermination argument.

The features of these linking assumptions that made them suitable for an evidential argument also made them robust against changes in the underlying physics. In the decades after Perrin's work, the picture of the atom changed dramatically with the advent of quantum mechanics. Yet the determinations of N did not require revision. The linking assumptions these arguments relied on did not depend on the internal structure of the atom, on the character of molecular interactions, or on the detailed dynamics governing collisions. They depended instead on quite general results that held under quantum mechanics just as they had under classical kinetic theory. The linking assumptions were stable in a strong sense: they survived a revolutionary change in the underlying physics. In fact, the availability of the precise values of these parameters played an essential role in the development of this new physics — to take one example, Bohr's calculation of the Balmer series relied on precise values of molecular scale parameters, including N.

This example illustrates the value of *specificity*, in the following sense: the linking

[12] See, in particular, **SmithSeth2020** for a masterful assessment, which we rely on here.

assumptions yielded determinate, quantitatively precise values of N that could be used as premises for further inquiry. The agreement among these estimates — N pinned down to within a few percent across several distinct types of measurement — was specific enough to serve as a starting point for further work, and it underwrote calculations of other quantities which presupposed a value of N. And it was specific enough that the residual disagreements among the different methods could themselves be informative: a determination that fell systematically outside the range of others plausibly indicated a need to revise or reconsider the relevant linking assumption (as, ironically, proved to be the case with Perrin's own measurements). This is the sense of specificity that closing the loop requires — not that the framework's predictions are unique to it, which is a separate question about whether competing hypotheses could reproduce them, but that the framework is detailed enough to support further inquiry.

This is a more demanding standard than mere insensitivity to local modeling choices. It is stability against deep theoretical revision of the underlying physics — the kind of stability one wants in linking assumptions that are doing the work of converting observations into evidence about deeper structural features. A linking assumption for the case of inflation that depended sensitively on the detailed character of the UV completion would be a fragile bridge, vulnerable to revision whenever the UV theory was revised. There is also a complementary feature of Perrin's case worth noting. As **SmithSeth2020** have emphasized, Perrin's experiments established certain facts about Brownian motion — that the kinetic energy of suspended granules far exceeds what their visible motion would suggest, that this kinetic energy varies linearly with temperature and is independent of fluid properties, that fluids have a discrete microstructure on scales below the granule size — independently of the molecular hypothesis itself. The linking assumptions were not only stable against future revisions in atomic theory; they had support from below, in lower-energy regularities that could be experimentally established without committing to any particular account of what molecules were. This bidirectional anchoring — robustness against future revisions and support from independently established regularities — is what made the atomic hypothesis specific in the strong sense relevant here.

All of these features contrast sharply with the case of inflationary cosmology. The linking assumptions connecting features of the primordial perturbation spectrum to the inflaton Lagrangian are sensitive to new physics, in this case UV physics such as symmetries at higher energies, in ways that Perrin's were not.[13] Nor are the linking assumptions anchored in independently established lower-energy regularities, in the way Perrin's were by the kinetic-theoretic facts that his experiments established without commitment to the molecular hypothesis. And the framework lacks specificity: current observations are compatible with a number of distinct models, characterized by different values of $V(\phi)$. Several classes of models remain consistent with Planck constraints, and the data do not fix the relevant features of $V(\phi)$ tightly enough to underwrite further inferences in the way Perrin's value of N did. Tensions between specific models and observations typically prompt a move to a different model within the existing modeling space, rather than the identification of a physical source for the discrepancy. Both the stability that the Perrin case exemplifies and the specificity it required for iterative leverage are absent.

[13] We return to these issues below; see also **KoberinskiSmeenk2024** for further discussion.

There is, however, one prospect for partial overdetermination within the inflationary framework that is promising. The consistency relation linking the tensor-to-scalar ratio r to the tensor spectral index n_t would, if it were to be measured, provide a genuine internal check on the inflationary mechanism for amplifying perturbations: a detection of primordial gravitational waves, together with measurements of both r and n_t, would overdetermine the shape of $V(\phi)$ at a specific point. Even at this more modest scale — a single internal consistency check, not a convergence of diverse phenomena on the same theoretical quantity — the relation would provide a kind of evidential leverage that the current situation lacks. No primordial tensor signal has yet been detected, though, and current upper bounds ($r < 0.036$) push the consistency relation further from observational reach.[14] In the absence of this test, the empirical case for inflation rests on compatibility rather than overdetermination.

Obstacles to Iterative Refinement

The second dimension of closing the loop concerns iterative refinement: the ability of the theory to serve as a starting point for a progressive sequence of approximations in which discrepancies with observation are traced to identifiable physical sources. This requires a framework that is specific enough for discrepancies to be *informative* — indicating where the current model is incomplete — rather than merely signalling that a different model should be chosen from within a large space of alternatives.

Inflation as it currently stands does not support this kind of inquiry, for two interrelated reasons. Consider, first, a striking historical fact: in four decades of confrontation with increasingly precise CMB observations, as far as we are aware, no discrepancy between inflationary models and data has been used to identify a robust new physical feature of the early universe.[15] The pattern is rather the opposite. Specific models have been ruled out by observations — Guth's original proposal, many large-field potentials, classes of plateau models inconsistent with the most recent Planck bounds — but in response cosmologists have selected a different model from the existing space, rather than refining a model by identifying a physical source for the discrepancy. The lack of a canonical model anchored in independently constrained physics leaves no settled starting point from which discrepancies could serve a diagnostic role. (It is admittedly also the case that the inaccessibility of the relevant regimes would make it extremely challenging to obtain robust evidence for any sources that might be introduced to resolve a discrepancy.) One might think the EFT approach bypasses this problem by working at a more abstract level: rather than committing to a specific potential, one constrains the coefficients of an effective action common to a wide class of models. But as we argued in §8.3.4, the EFT framework's physical significance is itself in question. The flexibility that the EFT was supposed to manage is not transformed into a structured, physically grounded framework. Working at the EFT level inherits, rather than escapes, the difficulty.

[14] For small values of r, the measurement of n_t required to test the consistency relation becomes extremely challenging. See ... references ...

[15] This is not to deny that there has been progress in developing more detailed and sophisticated models of inflation; the claim is rather that there are no analogs of kinds of robust sources identified in celestial mechanics, such as new planets, the physical significance of configurational details of the orbits, and so on.

The second obstacle is the modular structure of the inflationary account. As discussed in §8.3.4, inflation effectively decomposes into three distinct dynamical epochs — the onset of inflation, the slow-roll amplification of perturbations, and reheating — each governed by different features of the Lagrangian and subject to largely different observational constraints. The EFT framework covers only the middle epoch; the onset and end fall outside its scope. This modularity has direct consequences for the overdetermination diagnosis of §8.4.2. The available observations do not bear on common theoretical quantities in the way overdetermination requires; rather, they constrain different segments of inflation's dynamical history. In the absence of a canonical model that ties these three domains together, there is essentially no mutual check between these different aspects. Progress in understanding one epoch does not constrain or inform the others. The interlocking web of theoretical relationships that characterizes a mature physical theory — where understanding one aspect tightens constraints on others — is almost entirely absent.

8.4.3 Alternatives

The obstacles to closing the loop identified above are internal to the inflationary program: they concern the flexibility of the modeling space and the sensitivity of the framework to unknown physics. But there is a further, external challenge. Over the past several decades, a diverse space of alternatives to inflation has been developed, motivated by different physical frameworks and employing distinct mechanisms for generating the primordial perturbation spectrum. These alternatives matter for the present argument not because any one of them is clearly superior to inflation, but for two other reasons: they provide a reference class against which the distinctiveness of inflation's observational successes can be assessed, and they help to clarify what a viable theory of the early universe actually needs to accomplish.

Consider first the question of distinctiveness. Several of inflation's most widely cited observational successes — the near scale-invariance, Gaussianity, and adiabaticity of the primordial perturbation spectrum — turn out to be features shared by a broader class of proposals. **HollandsWald2002** construct a simple model that produces a nearly scale-invariant spectrum without any inflationary phase, based on a different ansatz for the initial conditions of the perturbation modes: modes are taken to be "born" in their ground state when their proper wavelength equals the Planck scale, motivated by considerations regarding the domain of applicability of semiclassical quantum gravity. The dynamical mechanism, overdamping of modes with wavelengths much larger than the Hubble radius, is the same as in inflation, but there is no horizon crossing, and the result depends critically on assumptions about initial conditions that inflation is often claimed to avoid. More developed alternatives, including bouncing cosmologies and string gas cosmology, reproduce these broad spectral features through physically distinct mechanisms. In bouncing cosmologies — a family that includes the ekpyrotic and cyclic scenarios (**Lehners2008**), the matter-bounce scenario (**Brandenberger2012**), and pre-Big Bang cosmology (**GasperiniVeneziano2003**) — a contracting phase preceding a bounce replaces the inflationary expansion. Perturbation modes that start at sub-Hubble scales during contraction are stretched past the Hubble radius and amplified, providing an alternative route to a nearly scale-invariant, Gaussian, adiabatic spectrum. In string gas cosmology (**BrandenbergerVafa1989**), the early universe begins in a quasi-static, string-scale phase, and thermal fluctuations of a gas of strings

seed structure formation through a mechanism entirely different from the horizon-crossing amplification of vacuum fluctuations.

Where the alternatives diverge most sharply from inflation is in their predictions for the features of the CMB that have not yet been measured with sufficient precision. The tensor perturbation spectrum is the most important discriminant: ekpyrotic models typically predict a much smaller tensor-to-scalar ratio r than standard slow-roll inflation, and string gas cosmology predicts a slight *blue* tilt for tensor perturbations, in contrast with inflation's red tilt. The alternatives also make distinct predictions for the shape and amplitude of non-Gaussianities, and for features of the perturbation spectrum at the very largest angular scales. The significance of this pattern should be clear: the observations that inflation fits — the scalar power spectrum, Gaussianity, adiabaticity — are observations that any viable account of structure formation would need to fit, and several physically distinct proposals do fit them. The observations that would discriminate among these proposals — the tensor spectrum, non-Gaussian signatures, the detailed form of departures from exact scale-invariance — still remain out of reach.

This pattern has a direct bearing on our Keplerian assessment. Where the EFT analysis of §8.3.4 applied the first strategy within the inflation research program — looking for common structure across models that all assume an inflationary mechanism — the alternatives lead to a different assessment, looking for common structure across rival programs with quite different underlying physics. The diagnostic verdict in this case is of the second kind flagged in the introduction. The EFT-of-inflation framework does not apply to Hollands-Wald, bouncing cosmologies, or string gas cosmology; these proposals do not share the symmetry structure that organizes the inflationary EFT, nor any other comparable structural features. What they share with inflation is only reproducing a nearly scale-invariant, Gaussian, adiabatic spectrum of primordial scalar perturbations, which is a criteria of adequacy for any account of structure formation. The different programs produce this phenomenology through physically distinct mechanisms, which could in principle be distinguished by observations (of the tensor spectrum, non-Gaussianities, and other features) that are not yet available. There is no shared structural framework that covers these rival programs, and only the observable output is held in common. The apparent underdetermination is real, although perhaps only transient if further observations provide access to the features of the early universe that would discriminate among them.

The comparison with alternatives also helps to clarify what a viable theory of the early universe needs to accomplish. Inflation is typically presented as a package deal: it simultaneously addresses the horizon and flatness problems and provides a mechanism for generating primordial perturbations. This packaging has shaped how inflation is evaluated, with successes in one domain taken to support the entire program, including its fine-tuning motivations. But as Brandenberger has emphasized in a series of review articles, these two achievements can and should be assessed independently.[16] A theory can provide a successful account of structure formation without addressing the horizon or flatness problems at all (as in the Hollands-Wald case), or it can address both but through physically distinct mechanisms (as in some bouncing cosmologies, where the contraction phase resolves the horizon problem and a separate mechanism generates perturbations). This decomposition

[16] See, e.g., **Brandenberger2012**; Brandenberger (2014).

matters because it clarifies where the observational evidence actually bears.

Once the inflationary package is decomposed in this way, the question of how to evaluate competing proposals shifts accordingly. The essential desideratum is which theory provides an account of structure formation that is sufficiently specific and stable to serve as a starting point for further inquiry. On this score, inflation remains the most developed proposal: no alternative has been subjected to the same sustained program of observational confrontation, and none has achieved the same level of theoretical articulation.[17] But development and articulation are not the same as specificity and stability. The existence of physically distinct alternatives that reproduce inflation's main observational successes makes it particularly difficult to maintain that these successes are diagnostic of the inflationary mechanism, rather than generic features of any viable account of structure formation. And the features that *would* be diagnostic (the consistency relation, the tensor spectrum, the detailed form of non-Gaussianities) remain untested.

8.5 Conclusions

We began this chapter by introducing two strategies, drawn from Kepler, for responding to skeptical challenges about the reach of theoretical reasoning. The first seeks to identify common structure across allegedly competing hypotheses, showing that what is actually doing the evidential work is shared among the rivals, so that the apparent underdetermination dissolves. The second insists on going beyond mere compatibility with the data, demanding that hypotheses connect to a broader account of physics and underlying causes. Together, these strategies constitute a Keplerian defense of the physical significance of a theory that saves the phenomena. Using the theory as a starting point for a progressive, self-correcting line of inquiry provides the most compelling response to Duhemian skepticism. The question we have pursued in this chapter is whether, and to what extent, inflationary cosmology supports such a defense.

Our assessment is mixed, in a way that is instructive about the relationship between saving the phenomena and establishing the physical significance of a theory. Inflation has remained the consensus account of the early universe through four decades of increasingly precise observations. It provides a compelling account of structure formation that was empirically favored over its principal rival, topological defects, on the basis of its compatibility with the CMB power spectrum. This is a genuine and substantial achievement. The success has not been accompanied, however, by the identification of a clear physical source for inflationary expansion. Pursuing the two Keplerian strategies reveals a significant gap between this phenomenological success and the kind of evidence that would warrant treating inflation as a secure foundation for further inquiry.

The observations that inflation fits — a nearly scale-invariant, Gaussian, adiabatic spectrum of primordial scalar perturbations — turn out to be shared by physically distinct proposals for the early universe. The common structure here is not reassuring in the way Kepler intended: rather than showing that apparent rivals share the same evidential core, it shows that the features inflation successfully accounts for are generic, in the sense that

[17] This asymmetry may be partly sociological — inflation's dominance has shaped the allocation of theoretical and observational effort — but it also reflects genuine features of the theory, including the relative simplicity of the inflationary account and how it can be treated with QFT methods.

any viable account of structure formation would need to reproduce them. The observations that *would* discriminate among the proposals are precisely those that remain beyond observational reach. The second strategy, going beyond the numbers to establish a connection with underlying physics, has fared no better. In its initial formulation, inflation was closely connected to high energy physics, and inflation resulted from fields already present in extensions of the Standard Model. Had this connection held, it would have provided exactly what the second strategy demands: a link between the cosmological hypothesis and an independently constrained account of physics. Even if inflaton particles could never be produced at CERN, constraints on the relevant fields from collider experiments, precision electroweak measurements, or proton decay searches would have provided non-cosmological constraints on the inflaton Lagrangian, partially breaking the isolation of early universe cosmology from the rest of physics.

This early promise has not been fulfilled. The amplitude of the primordial perturbation spectrum ruled out the original candidates, and the inflaton became a phenomenological placeholder — a scalar field whose properties are chosen to fit cosmological data, with its microphysical identity left open. The result has been a striking shift: from substantive unification, in which the same physical model had consequences across multiple domains, to methodological unification, in which early universe theorists employ the techniques of quantum field theory without the tight physical constraints that originally motivated the program.

We can now pull these threads together and summarize our diagnosis. On the theoretical side, the inflationary framework is not stable or specific in the senses that closing the loop requires. The EFT framework, as we argued in §8.3.4, is sensitive to UV physics through the η problem, the trans-Planckian problem, and — most pointedly — the cosmological constant problem. The modular structure of the inflationary account, with different dynamical epochs governed by different features of the Lagrangian and constrained by disjoint observational windows, prevents the kind of mutual constraint that would let observations of one regime inform theoretical commitments about another (§8.4.2). On the observational side, two facts about our epistemic situation reinforce these theoretical features: the inaccessibility of the relevant energy scales to terrestrial experiment, and the restriction of the cosmological window to a handful of features of the CMB and large-scale structure. These are, strictly speaking, contingencies, reflecting our position in the universe and what is accessible to us. Better experiments and broader observational coverage could in principle relax them, but in practice it would be enormously difficult to overcome the constraints imposed by our location and capabilities. The combination of these theoretical features with the observational limits just noted is what makes going past merely phenomenological success so challenging.

The point comes into sharper relief if we compare inflation with another novel physical component introduced into modern cosmology to explain otherwise puzzling features of the universe. Dark matter, like the inflaton, was introduced as a posited physical component whose detailed properties remain unknown. But there is a crucial structural difference. Dark matter, once present in the early universe, *remains present* throughout the universe's subsequent history. It continues to have dynamical effects across an enormous range of scales and epochs: it shapes galaxy rotation curves, produces gravitational lensing signatures, leaves a characteristic imprint on the CMB anisotropy spectrum, and influences the baryon

acoustic oscillation signal in large-scale structure. These diverse, independent consequences are what makes closing the loop possible in the dark matter case, as we have argued above: they provide multiple lines of evidence that bear on the same underlying quantities (the dark matter density and its distribution), enabling the kind of overdetermination and iterative refinement that we have argued constitutes genuine empirical success.

The inflaton's role in the universe's history is fundamentally different. It dominates the dynamics during a transient phase of exponential expansion, seeds the primordial perturbation spectrum, and then, in effect, disappears from the scene: the inflaton field itself has no ongoing presence, no continuing dynamical effects that could be probed by independent means at later times. This is not merely a practical inconvenience that future experiments might overcome. It is a structural feature of inflationary cosmology: the physics responsible for inflation is, by construction, confined to an epoch that is accessible only through its imprint on features of the early universe. The contrast with dark matter thus illustrates a general point about the conditions under which closing the loop is feasible: the posited physics must have ongoing, or at least multiply accessible, consequences, so that diverse lines of evidence can provide the mutual constraints needed for overdetermination and iterative refinement.

None of this entails that inflation is false, or that it should be abandoned, or that the account of structure formation it provides is without significant epistemic value. Inflation demonstrates that the initial conditions revealed by the CMB can be generated through a dynamical process in the early universe. In particular, it provides an account of how the coherent, superhorizon perturbations required to seed structure formation could have arisen through causal processes — quantum fluctuations amplified during a phase of accelerated expansion — rather than being stipulated as brute features of an initial state whose causal structure, in the standard FLRW models, seems to preclude any such account. This is a genuine achievement. It resolves what was, prior to inflation, a serious conceptual puzzle: the apparent conflict between the causal structure of the FLRW models and the existence of coherent perturbations on scales far larger than the Hubble radius. A "how possibly" explanation of this kind has real value, precisely because it shows how to resolve the apparent tension between what is required by the standard model of cosmology and what we might take to be a physically plausible initial state.

But a "how possibly" explanation does not have the same status as other claims in physics secured by closing the loop. Showing that certain observed features of the CMB *could* have arisen through a particular mechanism does not establish that they *did* arise in that way, nor does it provide the kind of progressive, self-correcting line of inquiry through which the details of the mechanism could be pinned down with increasing precision. For that, one would need the overdetermination of theoretical parameters and the capacity for iterative refinement that we have argued inflation currently lacks. A detection of primordial gravitational waves, or of non-Gaussianities in the CMB, could significantly alter this assessment: the former would provide a genuine internal check on the inflationary mechanism through the consistency relation, and the latter would constrain the higher-order terms in the effective field theory that currently remain entirely unconstrained. A breakthrough in connecting the inflaton to independently motivated particle physics would be equally transformative, by partially breaking the isolation of early universe cosmology from the rest of physics. But absent such developments, inflation remains at a stage comparable to Ke-

pler's achievement in the *Astronomia Nova*: a physically motivated proposal that saves the phenomena and suggests a promising direction for further inquiry, but whose underlying physics has not yet been pinned down with sufficient precision or stability to serve as a secure foundation for closing the loop.