Dark Energy Versus Modified Gravity

Austin Joyce,[1] Lucas Lombriser,[2] and Fabian Schmidt[3]

[1]Enrico Fermi Institute and Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637; email: ajoy@uchicago.edu

[2]Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, United Kingdom; email: llo@roe.ac.uk

[3]Max Planck Institute for Astrophysics, D-85748 Garching, Germany; email: fabians@mpa-garching.mpg.de

Annu. Rev. Nucl. Part. Sci. 2016. 66:95–122

First published online as a Review in Advance on June 17, 2016

The *Annual Review of Nuclear and Particle Science* is online at nucl.annualreviews.org

This article's doi: 10.1146/annurev-nucl-102115-044553

Keywords

cosmology, dark energy, modified gravity, structure formation, large-scale structure

Abstract

Understanding the reason for the observed accelerated expansion of the Universe represents one of the fundamental open questions in physics. In cosmology, a classification has emerged among physical models for this acceleration, distinguishing between dark energy and modified gravity. In this review, we provide a brief overview of models in both categories as well as their phenomenology and characteristic observable signatures in cosmology. We also introduce a rigorous distinction between dark energy and modified gravity based on the strong and weak equivalence principles.

Contents

1. INTRODUCTION

The accelerated expansion of the Universe, discovered in 1998 (1, 2), has raised fascinating questions for cosmology and physics as a whole. Elucidating the physics behind the acceleration is one of the main science drivers for a major experimental effort using large-scale structure (LSS) surveys. All current observations are consistent with a cosmological constant (CC); whereas the CC is in some sense the most economical possibility, it has its own theoretical and naturalness problems (3, 4), so it is worthwhile to consider alternatives. A convenient classification scheme has emerged in the field of cosmology, separating different physical scenarios of accelerating cosmologies into the categories of dark energy and modified gravity. Essentially, dark energy models modify the stress-energy content of the Universe, adding another component with equation of state $w \simeq -1$. That is, we modify the right-hand side of the Einstein equations. The modified gravity category corresponds to modifying the left-hand side—that is, general relativity (GR) itself—for example, by modifying the Einstein–Hilbert action. This review intends to provide an overview of these two categories, and describe the cosmological observables and tests that are able to distinguish experimentally between these two scenarios of physics driving the acceleration.

As discussed below, although there are models that unambiguously belong to one category or the other, in reality there is a continuum of models between the two extremes of "pure" dark energy and modified gravity such that a strict division into these two categories is to some extent a matter of personal preference. In this review, we base the distinction in theory space upon the strong equivalence principle (SEP). A compatible, slightly less rigorous, but more practical observational distinction in cosmology is to call modified gravity those models that feature a light universally coupled degree of freedom mediating a fifth force.

Throughout this review, we restrict ourselves to observations on extragalactic scales, which translates to scales of order 1 Mpc and above. Furthermore, we consider only scenarios for which

effects become relevant at late times. That is, we disregard the possible presence of such effects in the early Universe (often referred to as early dark energy).

We begin in Section 2 with an overview of the landscape of dark energy and modified gravity models. Section 3 then describes the cosmological phenomenlogy of these models. Section 4 provides a summary of the observables and experimental methods used in cosmology to test dark energy and modified gravity. We conclude with an outlook in Section 5. Throughout, we work in units where $c = \hbar = 1$.

Quintessence models: the phenomenology of these models is rich; however, their qualitative behavior can be of two types (11): models that are just starting to deviate from $w \simeq -1$ (thawing) and models that are approaching $w \simeq -1$ today (freezing)

2. OVERVIEW: DARK ENERGY AND MODIFIED GRAVITY

In this section, we briefly review dark energy and modified gravity models before turning to the tricky issue of drawing a boundary between these two paradigms. We do not attempt to be comprehensive—there exist many excellent reviews on both subjects (see, e.g., References 5 and 6)—but rather want to give a flavor of the types of models and their phenomenology.

We note at the outset that the models we describe are not per se solutions of the CC problem. Generally, it is still necessary to invoke some other physics or symmetry principle to explain why the expected contribution to the CC from Standard Model (SM) fields is absent. Despite lacking a complete solution of the CC problem, the general approaches considered here are promising and represent reasonable possibilities for how physics in the gravitational sector may behave.

2.1. Dark Energy

Perhaps the most natural direction to explore to explain the observed value of the CC is to posit that the CC is itself a dynamical field, which relaxes to its present value through some mechanism (7–9). There are many possible incarnations of this idea; what they share is the presence of a degree of freedom which develops a condensate profile that drives the background cosmological evolution. The simplest one is to imagine that the CC is a canonical scalar field with a potential[1]

$$S = \int \mathrm{d}^4x\sqrt{-g}\left(\frac{M_{\rm Pl}^2 R}{2} - \frac{1}{2}(\partial\phi)^2 - V(\phi)\right). \qquad 1.$$

We refer to this model as quintessence[2] (10). A homogeneous condensate of this field, $\phi = \phi(t)$, behaves as a perfect fluid, with equation of state $w = P/\rho$, given by

$$w = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)}, \qquad 2.$$

where (here and throughout) dots denote derivatives with respect to time. Observations restrict w to be very close to -1 in the present Universe (note that this need not be the case for all times). This is equivalent to the requirement that the evolution of the field be potential dominated. In this sense, the present stage of cosmic acceleration in quintessence models is similar to a period of very low scale inflation. It is also important to note that the quintessence equation of state is generally time dependent.

An intriguing aspect of quintessence models is that a large class of them exhibits so-called tracking behavior (12–15), where the energy density in the field closely traces the energy density

[1] We define the reduced Planck mass as $M_{\rm Pl}^2 = (8\pi G)^{-1}$.

[2] In his influential review, Weinberg (3) showed that models of quintessence necessarily have to be tuned in order to make the CC small.

DARK ENERGY VERSUS MODIFIED GRAVITY

In the text, we draw a distinction between dark energy and modified gravity by means of the strong equivalence principle (SEP). We classify any theory which obeys the SEP as dark energy, and any theory which violates it as modified gravity (see Section 2.3). Heuristically, the SEP forbids the presence of fifth forces.

in dark matter until the near past, providing a possible explanation of the coincidence problem—that is, why the current density of dark energy and dark matter are comparable. The canonical example is the Ratra–Peebles potential (9),

$$V(\phi) = \frac{M^{4+n}}{\phi^n}, \qquad 3.$$

with $n > 0$.

Another consideration to keep in mind with dark energy (and even to some extent modified gravity) models is to be sure that we are not secretly reintroducing a CC. As a simple example, a scalar field with an exactly flat potential could plausibly be thought of as dark energy, but in this case we should really absorb this constant potential into the bare CC in the Lagrangian. Another compelling possibility to consider is that dark energy and dark matter may be unified into a single component. However, this turns out to be so fine-tuned that one effectively returns to a two-fluid model (16).

2.1.1. More general dark energy models. Even within the paradigm of the CC as a dynamical scalar field, there is a plethora of different models, with varied phenomenology. Perhaps the most obvious generalization away from Equation 1 is to consider a noncanonical scalar field—one that possesses derivative self-interactions. Concretely, these models may be written as

$$S = \int \mathrm{d}^4x\sqrt{-g}\Lambda^4 K(X), \qquad 4.$$

where we have defined $X \equiv -\frac{1}{2\Lambda^4}(\partial\phi)^2$ and $K(X)$ is an arbitrary function. Models of this type are often referred to as k-essence (17, 18). Similar to the canonical case, for homogeneous field profiles, $\phi = \phi(t)$, these models behave as a perfect fluid, though of a more general type, where now the equation-of-state parameter w is given by

$$w = \frac{K}{2XK_{,X} - K}. \qquad 5.$$

By suitably choosing a functional form for K, one can reproduce the cosmic expansion history. Field perturbations away from the background evolution no longer propagate luminally in k-essence models; instead, they propagate with a speed of sound given by

$$c_s^2 = \frac{K_{,X}}{K_{,X} + 2XK_{,XX}}. \qquad 6.$$

This allows structure formation to be noticeably different with respect to canonical dark energy models, as we discuss in Section 3.3.

Dark energy models of the late-time acceleration are similar to inflation, and there has been an effort to adapt the effective field theory of inflation formalism (19) to the problems of dark energy and modified gravity (20–24).

2.2. Modified Gravity

We begin this section by reviewing some commonly studied IR modifications of GR. These examples allow us to draw some more general lessons: Generic modifications lead to new physics on small scales (25). Indeed, all known modifications to Einstein gravity introduce new degrees of freedom in the gravitational sector.[3] These new particles mediate a fifth force, so the theory must employ some screening mechanism in order to evade local tests of gravity, which are very constraining.

2.2.1. Scalar–tensor theories. Scalar–tensor theories are probably the best-studied modifications to Einstein gravity. Restricting to canonical scalar fields for the moment, these theories can be cast in Einstein frame as

$$S = \int \mathrm{d}^4x\sqrt{-g}\left(\frac{M_{\rm Pl}^2 R}{2} - \frac{1}{2}(\partial\phi)^2 - V(\phi)\right) + S_{\rm matter}\left[A^2(\phi)g_{\mu\nu}, \psi\right]. \qquad 7.$$

Here, R is the Ricci scalar constructed from the metric $g_{\mu\nu}$. Matter fields, ψ, couple to the Jordan-frame metric $\tilde{g}_{\mu\nu} = A^2(\phi)g_{\mu\nu}$. This coupling induces interactions between matter fields and ϕ, causing test particles in the Newtonian regime to feel a force,

$$\boldsymbol{a} = -\boldsymbol{\nabla}\Big(\Phi + \ln A(\phi)\Big), \qquad 8.$$

sourced by both the Einstein-frame potential, Φ, and the scalar, ϕ.

The prototypical scalar–tensor theory is Brans–Dicke theory (26), which corresponds to $V(\phi) = 0$ and $A^2(\phi) = \exp[-\phi/(M_{\rm Pl}\sqrt{3/2+\omega})]$ in Equation 7. This theory is usually cast in Jordan frame, $g_{\mu\nu} \mapsto \tilde{g}_{\mu\nu}$, where it takes the form

$$S = \frac{M_{\rm Pl}^2}{2}\int \mathrm{d}^4x\sqrt{-\tilde{g}}\left(\phi\tilde{R} - \frac{\omega}{\phi}(\partial\phi)^2\right) + S_{\rm matter}\left[\tilde{g}_{\mu\nu}, \psi\right], \qquad 9.$$

with ω a constant parameter. Solar System tests of gravity place the bound $\omega \gtrsim 4\times10^4$ (6, 27). This bound corresponds to taking the coupling to matter to be very weak, making the Brans–Dicke theory essentially equivalent to a dark energy model.

In more general scalar–tensor theories—by suitably choosing $V(\phi)$ and $A(\phi)$—it is possible to evade Solar System constraints, while still having interesting phenomenology for the scalar. To see how this can work, we focus on the usual case where the scalar field ϕ couples to the trace of the Jordan-frame stress tensor, $\tilde{T} = A^{-4}T$, where T is the trace of the stress tensor in Einstein frame. If we consider nonrelativistic sources, $T = -\rho$, and define $\bar{\rho} \equiv A^{-1}\rho$, the scalar obeys the equation of motion

$$\Box\phi = \frac{\mathrm{d}V_{\rm eff}}{\mathrm{d}\phi}, \qquad V_{\rm eff}(\phi) = V(\phi) + A(\phi)\bar{\rho}. \qquad 10.$$

The key point is that ϕ responds to an effective potential, which depends on the external matter sources. This makes it possible to engineer situations where the field behaves differently in different environments.

[3] In this review, we focus on modified gravity applications to cosmic acceleration. There has also been research on modifying gravity to obviate the need for dark matter, but we do not discuss that topic here.

The best-known example is the chameleon field (28, 29), where the potential and matter coupling are chosen so that the effective mass of the scalar field,

$$m_{\rm eff}^2(\phi) = \frac{{\rm d}^2 V_{\rm eff}}{{\rm d}\phi^2} = \frac{{\rm d}^2 V}{{\rm d}\phi^2} + \frac{{\rm d}^2 A}{{\rm d}\phi^2}\bar{\rho}, \qquad 11.$$

increases in regions of high density, like the Solar System. A concrete example is the Ratra–Peebles potential and a linear coupling (28, 29):

$$V(\phi) = \frac{M^{4+n}}{\phi^n}, \qquad A(\phi) \simeq 1 + \xi\frac{\phi}{M_{\rm Pl}}. \qquad 12.$$

Another popular scalar–tensor model with a screening mechanism is the symmetron model (30), which can also be cast as Equation 7, but with

$$V(\phi) = -\frac{\mu^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4, \qquad A(\phi) \simeq 1 + \frac{\phi^2}{2M^2}. \qquad 13.$$

In regions of low density, the $\mathbb{Z}_2$ symmetry of the model is spontaneously broken and ϕ mediates a fifth force; in regions of high density, the symmetry is restored and the additional force turns off.

More precisely, in both the chameleon and the symmetron a screened field is sourced only by a thin shell near the surface of a dense object. Whether or not the screening operates depends on the Newtonian potential of the object exceeding a certain threshold (29, 30); see Section 2.2.4.

One of the most-studied scalar–tensor theories is so-called $f(R)$ gravity, which on its face appears not to be a scalar–tensor theory at all. This model consists of higher-curvature corrections to the Einstein–Hilbert action

$$S = \frac{M_{\rm Pl}^2}{2}\int {\rm d}^4x\sqrt{-g}\Big(R + f(R)\Big) + S_{\rm matter}[g_{\mu\nu}, \psi], \qquad 14.$$

where $f(R)$ is a function only of the Ricci scalar, chosen to become significant in the low-curvature regime $R \to 0$. $f(R)$ models have been used to drive cosmic acceleration (31–33). However, these original models are actually in conflict with precision tests of gravity (34), essentially because these models are scalar–tensor theories in disguise (35, 36). One can see this by performing a field redefinition and conformal transformation (see, e.g., Reference 37 for details) to cast the theory as in Equation 7 with

$$V(\phi) = \frac{M_{\rm Pl}}{2}\frac{(\phi f_{,\phi} - f)}{(1 + f_{,\phi})^2}, \qquad A(\phi) = e^{\frac{\phi}{\sqrt{6}M_{\rm Pl}}}, \qquad 15.$$

which is equivalent to the Brans–Dicke theory in Equation 9 with $\omega = 0$ and a potential.

$f(R)$ models compatible with local tests of gravity were constructed in References 38 and 39. For example, in Reference 38 $f(R)$ is given by

$$f(R) = -m^2\frac{c_1\left(\frac{R}{m^2}\right)^n}{c_2\left(\frac{R}{m^2}\right)^n + 1} \approx \rho_{\Lambda,{\rm eff}} + f_{R0}\left(\frac{R}{\bar{R}_0}\right)^{-n}, \qquad 16.$$

where the second form is the leading expression for $R/m^2 \ll 1$, which is required by cosmological and local tests (Section 4). The parameters c_1 and c_2 can be adjusted so that $\rho_{\Lambda,{\rm eff}}$ supplies a CC consistent with observations, leaving f_{R0} as a free parameter. $\bar{R}_0$ is the background Ricci scalar today, so that f_{R0} corresponds to the field value in the background today. In scalar–tensor language, the model is consistent with experiment because of the chameleon mechanism. $f(R)$ gravity has been explored in a variety of contexts, and we refer the reader to Reference 40 for more details.

2.2.2. More general scalar–tensor theories: Horndeski and generalizations. The space of scalar–tensor theories is far broader than the example presented in Equation 7. In fact, in recent years, there has been an effort to map out the most general consistent theory of a metric interacting with a scalar. In this context, consistency requires that the theory be absent of ghosts. Typically, theories with equations of motion that are higher than second order in time derivatives have a ghost.[4] Therefore, theories that have second-order equations of motion have attracted a great deal of interest.

Ghost: a quantum with either negative energy density or negative norm. The existence of propagating ghosts in a theory poses problems for its quantization; in particular, the vacuum is unstable to decay into ghosts and healthy particles (43, 44)

The most general theory of a scalar field interacting with a metric that has second-order equations of motion has been derived (42); it corresponds to the Lagrangian

$$\begin{aligned}\mathcal{L}_{\text{gen.gal.}} &= K(\phi, X) - G_3(\phi, X)\Box\phi + G_4(\phi, X)R \\ &\quad + G_{4,X}(\phi, X)[(\Box\phi)^2 - (\nabla_\mu\nabla_\nu\phi)^2] + G_5(\phi, X)G_{\mu\nu}\nabla^\mu\nabla^\nu\phi \\ &\quad - \frac{1}{6}G_{5,X}(\phi, X)[(\Box\phi)^3 - 3(\Box\phi)(\nabla_\mu\nabla_\nu)^2 + 2\nabla^\mu\nabla_\alpha\phi\nabla^\alpha\nabla_\beta\phi\nabla^\beta\nabla_\mu\phi].\end{aligned} \quad 17.$$

This theory has four arbitrary functions, K, G_3, G_4, and G_5, and X is defined as in Equation 4. Interestingly, an equivalent theory had been derived much earlier by Horndeski (45, 46), but had gone essentially unnoticed in the literature.

A particularly well studied limit of this theory is the so-called galileon (47). The simplest version of this theory—the cubic galileon—is a special case of Equation 17:

$$\mathcal{L}_{\text{gal.}} = \frac{M_{\text{Pl}}^2 R}{2} - \frac{1}{2}(\partial\phi)^2 - \frac{1}{\Lambda^3}\Box\phi(\partial\phi)^2. \quad 18.$$

In the limit where gravity is taken to be nondynamical, this theory has the symmetry $\delta\phi = c + b_\mu x^\mu$. In general dimensions, there are only a finite number of terms invariant under this symmetry, which also have second-order equations of motion. The galileon appears in many constructions; in particular, it describes the decoupling limits both of massive gravity (48) and of the Dvali–Gabadadze–Porrati (DGP) model (49, 50). For a review of many interesting properties of the galileon, see Reference 51.

Recently, an effort has been made to relax the assumption of second-order equations of motion in order to construct the most general scalar–tensor type theory that propagates three degrees of freedom nonlinearly. In such theories, the expected ghostly degrees of freedom from higher-order equations of motion are projected out by constraints (see Reference 52 and references therein).

2.2.3. Massive gravity and brane-worlds. Einstein gravity is the theory of a massless spin-2 particle. A natural question to ask is whether it is possible for the graviton to be a massive particle. The intuition behind this explanation for cosmic acceleration is that massive fields induce potentials of the Yukawa type,

$$V(r) \sim \frac{e^{-mr}}{r}, \quad 19.$$

which shut off at distances of order m^{-1}. The idea is that the presently observed cosmic acceleration could be due to a weakening of gravity at large scales, with $m \sim H_0$.

Massive gravity has been studied since the initial investigations of Fierz & Pauli (53), but much of the modern interest is due to a ghost-free theory by de Rham, Gabadadze, and Tolley (dRGT) (48, 54), built on earlier research (55, 56), which showed how to overcome the traditional difficulties associated with nonlinearly completing the Fierz–Pauli theory. In particular, generic nonlinear theories propagate a ghost in addition to the graviton degrees of freedom (57). The

[4]This is the content of Ostrogradsky's theorem (41).

absence of this Boulware–Deser ghost in the dRGT theory was shown in Reference 58. Recovery of GR in the massless limit relies on the Vainshtein mechanism (see Section 2.2.4). For reviews of the theoretical background and recent developments, see References 59 and 60.

ϕ and Φ_N: on subhorizon scales in cosmology, and within the Compton wavelength of the field, ϕ, and the Newtonian potential, Φ_N, are proportional in the absence of screening, as may be deduced using the scalar equation of motion and the Poisson equation

Another avenue to massive gravity is as a resonance (as opposed to the hard mass above). This idea has been very influential in cosmology; the most-studied example is the DGP model (49). The setup consists of a four-dimensional (4D) brane (on which matter fields live) embedded in an infinitely large five-dimensional (5D) bulk space-time. The model consists of dynamical gravity both on the brane and in the bulk:

$$S = \frac{M_5^3}{2}\int \mathrm{d}^5X\sqrt{-G}\mathcal{R} + \frac{M_4^3}{2}\int \mathrm{d}^4x\sqrt{-g}R, \qquad 20.$$

where M_5 and M_4 are the 5D and 4D Planck masses; G and $\mathcal{R}$ are the 5D metric and Ricci scalar; and g and R are the 4D metric and Ricci scalar. From the perspective of an observer living on the brane, gravity is mediated by a continuum of gravitons, and at short distances, gravity appears to be 4D, but at large distances, the gravitational potential "leaks" off the brane and is that of a 5D theory. The potential due to a mass M source behaves as (59, 61)

$$V(r) \simeq \begin{cases} \dfrac{M}{M_4^2 r} & r \ll r_c \\ \dfrac{M}{M_4^2 r}\dfrac{r_c}{r} & r \gg r_c \end{cases}, \qquad 21.$$

where $r_c = M_4^2/2M_5^3$. Therefore, the potential is weaker at large distances than it would be in pure GR. This weakening allows for de Sitter solutions on the brane absent a bare CC (61, 62). Unfortunately, this self-accelerating branch of solutions is unstable—perturbations are ghostly (50, 63). The DGP model remains quite interesting as a benchmark model; it is one of the better-studied modified gravity models in the literature. For a review of DGP, see Reference 64.

Interestingly, a massive graviton could address the CC problem by means of degravitation (65, 66): A large CC nevertheless does not gravitate, because a massive graviton behaves as a high-pass filter (67). However, a realistic degravitating solution has not yet been found in either dRGT or brane-world.

2.2.4. Screening mechanisms. In all of the models we have discussed, new degrees of freedom appear in the gravitational sector. This is in fact a general feature: GR is the unique low-energy theory of a massless spin-2 particle (68, 69), so essentially any departure from GR introduces new degrees of freedom, typically with a mass of order the Hubble parameter today: $m \sim H_0$. Because they couple to the SM, these light fields mediate a long-range force between matter sources.

The presence of additional forces is strongly constrained by laboratory and Solar System tests (27). Therefore, the additional degrees of freedom must hide themselves locally. The screening mechanisms by which this is achieved fall into several broad classes, activating in regions where the Newtonian potential Φ_N or successive gradients become large. Φ_N is defined as the potential that solves the Poisson equation (Equation 26). For concreteness, we assume that the additional degree of freedom is a scalar in the following discussion, but the general philosophy is much broader. For a more complete discussion, see Reference 37.

These classes of models are described as follows.

1. Screening at large field values. The first type of screening we consider activates in regions where the Newtonian potential exceeds some threshold value: $\Phi_N > \Lambda$. Generally, this leads to the additional degree of freedom itself developing a large vacuum expectation value—which causes the coupling to matter to weaken, the mass of the field to increase, or the self-coupling of the field to become large—leading to a diminishing of the force mediated. Notable examples are the chameleon (28, 29), symmetron (30, 70, 71), and dilaton (72, 73). From a phenomenological viewpoint, these mechanisms activate in regions of large potential, Φ_N, so regions of small Newtonian potential should exhibit the largest deviations from GR.
2. Screening with first derivatives. This mechanism turns on when the local gravitational acceleration exceeds some critical value: $\nabla\Phi_N > \Lambda^2$. This condition roughly corresponds to first gradients of the scalar field becoming large; screening occurs due to kinetic self-interactions: $\partial\phi/\Lambda^2 \gg 1$. General $P(X)$ models can display this kind of screening (74); a broad class of these models is referred to as k-mouflage (75). Because these models screen in regions of large acceleration, it is intriguing to imagine that they could be relevant for reproducing the phenomenological successes of MOND (76) in a more complete framework (77, 78).
3. Screening with second derivatives. The last category of screening mechanisms we discuss is those that become active in regions of large curvature: $\nabla^2\Phi_N > \Lambda^3$, which is equivalent to high density. These mechanisms rely on nonlinearities in the second derivatives: $\Box\phi/\Lambda^3 \gg 1$. The most commonly studied example in this class is the Vainshtein mechanism (79), which operates in the galileon and in massive gravity. In these models, the largest deviations from GR are expected in low-curvature regimes. Additionally, there is evidence that Vainshtein screening is less efficient than naïve estimates in time-dependent scenarios (80, 81), making this a promising avenue to search for deviations from GR.

Test bodies: objects that are sufficiently small that we may neglect gravitational tidal forces in deriving their motion; a more precise definition can be found in Reference 82

2.3. Drawing a Boundary

In order to meaningfully discuss observationally discriminating between modified gravity and dark energy, we must draw a distinction between the two scenarios. This is a somewhat aesthetic choice, but nonetheless it helps us organize our thinking about exploring and testing the various possibilities.

The distinction we make relies essentially upon the motion of bodies in the theory. To begin, recall the weak equivalence principle (WEP). The WEP is the statement that there exists some (usually Jordan-frame) metric to which all matter species couple universally. Then, test bodies—regardless of their composition—fall along geodesics of this metric. This is usually stated as the equivalence of inertial mass and gravitational mass. In this review, we focus on theories that satisfy the WEP at the level of the action.

To distinguish dark energy versus modified gravity, we further invoke the SEP. The SEP extends the universality of free fall to be completely independent of a body's composition, including gravitational binding energy, so compact objects like black holes also follow geodesics (83). We call anything that obeys the SEP dark energy, and anything that does not, modified gravity. The motivation for this definition is to classify models that influence ordinary matter only gravitationally as dark energy. In these models, the force felt between two bodies is only that of GR (and possibly other SM forces). However, in models of modified gravity two isolated bodies may carry additional charges (e.g., scalar charge), which causes them to experience an additional force beyond that of gravity. The appeal to the SEP is an attempt to make this intuition precise. A

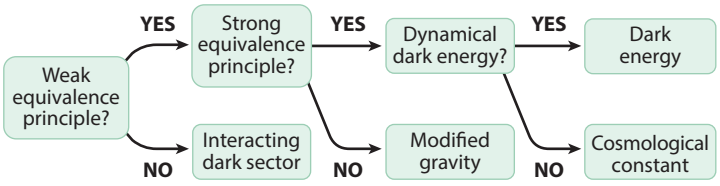

Figure 1

A flowchart of the distinctions made in the text.

theoretical motivation for this distinction based upon the SEP is that it is believed—although not proven—that GR is the only metric theory that obeys the SEP (27, 83).

The preceding discussion is somewhat abstract, so it is useful to illustrate the main points with scalar–tensor theory. We again consider the action Equation 7, but allow for each matter species to have a different coupling to ϕ:

$$S = \int \mathrm{d}^4x\sqrt{-g}\left(\frac{M_{\rm Pl}^2}{2}R - \frac{1}{2}(\partial\phi)^2 - V(\phi)\right) + S_{\rm matter}\left[A_i^2(\phi)g_{\mu\nu}, \psi_i\right]. \qquad 22.$$

Here, the notation $A_i^2(\phi)$ captures the fact that the individual matter fields, ψ_i, do not necessarily all couple to the same Jordan-frame metric. The first restriction we make is to demand that the model obey the WEP[5] (**Figure 1**), which restricts the couplings to be the same: $A_i^2(\phi) \equiv A^2(\phi)$. Models with $A^2(\phi) = 1$ satisfy the SEP and, hence, are models of dark energy. In these models, the scalar field, ϕ, affects the motion of matter only through its gravitational influence; it is a decoupled source of stress energy. Cases where $A^2(\phi)$ is some nontrivial function are models of modified gravity. In modified gravity models, the force mediated by ϕ does not affect all objects universally. As an extreme example, consider the motion of some diffuse object compared with that of a black hole. Due to the no-hair theorem (85), black holes carry no scalar charge and therefore feel no fifth force from ϕ. However, a more diffuse object, such as a star or planet, will feel such a fifth force, leading to a large violation of the SEP.

Note that although the theories we consider satisfy the WEP at the microscopic level (all matter couples to the same metric), it is nevertheless possible to have apparent violations of the WEP for macroscopic objects (86, 87). This occurs in theories with screening mechanisms; whether or not the mechanisms operate depends mainly on the masses of the objects involved, so large-mass objects can fall at a different rate than small-mass objects. This is essentially because there is not necessarily a relationship between the scalar charge carried by objects and their gravitational potential, as there is in scalar–tensor theories without screening mechanisms, which generally exhibit violations of the equivalence principle only when $\Phi_{\rm N}$ is large (86).

Although the distinction based on the SEP is theoretically clean and satisfying, in practice it is not very useful phenomenologically. A more pragmatic distinction (relevant for the tests discussed below) is to call anything which has a fifth force modified gravity, and anything else, dark energy.

3. COSMOLOGICAL PHENOMENOLOGY

In order to describe the phenomenology of dark energy and modified gravity models, it is convenient to first consider the unperturbed universe, described by the scale factor $a(t)$, before turning

[5]Interactions in the dark sector—that is, a coupling between dark matter and dark energy—are a further scenario to consider. Because dark energy cannot strongly interact with visible matter, this approach violates the WEP. We do not consider this possibility in depth, but see Reference 84 for a review of the phenomenology.

to the equations governing the evolution of cosmological perturbations. The phenomenology of the latter is much richer, and more likely to yield insights distinguishing modified gravity and dark energy.

3.1. Background Expansion

In GR, applying the Einstein equations to the Friedmann–Robertson–Walker (FRW) metric yields the Friedmann equations

$$H^2 = \frac{1}{3M_{\rm Pl}^2}\sum_i \rho_i = \frac{1}{3M_{\rm Pl}^2}\left[\overline{\rho}_m + \overline{\rho}_{\rm DE}\right],$$
$$\dot{H} + H^2 = -\frac{2}{3M_{\rm Pl}^2}\sum_i [\bar{\rho}_i + 3\bar{p}_i] = -\frac{2}{3M_{\rm Pl}^2}[\overline{\rho}_m + (1+3w_{\rm DE})\overline{\rho}_{\rm DE}]. \quad 23.$$

Here, we have specialized to the case of a spatially flat universe and two stress-energy components, nonrelativistic matter [dominated by cold dark matter (CDM)] and dark energy with equation of state $w(t)$ (note that w is time dependent), which is the relevant case in the late Universe.

Modified gravity models lead to correspondingly modified Friedmann equations when applied to a homogeneous cosmology. As an example, consider the DGP model (see Section 2.2.3), where the first Friedmann equation is modified to

$$H^2 \pm \frac{H}{r_c} = \frac{1}{3M_{\rm Pl}^2}\overline{\rho}_m \quad 24.$$

in the absence of dark energy. The negative sign corresponds to the self-accelerating branch, which admits a de Sitter solution $H =$ constant at late times. However, this expansion history is easily mimicked in GR by a suitably chosen quintessence potential that yields $\overline{\rho}_{\rm DE}(t)/3M_{\rm Pl}^2 = H(t)/r_c$. This clearly illustrates that, while constraints on the background expansion are crucial in order to interpret LSS observables, they do not allow us to cleanly distinguish between modified gravity and dark energy.

3.2. Structure Formation with Quintessence

We now turn to the growth of structure in the context of a quintessence dark energy model, that is, a canonical light scalar field with an effective sound speed of one. The CC, and thus lambda cold dark matter (ΛCDM), is included here as a limiting case. Current constraints already imply that, for $z \lesssim 1$, the equation of state has to be close to -1. In these simplest dark energy models, density and pressure perturbations of the dark energy component are of order $(1+w)\Phi$, where typical cosmological potential perturbations are of order $\Phi \sim 10^{-4}$. Furthermore, anisotropic stress is negligible in these models. Thus, dark energy perturbations have a very small effect on LSS.

We continue to assume a spatially flat background and work in the conformal-Newtonian gauge, so that the metric is given by

$$ds^2 = -(1+2\Psi)dt^2 + 2a(t)B_i dt dx^i + a^2(t)[(1-2\Phi)\delta_{ij} + \gamma_{ij}]dx^i + dx^j. \quad 25.$$

Assuming that the scalar potential perturbations Ψ and Φ are much smaller than one, we can work to linear order in them. This is accurate almost everywhere in the universe. The potential Ψ governs the dynamics of nonrelativistic objects, whereas the combination $\Phi_\gamma \equiv (\Phi+\Psi)/2$ determines the null geodesics, namely light propagation. One way to see this is to note that null

geodesics are conformally invariant. Then, consider a conformal transformation $g_{\mu\nu} \mapsto e^{2\omega} g_{\mu\nu}$. At linear order, this transformation corresponds to $\{\Psi, \Phi\} \mapsto \{\Psi + \omega, \Phi - \omega\}$, so that Φ_γ is the unique linear combination of potentials that is conformally invariant.

Subhorizon scales $k \gg aH$: most LSS measurements are on these scales; in this limit, time derivatives are of order H, such that $\dot{\Phi} \sim H\Phi$, can be neglected in comparison to spatial derivatives

In Equation 25, B_i is a transverse vector capturing the vector modes, whereas γ_{ij} is a transverse trace-free tensor corresponding to gravitational waves (tensor modes). We neglect the vector modes throughout, as they decay and do not propagate. We briefly discuss tensor modes in Section 3.4.3.

In GR, the difference between the potentials, $\Phi - \Psi$, is sourced by the anisotropic stress, namely the trace-free part of T_{ij}. In the models considered in this section, this part is negligible, so we can set $\Phi = \Psi$ (see Section 3.3 for generalizations). On subhorizon scales $k \gg aH$, the 00 component of the Einstein equations reduces to the Poisson equation

$$\nabla^2 \Phi = 4\pi G \overline{\rho}_m a^2 \delta, \quad \text{or} \quad \nabla^2 \Phi = \frac{3}{2} \Omega_m(a) (aH)^2 \delta, \tag{26.}$$

where $\delta = \rho/\overline{\rho}_m - 1$ is the fractional matter overdensity, and $\Omega_m(a) = \overline{\rho}_m(a)/3H^2 M_{\rm Pl}^2$. For later reference, we define the solution to Equation 26 as Newtonian potential $\Phi_{\rm N}$. The continuity and Euler equations for the collisionless dark matter fluid are

$$\dot{\delta} + \frac{1}{a} \partial_k [(1+\delta) v^k] = 0,$$
$$\dot{v}^i + H v^i + \frac{1}{a} v^k \partial_k v^i = -\frac{1}{a} \partial^i \Psi. \tag{27.}$$

Here, $\boldsymbol{v} = a\,\mathrm{d}\boldsymbol{x}/\mathrm{d}t$ is the fluid peculiar velocity. On large scales, for Fourier modes $k \lesssim 0.05\, h\,\mathrm{Mpc}^{-1}$, both the density perturbation δ and the peculiar velocities $|\boldsymbol{v}|$ are much less than one, and we can linearize these equations. They can then be combined to yield the linear growth equation

$$\ddot{\delta}(\boldsymbol{k}, t) + 2H\dot{\delta}(\boldsymbol{k}, t) = -k^2 \Psi(\boldsymbol{k}, t) = \frac{3}{2} \Omega_m(a) H^2 \delta(\boldsymbol{k}, t), \tag{28.}$$

while the velocities obey $\boldsymbol{v} = -ia\boldsymbol{k}/k^2 \dot{\delta}$. Thus, individual Fourier modes of the density field evolve independently and equally on different scales. Furthermore, the evolution of density and velocity (given initial conditions) is completely controlled by the expansion history $H(t)$ of the universe. Thus, quintessence predicts a definite relation between the observed expansion history and growth of structure, allowing for consistency tests (see Section 4.2). Generally, for fixed initial conditions and an approximately constant equation of state w, the amount of late-time structure monotonically decreases with increasing w (**Figure 2**), as the accelerated expansion sets in earlier for less-negative w.

The same conclusions remain valid when considering the full Euler–Poisson system (Equations 26–27) and, indeed, the exact Vlasov (collisionless Boltzmann) equation for dark matter. The only impact of quintessence dark energy is through the background expansion history.

The nonlinear regime of LSS is frequently described by the halo model reviewed in Reference 88, which is motivated by the fact that a significant fraction of the dark matter resides in self-gravitating collapsed structures called halos. Moreover, most observed LSS tracers, such as galaxies and clusters, are found in these halos. In contrast to the complicated processes governing galaxy formation, collisionless N-body simulations are able to accurately predict abundance and clustering of halos.

The mass function of halos—that is, their number density in a logarithmic mass interval—is described to within ∼5% [in the context of ΛCDM (90)] by the universal form

$$\frac{\mathrm{d}n(M, Z)}{\mathrm{d}\ln M} = \frac{\overline{\rho}_m}{M} f(\nu) \left| \frac{\mathrm{d}\ln \sigma(M, z)}{\mathrm{d}\ln M} \right|, \quad \nu \equiv \frac{\delta_c(M)}{\sigma(M)}, \tag{29.}$$

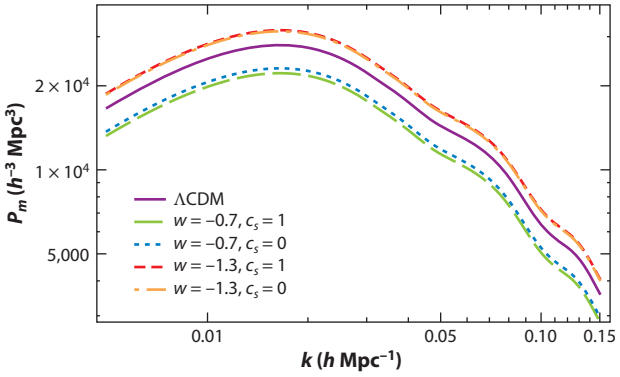

Figure 2

Linear matter power spectrum in quintessence dark energy with different values of equation of state w and speed of sound c_s. For the values of w shown, clustering dark energy with vanishing speed of sound modifies the matter power spectrum by a few percent. It can be probed by weak lensing measurements of the cosmic microwave background and galaxies. Modified from Reference 96.

where $\sigma(M, z)$ is the variance of the linear density field at redshift z when smoothed with a top-hat filter of radius $R(M)$, which contains the mass M at the background density $\overline{\rho}_m$. $f(\nu)$ is in general a free function, and δ_c is the initial overdensity of a spherical top-hat overdensity that collapses (i.e., reaches radius $R_{\rm th} = 0$) at a given redshift z, extrapolated from the initial time to redshift z through linear growth. For quintessence dark energy, the top-hat radius obeys (see, e.g., Reference 89).

$$\frac{\ddot{R}_{\rm th}}{R_{\rm th}} = -\frac{4\pi G}{3}\left[\overline{\rho}_m + (1+3w)\overline{\rho}_{\rm DE}\right] - \frac{4\pi G}{3}\delta\rho_m, \qquad 30.$$

where $\delta\rho_m = \rho_m(<R_{\rm th}) - \overline{\rho}_m$ is the overdensity in the interior of the top hat. Equation 29 is inspired by the excursion set approach (91), and the well-known Press–Schechter mass function (92) is a special case. Massive halos correspond to small variances $\sigma(M)$ and, hence, to high ν. In the high-ν limit, $f(\nu)$ asymptotes to $\exp(-q\,\nu^2)$, $q = \mathcal{O}(1)$, corresponding to exponentially rare high-mass halos. The abundance of halos depends on the growth history through both $\sigma(M)$ and δ_c, although the dependence of δ_c on cosmology is quite weak. The ansatz (Equation 29) is not guaranteed to be correct, even for a free function $f(\nu)$. Nevertheless, it is accurate at the $\sim 10\%$ level for ΛCDM (93).

The large-scale distribution of halos, described statistically by their N-point functions, can also be derived in this picture. The most important statistic on large scales is the two-point function or power spectrum $P(k)$. On scales where linear perturbation theory applies, the halo two-point function is related to that of matter by $P_h(k, z) = b_1^2(z)P_m(k, z)+C$, where C is a constant corresponding to white noise, and the linear bias parameter b_1 can be derived from the halo mass function (Equation 29). b_1 describes the response of the abundance of halos to a long-wavelength density perturbation δ_l or, equivalently, a change in the background density $\overline{\rho}_m \mapsto \overline{\rho}_m(1+\delta_l)$. This corresponds to lowering the threshold $\delta_c \mapsto \delta_c - \delta_l$, so that the linear halo bias becomes (94)

$$b_1(M, z) = \frac{\overline{\rho}_m}{n(M, z)}\frac{\partial n(M, z)}{\partial \overline{\rho}_m} = -\frac{1}{\sigma(M)f(\nu)}\frac{{\rm d}f(\nu)}{{\rm d}\nu}. \qquad 31.$$

Finally, assuming a universal density profile for halos $\rho(r|M, z)$, such as the NFW profile $\rho(r|M, z) \propto [r(1 + r/r_s)^2]^{-1}$ (95), the halo model provides a description of the statistics of the matter density field on nonlinear scales.

3.3. Beyond Quintessence Dark Energy

We now consider three relevant generalizations of the quintessence dark energy case considered in the previous section: small speed of sound $c_s = (\delta p/\delta\rho)^{1/2} \ll 1$, anisotropic stress, and an interacting dark sector. Dark energy density perturbations are negligible on scales that are much smaller than the sound horizon of the dark energy: $R_{s,\mathrm{DE}} \simeq c_s H^{-1}$ in physical units. Let us now consider the opposite case, namely what happens on scales much greater than $R_{s,\mathrm{DE}}$. On those scales, the pressure perturbations of the dark energy component are negligible compared with its density perturbations. This means that the dark energy fluid moves on the same trajectories (geodesics) as the pressureless CDM component. Thus, density perturbations increase on small scales in both components, although the dark energy perturbations are suppressed by $1 + w$ (on intermediate scales, $c_s k/aH \sim 1$, dark energy pressure perturbations need to be taken into account). Specifically, the Euler–Poisson system is augmented by the continuity equation for the dark energy component and becomes

$$\dot{\delta}_{\mathrm{DE}} - 3Hw\,\delta_{\mathrm{DE}} + \frac{1}{a}\partial_k[(1+w+\delta)v^k] = 0; \qquad \nabla^2\Phi = 4\pi G(\overline{\rho}_m\delta + \overline{\rho}_{\mathrm{DE}}\delta_{\mathrm{DE}}).$$

It is also possible to extend the spherical collapse Equation 30 to this case (96):

$$\frac{\ddot{R}_{\mathrm{th}}}{R_{\mathrm{th}}} = -\frac{4\pi G}{3}\left[\overline{\rho}_m + (1+3w)\overline{\rho}_{\mathrm{DE}}\right] - \frac{4\pi G}{3}(\delta\rho + \delta\rho_{\mathrm{DE}}), \tag{32.}$$

$$\dot{\rho}_{\mathrm{DE}}(<R_{\mathrm{th}}) = -3\frac{\dot{R}_{\mathrm{th}}}{R_{\mathrm{th}}}\left[\rho_{\mathrm{DE}}(<R_{\mathrm{th}}) + w\overline{\rho}_{\mathrm{DE}}\right]. \tag{33.}$$

Again, we see that if $1 + w \ll 1$, then initially small perturbations in the dark energy density will stay small. **Figure 2** illustrates the effects of $c_s = 0$ dark energy for the linear matter power spectrum. For $w = -0.9$, the effect on the matter power spectrum is of order 1% at $z = 0$ (and smaller at higher redshifts). By contrast, the halo abundance, calculated using Equation 29 with the modified $\sigma(M)$ and δ_c, is larger by 5–10% at the very high mass end for $c_s = 0$ as compared with $c_s = 1$ (96). The quantitative effect of a small dark energy sound speed can be increased if there is a significant "early dark energy" component (97).

We now turn to the case of a dark energy component with scalar anisotropic stress $\Pi_{ij}^{\mathrm{DE}} \equiv T_{ij}^{\mathrm{DE}} - \delta_{ij}T_{kl}^{\mathrm{DE}}\delta^{kl}/3$. Π_{ij}^{DE} sources a difference between the two metric potentials following

$$\left[\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right](\Phi - \Psi) = 8\pi G a^2\,\Pi_{ij}^{\mathrm{DE}}. \tag{34.}$$

The fact that $\Phi \neq \Psi$ is relevant to the cause of distinguishing between modified gravity and dark energy, as the difference between Φ and Ψ is otherwise a distinctive feature of modified gravity (see the following section). Dark energy models that involve a single scalar field do not lead to anisotropic stress, whereas models involving vector fields generally do (98, 99). Koivisto & Mota (100) and Mota et al. (101) investigated the impact on structure formation of dark energy with anisotropic stress on the basis of the effective parameterization by Hu (102) via a viscous parameter c_{vis}. The authors showed that the effect of dark energy anisotropic stress observable via $\Phi - \Psi$ is significant only for horizon-scale perturbations. By contrast, modified gravity generally leads to a significant $\Phi - \Psi$ on all scales, so that one can still hope to effectively distinguish between modified gravity and dark energy via this probe.

Finally, one can also couple the dark energy component to the nonbaryonic components, such as CDM or neutrinos (103, 104). Following our distinction in Section 2.3, these are not modified gravity models, given that the fifth force does not obey the WEP. The absence of coupling to baryons and radiation lets these models evade local tests, so the effects on the LSS can become

quite significant (105). Interestingly, a generic signature of this model is a different clustering amplitude (bias) of baryons relative to the dark matter on large scales, which can be used to place constraints on this type of model.

Parameterized post-Friedmann (PPF) approach: on linear scales and for adiabatic initial conditions, any theory of gravity can be parameterized by two free functions of scale and time (Equation 36), with an additional constraint on superhorizon scales (Equation 35)

3.4. Structure Formation in Modified Gravity

In this section, we briefly review the rich phenomenology of structure growth in modified gravity, separately for the linear and nonlinear regimes.

3.4.1. Scalar perturbations on linear scales. We begin with structure formation in modified gravity theories on sufficiently large scales, where linear perturbation theory applies, and assume that stress-energy perturbations are only due to a CDM component. It is convenient to work in Fourier space, where individual modes evolve independently at linear order. Then, given that space-time is described by two potentials, $\Phi(\boldsymbol{k}, t)$ and $\Psi(\boldsymbol{k}, t)$, and the matter sector is characterized completely by $\delta(\boldsymbol{k}, t)$ (the velocity is determined via the continuity equation), any theory of gravity can be described in this regime by two free functions of $\boldsymbol{k}$ and t, which parameterize the relation between the three fields Φ, Ψ, and δ.

On superhorizon scales $k \ll aH$ and for adiabatic perturbations, further constraints are posed by diffeomorphism invariance (106–108). Specifically, a superhorizon adiabatic mode behaves as a separate, curved universe, so that its evolution is constrained by the solution to the background expansion in the given theory. Bertschinger (106) derived

$$\ddot{\Phi} - \frac{\ddot{H}}{\dot{H}}\dot{\Phi} + H\dot{\Psi} + \left(2\frac{\dot{H}}{H} - \frac{\ddot{H}}{\dot{H}}\right)H\Psi = 0, \qquad 35.$$

which, given one function of time specifying the relation between Φ and Ψ for $k \ll aH$, determines the superhorizon evolution of the potentials.

We now turn again to the subhorizon limit and assume adiabatic initial conditions. In the absence of preferred directions, we can parameterize the relation between δ and Φ, Ψ via

$$\begin{aligned} -k^2\Psi(\boldsymbol{k}, t) &= \frac{3}{2}\Omega_m(a)(aH)^2\,\mu(k, t)\delta, \\ \frac{\Phi(\boldsymbol{k}, t)}{\Psi(\boldsymbol{k}, t)} &= \gamma(k, t). \end{aligned} \qquad 36.$$

These equations reduce to GR for $\mu = \gamma = 1$. Given adiabatic initial conditions and assuming the WEP, Φ, Ψ, and δ are, at linear order, all related to the initial conditions (i.e., a single stochastic variable) through transfer functions T_Φ, T_Ψ, and $T_\delta(k, t)$, so that $\mu \sim (k/aH)^2 T_\Psi/T_\delta$ and $\gamma = T_\Phi/T_\Psi$. Parameterizations based on Equations 35 and 36 are referred to as the PPF approach. Moreover, for a local 4D theory of gravity, T_i are functions of k^2 only, so that μ and γ reduce to rational functions of k^2. If we further assume that no higher than second derivatives appear in the equations for Φ and Ψ, then μ and γ are completely described by five functions of time only (109):

$$\mu(k, t) = \frac{1 + p_3(t)k^2}{p_4(t) + p_5(t)k^2}, \quad \gamma(k, t) = \frac{p_1(t) + p_2(t)k^2}{1 + p_3(t)k^2}. \qquad 37.$$

Note that higher powers of k can appear in the equations for Φ and Ψ after the additional degrees of freedom are integrated out, even if the fundamental equations are all second order in derivatives. Using suitable interpolation, one can connect these subhorizon limit results to superhorizon scales in a parameterization that enforces Equation 35 (107, 110, 111), further reducing the number

of free functions. Finally, Gleyzes et al. (23) and Bellini & Sawicki (24) found that four free functions of time, along with a free background expansion history, completely describe the linear growth of perturbations for the Horndeski Lagrangian (Equation 17). While different equivalent parameterizations exist, the specific choice of μ and γ adopted here is motivated by the fact that μ can be directly constrained by the growth of structure, as Ψ is the potential governing the motion of matter. From Equation 28, the growth equation is modified to

$$\ddot{\delta}(\boldsymbol{k}, t) + 2H\,\dot{\delta}(\boldsymbol{k}, t) = \frac{3}{2}\Omega_m(a)H^2\mu(k, t)\delta(\boldsymbol{k}, t). \qquad 38.$$

Specifically, in modified gravity, the linear growth is in general scale (k) dependent, unlike the quintessence case or the case for dark energy with $c_s = 0$. Measuring the growth history as a function of k and t in principle enables a measurement of $\mu(k, t)$.

By contrast, γ, also referred to as gravitational slip, quantifies the departure between the two space-time potentials. The propagation of photons is, in the coordinate frame of Equation 25, governed by the combination of potentials

$$\Phi_\gamma = \frac{1}{2}(\Psi + \Phi) = \frac{1}{2}(1 + \gamma)\Psi. \qquad 39.$$

This is in exact analogy to the $\gamma_{\rm PPN}$ parameter in the parameterized post-Newtonian (PPN) framework (27), which is similarly constrained by combining photon propagation (Shapiro delay) with dynamics (the Earth's orbit). The fact that Solar System tests constrain $|\gamma_{\rm PPN} - 1| \lesssim 10^{-5}$ clearly shows that a consistent modification of gravity in cosmology has to be scale or environment dependent. Gravitational lensing observables in cosmology are then approximately related to the matter density through the combination $(1 + \gamma)\mu$ of the modified gravity parameterization (see Section 4).

3.4.2. Scalar perturbations on nonlinear scales. The description of nonlinear structure formation in the context of modified gravity is complicated by the necessity of screening mechanisms. In other words, although we can in principle extend the parameterization of Equation 36 to small, nonlinear scales, that will in general violate Solar System constraints on gravity, which tightly constrain μ and γ. The most interesting modified gravity models can circumvent these constraints by employing nonlinear screening mechanisms. By definition, they are not captured by a linear parameterization of the form of Equation 36. For this reason, cosmological constraints on modified gravity from nonlinear scales are generically model dependent. In this context, it is useful to classify models by their screening mechanism (see Section 2.2.4).

In models of the chameleon or symmetron/dilaton type, a certain depth of the gravitational potential is necessary to pull the field away from its background value $\bar{\varphi}$ and activate the screening mechanism. Specifically, if the Newtonian potential $\Phi_{\rm N}$, which is the solution of Equation 26, satisfies $-\Phi_{\rm N} \gtrsim C\bar{\varphi}/M_{\rm Pl}$, then φ becomes locally suppressed compared with $\bar{\varphi}$, and the deep potential region is screened. The constant C depends on the model but is typically of order unity; for $f(R)$ gravity, $C = 3/2$. In order for the Solar System to be screened, $C\bar{\varphi}/M_{\rm Pl} \lesssim 10^{-5}$, which can be shown to place an upper limit on the mass $\bar{m}_\varphi \gg H$ of the field in the background (112). The result is that viable models with this screening mechanism modify structure formation only on scales below ~30 Mpc.

This screening behavior further means that there will be a characteristic threshold halo mass $M_{\rm scr}(\bar{\varphi})$; halos above this mass will be screened, whereas halos below this threshold will be unscreened (because halo profiles are nearly universal, there is a well-defined mapping between central potential and halo mass). This transition effect can be clearly observed in N-body simulations of chameleon-type models; **Figure 3*a*** shows the gravitational acceleration within dark

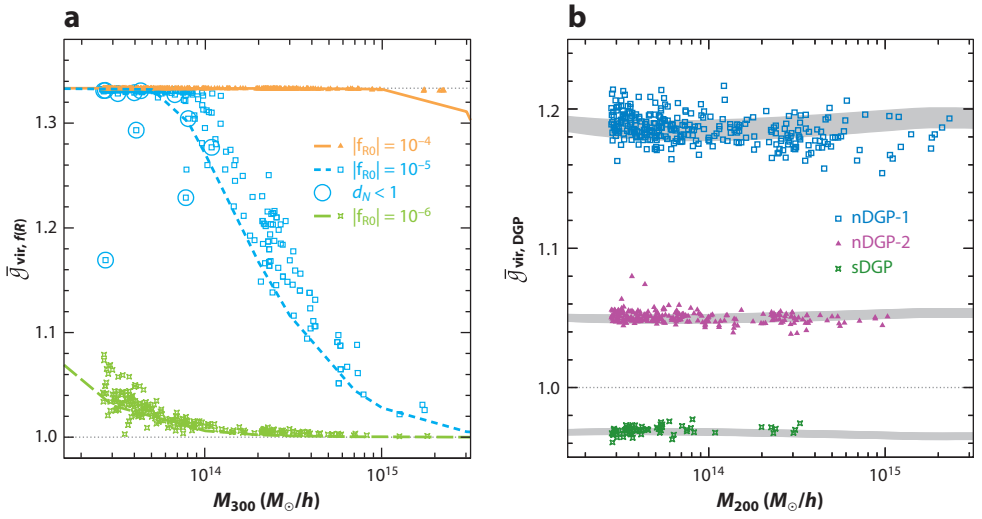

Figure 3

Mean mass-weighted gravitational acceleration (relative to general relativity) $\bar{g}$ within dark matter halos, as measured in (*a*) modified gravity simulations for $f(R)$, corresponding to chameleon screening, and (*b*) the Dvali–Gabadadze–Porrati model (DGP), showing Vainshtein screening. The lines show the expectations based on a simple spherical model of the halos (114). The circled points show halos that are screened due to the potential of a massive halo in the vicinity, rather than their own potential well. The ratio of dynamical to lensing mass of these halos is given by $M_{\mathrm{D}}/M_{\mathrm{L}} = \bar{g}^{3/5}$. It can be probed by comparing dynamical mass estimates of galaxy clusters with their gravitational lensing signal. The $f(R)$ model is of the Hu–Sawicki type (Equation 16) (38). The sDGP model corresponds to the self-accelerating branch where gravity is weakened, whereas the nDGP models show normal-branch DGP with a quintessence component added to produce an expansion history identical to ΛCDM (with $r_c = 500$ or 3,000 Mpc for nDGP-1 or nDGP-2, respectively) (115). Modified from Reference 114.

matter halos in $f(R)$ simulations, which consistently include the chameleon mechanism (113). Halos with a potential depth larger than the field value $-\Phi_{\mathrm{N}} \gtrsim 2|f_{R0}|/3$ become screened, whereas lower-mass halos are unscreened. Because unscreened halos accrete mass at a rate higher than in GR, this transition effect is also manifested in the halo mass function (89). The halo of our own Galaxy has to be screened in order for a chameleon/symmetron model to satisfy Solar System tests, so these models can be expected to show modified gravity effects only in lower-mass halos. This effect motivates the search for modified gravity effects in nearby dwarf galaxies (see Section 4.1.3).

We now turn to the Vainshtein screening mechanism. For simplicity, we restrict the discussion to the cubic galileon interaction term (Equation 18). In the case of spherical symmetry, the equation of motion in the subhorizon limit can be integrated once to yield

$$\frac{\mathrm{d}\varphi}{\mathrm{d}r} \propto \frac{M(<r)}{M_{\mathrm{Pl}}\, r^2}\, g\left(\frac{r}{r_*(r)}\right), \quad \text{where} \quad g(\xi) = \xi^3\left(\sqrt{1+\xi^{-3}} - 1\right). \qquad 40.$$

Here, $M(<r)$ is the mass (over the background density) enclosed within r. The scale $r_*(r) \propto [M(<r)/M_{\mathrm{Pl}}\Lambda^3]^{1/3}$ is the r-dependent Vainshtein radius (this scaling also holds for higher-order galileon interactions). In the case of the DGP model, the scale is given by

$$r_*(r) = \left[\frac{16M(<r)r_c^2}{9M_{\mathrm{Pl}}^2\beta^2}\right]^{1/3}, \quad \beta = 1 \pm 2Hr_c\left[1 + \frac{\dot{H}}{3H^2}\right]. \qquad 41.$$

We see that $d\varphi/dr$ is suppressed in comparison to the Newtonian gradient $M_{Pl}\, d\Phi_N/dr$ for $r/r_* \ll 1$, where the quantity $(r/r_*)^3$ is directly proportional to the inverse of the interior density $\rho^{-1}(<r)$. Specifically, Vainshtein screening occurs at fixed interior density, in contrast to the chameleon type, which occurs at fixed mass. The threshold density for screening is of order the background matter density today for natural model parameters (in the case of DGP, $r_c \sim H_0^{-1}$), which for nonlinear LSS leads to a qualitatively different behavior compared with the chameleon case (**Figure 3*b***): Halos of all masses are screened within a fixed fraction of their scale radius. Thus, Solar System constraints do not force us to look to certain types of objects for interesting signatures of Vainshtein-type models.

CMB: cosmic microwave background

BAO: baryon acoustic oscillations

3.4.3. Tensor perturbations. In addition to modifying the scalar perturbations, modified gravity generally alters the propagation of tensor modes. More specifically, a running of the gravitational coupling causes a change in the decay of the gravitational wave amplitude in an expanding universe. Furthermore, the G_4 and G_5 terms of a Horndeski scalar–tensor theory alter the propagation speed of tensor modes (23, 116).

4. COSMOLOGICAL TESTS

In this section, we provide an overview of the observables and experimental methods that are used to probe gravity and dark energy on cosmological scales.

4.1. Observables

Modified gravity and dynamical dark energy models in general change the background expansion history of our Universe with respect to ΛCDM. The predicted deviation can be tested by measuring distances to astronomical objects and cosmic "standard rulers." We refer to these measurements as geometric probes. For instance, a comparison between the magnitudes of high-redshift and low-redshift Type Ia supernovae, serving as standard candles, yields a measurement of the evolution of the luminosity distance $d_L(z) = (1+z)r(z)$, where $r(z) \equiv (1+z)d_A(z)$ is the comoving angular diameter distance. For a spatially flat universe, $r(z)$ is simply the comoving radial distance $\chi(z) = \int_0^z dz'/H(z')$. This is a relative distance measurement, although calibration with the low-redshift "distance ladder" also yields an absolute distance measurement. Observations of Cepheid stars in supernova host galaxies yield $H_0 = (73.8 \pm 2.4)$ km s^{-1} Mpc^{-1} (117) for the Hubble constant today. Measurements of the acoustic peaks in the cosmic microwave background (CMB), and their imprint on LSS—the baryon acoustic oscillations (BAO) observed in galaxy clustering—serve as a standard ruler, and provide complementary information about the absolute distance scale to $z \gtrsim 0.1$. These geometric probes, together with recent Planck measurements, constrain a constant dark energy equation of state in a spatially flat universe at $w = -1.006 \pm 0.045$ (118), consistent with the CC in ΛCDM.

When dark energy is evolving or gravity is modified, it leaves imprints in the formation of the LSS, as shown by Equation 28. Modified gravity models can match the ΛCDM expansion history at an observationally indistinguishable level, whereas the growth of cosmic structure may still be significantly modified, rendering the latter a vital probe for modified gravity. Nonetheless, geometric probes provide important constraints on cosmological parameters, which limit degeneracies in the growth of structure constraints. In the following subsections, we review some of the important observables of cosmological structures that are used to test modified gravity and dark energy. **Figure 4** summarizes some of the cosmological constraints that have been inferred on $f(R)$ gravity, which serves here as a representative for typical modified gravity models that has

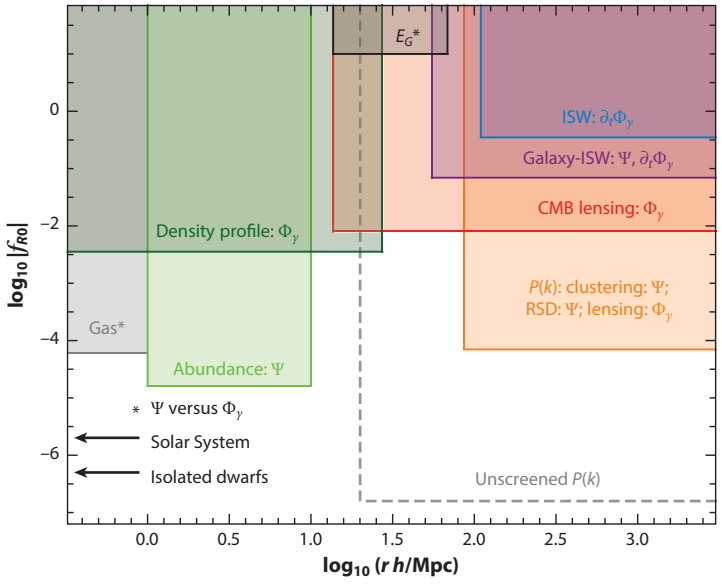

Figure 4

Current and forecasted constraints on $f(R)$ gravity, in particular the parameter f_{R0} of the Hu–Sawicki model (Equation 16), and potentials tested. Linear constraints: integrated Sachs–Wolfe (ISW) and galaxy–ISW cross-correlations (120); cosmic microwave background (CMB) lensing (121), the E_G probe (122), and power spectrum constraints from galaxy clustering; redshift-space distortions (RSD); and weak gravitational lensing (WL) (123). Nonlinear constraints: cluster abundance (120, 124, 125), cluster density profiles (119), and the comparison of cluster gas and weak lensing measurements (126). Also indicated are the currently tightest constraints on $|f_{R0}|$ from Solar System tests (38) and distance indicators in isolated dwarf galaxies (127). Future cosmological surveys measuring the power spectrum $P(k < 0.3\,h/\mathrm{Mpc})$ in unscreened regions at the 1% level will be able to outperform the astrophysical constraints (128). Modified from Reference 119.

also been particularly well tested. We show the range of scales covered by different cosmological observables and indicate which of the metric potentials in Equation 25 they probe. **Table 1** provides a summary of the current observational constraints on the modified gravity models discussed in Section 2.

4.1.1. Cosmic microwave background. The CMB radiation is the primary observable for obtaining information about our cosmos. Of particular importance for dark energy and modified

Table 1 Current observational constraints on modified gravity models

Screening mechanism	Representative model	Linear cosmology	Nonlinear cosmology	Submegaparsec
Chameleon	$f(R)$ gravity[a]	$\lvert f_{R0}\rvert \lesssim 10^{-4}$	$\lvert f_{R0}\rvert \lesssim 10^{-5}$	$\lvert f_{R0}\rvert \lesssim (10^{-7}$–$10^{-6})$
Symmetron	Equation 13[b]	None	None	$\chi \lesssim (10^{-7}$–$10^{-6})$
Vainshtein	nDGP[c]	$H_0 r_c \gtrsim (0.1$–$1)$	$H_0 r_c \gtrsim 0.1$	$H_0 r_c \gtrsim 10^{-4}$
Vainshtein	Cubic galileon[d]	Incompatible ISW	Incompatible voids	Compatible

[a]From References 38, 123, and 125–127.
[b]From References 30 and 127.
[c]From References 6 and 149–152.
[d]From References 47, 153, and 154. Abbreviations: ISW, integrated Sachs–Wolfe effect; nDGP, normal branch Dvali–Gabadadze–Porrati model.

gravity are constraints on the amplitude and adiabaticity of initial perturbations. However, late-time modifications manifest themselves in the CMB temperature and polarization only via secondary anisotropies. The presence of dark energy or a CC gives rise to the integrated Sachs–Wolfe (ISW) effect (129), a fluctuation in the temperature field at the largest scales due to the change in energy of the CMB photons $\propto \dot{\Phi}_\gamma$ when traversing the evolving metric potentials. This effect is modified in alternative theories of gravity or dynamical dark energy, as shown by Equation 36. A related secondary anisotropy in the CMB is due to the depths of the potentials Φ_γ, which determine the gravitational lensing of the CMB photons by the cosmological structure between the observer and the last-scattering surface. The directions of the observed CMB photons are displaced by the lensing deflection angle $\boldsymbol{d}$. The effect is often described by the convergence field $\kappa = \nabla_{\hat{\boldsymbol{\theta}}} \cdot \boldsymbol{d}$, where $\nabla_{\hat{\boldsymbol{\theta}}}$ is the gradient operator on the sphere. On smaller angular scales, the Sunyaev–Zel'dovich (SZ) effect, caused by the energy gain of CMB photons by collisions with electrons in the postreionization universe, provides a complementary signature that probes gravitationally induced motion of matter and, hence, the dynamical potential Ψ rather than the lensing potential Φ_γ. The fairly small kinetic SZ effect (130) measures the line-of-sight component of the large-scale electron momentum density. The larger thermal SZ effect (131) is induced by virialized structures. In order to extract these late-time signatures, it can also be useful to consider their cross-correlation with the foreground structure.

ISW effect: integrated Sachs–Wolfe effect

SZ effect: Sunyaev–Zel'dovich effect

WL: weak gravitational lensing

At linear order, the ISW temperature fluctuation, the lensing convergence, and the galaxy density perturbation are given as functions of the position on the sky $\hat{\boldsymbol{\theta}}$ by

$$\Delta T_{\rm ISW}(\hat{\boldsymbol{\theta}}) = -2\int_0^{\chi_*} \mathrm{d}\chi \frac{\mathrm{d}t}{\mathrm{d}\chi}\frac{\partial}{\partial t}\Phi_\gamma(\chi\hat{\boldsymbol{\theta}},\chi), \qquad 42.$$

$$\kappa(\hat{\boldsymbol{\theta}}) = \int_0^{\chi_*} \mathrm{d}\chi \frac{\chi_*-\chi}{\chi_*\chi}\nabla^2_{\hat{\boldsymbol{\theta}}}\Phi_\gamma(\chi\hat{\boldsymbol{\theta}},\chi), \qquad 43.$$

$$g(\hat{\boldsymbol{\theta}}) = \int_0^{\chi_*} \mathrm{d}\chi \frac{\mathrm{d}z}{\mathrm{d}\chi} b[z(\chi)]\Pi(z)\delta(\chi\hat{\boldsymbol{\theta}},\chi), \qquad 44.$$

respectively, where χ_* denotes the conformal radial distance to the last-scattering surface, the galaxy bias b is scale independent but dependent on redshift, and Π is a normalized selection function. For multipoles $\ell \gtrsim 10$, one can employ the subhorizon limit and Limber approximation to write the angular cross-power spectra of these quantities as

$$C_\ell^{XY} = \int \mathrm{d}z \frac{H(z)}{\chi^2(z)}\left[F_X(k,z)\,F_Y(k,z)\,P(k)\right]\Big|_{k=\ell/\chi(z)}, \quad \text{where} \qquad 45.$$

$$F_{\rm ISW}(k,z) \equiv T_{\rm CMB}\frac{3H_0^2\Omega_{m,0}}{k^2}\frac{\partial}{\partial z}G(k,z),$$

$$F_g(k,z) \equiv b(z)\Pi(z)D(k,z),$$

$$F_\kappa(k,z) \equiv \frac{3H_0^2\Omega_{m,0}}{2H(z)}\frac{\chi(\chi_*-\chi)}{\chi_*}G(k,z), \quad \text{and}$$

$$G(k,z) \equiv \frac{1}{2}[1+\gamma(k,z)]\mu(k,z)(1+z)D(k,z).$$

Here, we have used Equation 36, defined $D(z,k) \equiv \delta(z,k)/\delta(0,k)$ as the scale-dependent growth factor, and denoted $P(k)$ as the matter power spectrum today.

4.1.2. Large-scale structure. The bulk of cosmological information on dynamical dark energy and modified gravity is contained in the three-dimensional (3D) LSS at low redshifts $z \lesssim 5$. Galaxy shape correlations probe the lensing potential Φ_γ via WL, specifically the convergence κ (Equation 43, with χ_* replaced with the comoving distance to the source galaxies). Similarly, angular correlations of galaxies in redshift slices probe the matter power spectrum on large

scales (Equation 45). Furthermore, the 3D power spectrum of galaxies traces the underlying distribution of matter on large scales via

$$P_g(\boldsymbol{k}, z) = b^{2(z)} P(k, z) + 2b(z)\hat{k}_\parallel^2 P_{\delta\theta}(k, z) + \hat{k}_\parallel^4 P_{\theta\theta}(k, z), \quad 46.$$

RSD: redshift-space distortions

where $\hat{k}_\parallel$ is the cosine of $\boldsymbol{k}$ with the line of sight. The peculiar motion of galaxies leads to the terms involving the cross- and autocorrelation of the matter velocity divergence θ [redshift-space distortions (RSD) (132)], which can be used to measure the growth rate of structure $f \equiv d \ln D/d \ln a$. Note that the galaxy bias b cannot be predicted from first principles, and must be either constrained using lensing or marginalized over (see Section 4.5).

The statistics of matter and Φ_γ can be measured well into the nonlinear regime through galaxy clustering and WL, where Equations 45 and 46 are no longer correct. Nonlinear effects become important on scales of $k \gtrsim 0.1\,h\,\mathrm{Mpc}^{-1}$ (depending on redshift), and potentially even larger scales in modified gravity. Upcoming surveys will measure statistics of galaxy counts and lensing with high signal-to-noise ratios out to wave numbers $k > 1\,h\,\mathrm{Mpc}^{-1}$, so that the bulk of information is in the nonlinear regime. The interpretation of measurements in the nonlinear regime is complicated by nonlinear gravitational evolution, baryonic feedback effects, and screening in modified gravity. Because a fundamental description of the nonlinear structure for general dark energy and modified gravity theories is lacking, observational tests have therefore been limited to specific models.

A further, fully nonlinear probe is the abundance of galaxy clusters, governed by the halo abundance $\mathrm{d}/\mathrm{d} \ln M$ (Equation 29), which yields tight constraints on the allowed deviations from ΛCDM (**Figure 4**) (124); their density profiles have also been used to probe modified gravity (119). Clusters can be identified via their galaxy content, X-ray emission, or SZ signal. A crucial ingredient for cosmological constraints is the relation between observable and halo mass. This relation can be determined through gravitational lensing of background galaxies or the CMB, which yields the lensing mass $M_\mathrm{L}(<r) \propto r^2 \nabla \Phi_\gamma(r)$. The gas temperature T_gas and pressure P_gas profiles measured in X-rays and SZ (where SZ measures only P_gas) yield estimates of the dynamical mass $M_\mathrm{D}(<r) \propto r^2 \nabla \Psi(r)$. For this purpose, one assumes hydrostatic equilibrium, which in GR yields

$$\frac{1}{\rho_\mathrm{gas}} \frac{\mathrm{d}P_\mathrm{gas}(r)}{\mathrm{d}r} = -\frac{G\,M_\mathrm{D}(r)}{r^2}, \quad 47.$$

where for thermal pressure $P_\mathrm{gas} \propto \rho_\mathrm{gas} T_\mathrm{gas} \propto P_\mathrm{e}$. The dominant systematic uncertainty in M_D from these measurements is due to the poorly known nonthermal pressure components. Similarly, the velocity dispersion of galaxies within clusters probes M_D, although the interpretation is hampered by uncertainties in the velocity bias of galaxies. This uncertainty becomes smaller on larger scales, where galaxies exhibit coherent inflow motion onto clusters (133, 134).

As noted in Section 3.4, a difference between M_L and M_D is generic to modified gravity. Along with the screening effects discussed in Section 3.4.2, this difference clearly needs to be taken into account when constraining modified gravity using cluster abundance. The discrepancy can also be used as probe of modified gravity itself (see Section 4.5). The screening of modified gravity effects in dense regions motivates the use of abundance, clustering, and profiles of underdense regions (i.e., voids) as a modified gravity probe (86).

Finally, with the advent of gravitational wave astronomy, demonstrated with the first detection from aLIGO (135), and from aVIRGO and eLISA in the future, it will become feasible to test the modifications in the propagation of gravitational waves described in Section 3.4.3 (also see Reference 136). In particular, a direct comparison of the arrival times between gravitational waves and the electromagnetic emissions from a single event will place tight constraints on deviations between the propagation speed of tensor modes and the speed of light. Furthermore,

tight constraints on the G_4 and G_5 terms of Horndeski gravity (137) are placed by binary pulsar timing, as these are not screened by the Vainshtein mechanism (assuming no additional screening mechanism is active).

4.1.3. Submegaparsec scales. Although small scales lie beyond the scope of this review, we note that, in addition to the Solar System constraints, the tightest bounds on chameleon/symmetron modifications are currently inferred from stellar distance indicators (standard candles) in nearby dwarf galaxies residing in a low-density region of space (127). This test uses the combination of red giant stars, which are expected to be screened, with Cepheids, whose pulsating envelopes are still unscreened. The difference in screening introduces a systematic deviation between the distances inferred. These constraints depend crucially on a correct identification of unscreened environments for these stars, obtained from a reconstruction of the local density field.

4.2. Consistency Tests: Geometry and Growth

The most common method of obtaining cosmological constraints on dark energy is to combine various geometric probes of the expansion history and observations of the growth of structure to obtain joint constraints on the dark energy equation of state w. Because only quintessence unambiguously predicts the growth factor $D(a)$ once the expansion history is fixed, these constraints apply only to this type of dark energy. Moreover, as discussed in Section 2.1, a constant $w \neq -1$ is not necessarily a good assumption, so a more general parameterization (138, 139),

$$w(t) = w_0 + w_a[1 - a(t)], \qquad 48.$$

is often adopted. Any observational sign for $w \neq -1$ or $w_a \neq 0$ would be evidence against ΛCDM. In order to test for evidence beyond quintessence, Linder (140) proposed extending this parameterization by a parameter $\tilde{\gamma}$, motivated by the fact that the growth factor in a variety of quintessence models satisfies $f = d\ln D(a)/d\ln a = \Omega_m(a)^{\tilde{\gamma}}$, where $\tilde{\gamma} \approx 0.55$. Thus, finding $\tilde{\gamma} \neq 0.55$ would imply clustering dark energy, modified gravity, or an interacting dark sector (142). Mortonson et al. (141) and Ruiz & Huterer (143) devised more general consistency tests between geometry and growth (**Figure 5**).

Although the consistency tests are model independent, it is generally not true that findings consistent with ΛCDM rule out any modification of the concordance model, as these parameters do not encompass the entire space of possible dark energy and modified gravity models. Furthermore, it is not straightforward to turn a constraint on $\tilde{\gamma}$ into a constraint on a given modified gravity or

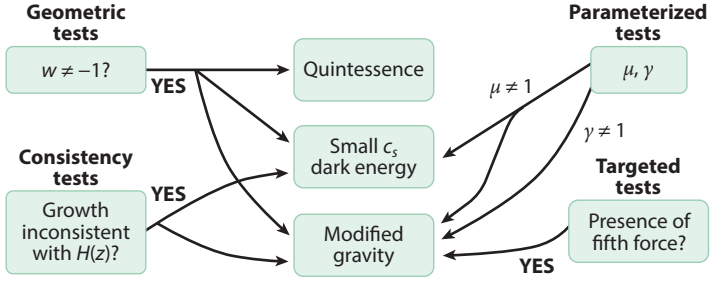

Figure 5

Summary of the various types of cosmological tests and their implications. Detection of deviations from ΛCDM phenomenology in these four classes of tests implies different physics beyond the standard cosmology (*arrows*).

non-quintessence-type dark energy model, as the parameterization is based on linear subhorizon scales; even if the constraint can be mapped onto a model parameter, it will generally not be an optimal constraint.

4.3. Parameterized Tests of Gravity

We can improve several of the disadvantages of the consistency tests discussed above by adopting the more physical parameterization introduced in Equation 36, which incorporates physical constraints such as mass and momentum conservation. Quintessence models correspond to $\mu = \gamma = 1$, while clustering dark energy can be captured by $\gamma = 1$ and $\mu \propto 1 + w$. Note however that this parameterization is still restricted to linear scales.

Without further constraints, this parameterization involves two free functions of both time and scale. Zhao et al. (144) performed a principal component analysis (PCA), for bins in k and z, of μ and γ to determine how well the data can constrain the eigenmodes of μ and γ along with redshift bins in $w(a)$. They found that the future LSST survey will be able to constrain the 10 best eigenmodes in μ and γ to ∼3% and ∼10%, respectively. Although PCA is useful to address the possible modifications in full generality, it may allow for too much freedom to detect a well-defined deviation predicted by a particular clustering dark energy or modified gravity model. As discussed in Section 3.4.1, assumptions such as, at most, second-order derivatives or a Horndeski Lagrangian allow us to significantly reduce the freedom in the parameterization, leading to, for example, Equation 37. One might further expect the modifications to scale as the effective dark energy density $\Omega_{\rm DE} = 1 - \Omega_{\rm K} - \Omega_m$, providing an even more restrictive parameterization of the five functions $p_{1-5}(t) = p_{i,0}\Omega_{\rm DE}(t)/\Omega_{\rm DE,0}$. Moreover, one could argue that if the gravitational modifications are to drive cosmic acceleration, then any scale appearing in the equations should be of order the Hubble scale today. Then, μ and γ are completely determined by one function of time each on subhorizon scales $k \gg aH$, which motivates the restriction of cosmological tests of μ and γ to the two constants $p_{3,0}/p_{5,0}$ and $p_{2,0}/p_{3,0}$. Note that this last restriction fails to capture $f(R)$ and other chameleon/symmetron models.

The main caveat to this type of parameterization is that it is limited to linear scales. Although more parameters can be introduced to model a screening-induced suppression of μ and γ on nonlinear scales, doing so clearly does not realistically model the intrinsically nonlinear screening effects. Finally, even on linear scales, constraints on μ and γ are weakened when allowing for other nonstandard ingredients, as discussed in the following section.

4.4. Cosmic Degeneracies

When comparing theoretical predictions of dark energy and modified gravity models against cosmological observations, it is important to discriminate the effects from other signatures of new fundamental physics and complex nonlinear processes. In particular, massive neutrinos or baryonic feedback effects can compensate for the effects of an enhanced growth of structure in the power spectrum or the halo mass function (145, 146). Similar effects can be produced by non-Gaussian initial conditions. It has also been pointed out that Horndeski scalar–tensor modifications are endowed with sufficient freedom to allow the recovery of the background expansion history and linear LSS of a ΛCDM universe (although these models require some fine-tuning) (147), which limits a fundamental discrimination between modified gravity and dark energy or a CC based on cosmological structure on large scales. Importantly, the modified propagation of tensor modes (Section 3.4.3) allows these models to be distinguished from ΛCDM.

4.5. Targeted Tests

E_G: combination of galaxy–lensing and galaxy–velocity cross-correlations that is sensitive to a fifth force while canceling the unknown galaxy bias (Equation 49)

In order to avoid known and unknown degeneracies, it is desirable to devise tests specifically targeting the distinction between dark energy and modified gravity that we have introduced in Section 2.3. More precisely, we would like to devise tests that constrain a universally coupled fifth force. The most promising such test uses the generically predicted difference between dynamical (Ψ) and lensing (Φ_γ) signatures. In the absence of a fifth force, this difference could be sourced only through a large anisotropic stress in the dark sector, which is difficult to obtain (Section 3.3).

On subhorizon linear scales, such a test can be performed by combining the cross-correlation between foreground galaxies and lensing, $P_{g(\nabla^2\Phi_\gamma)}$, and the galaxy density–velocity cross-correlation of the same population of galaxies, $P_{g\theta}$. The latter can be extracted from the 3D clustering (Equation 46). The ratio (148)

$$E_G \equiv \frac{P_{g(\nabla^2\Phi_\gamma)}}{P_{g\theta}} \simeq \frac{1}{2}(1+\gamma)\mu\frac{\Omega_m}{f} \qquad 49.$$

cancels the linear galaxy bias and isolates a combination of the μ and γ parameters that are induced by fifth forces, providing a robust test of gravity on linear scales. Note that f depends on μ as well.

On smaller, mildly nonlinear scales, another comparison between dynamics and lensing is possible through the use of the infall of galaxies onto massive clusters (133, 134). WL around clusters constrains the spherically averaged profile of Φ_γ. The two-dimensional cross-correlation between galaxies and clusters on scales ~5–20 Mpc, which measures a projection of the galaxy phase space, contains information about the infall velocity and thus Ψ. This probe goes beyond E_G in that it is not restricted to linear scales. However, effects such as a bias between galaxy velocities and those of the underlying matter distribution need to be carefully controlled.

On even smaller scales, one can compare the lensing mass $M_{\rm L}$ with the dynamical mass $M_{\rm D}$ of clusters (measured using X-rays or SZ; see Section 4.1.2) or galaxies (measured using stellar velocities). If γ is scale independent, and in the absence of screening, $M_{\rm D}/M_{\rm L} = (1+\gamma)^{3/5}$ for virialized objects (114). For viable models however, taking screening into account is essential (**Figure 3**) (114, 126).

5. OUTLOOK

We have provided a brief overview of different physical mechanisms to explain the accelerated expansion of the universe, and introduced theoretical and phenomenological distinctions between the two scenarios of dark energy and modified gravity. In the next 15 years, LSS and CMB surveys (AdvACT, eBOSS, DES, DESI, Euclid, HSC/PFS, LSST, POLARBEAR, SPT-3G, WFIRST, and others) will have the potential to constrain dynamical dark energy and departures from GR at the few-percent level. This will either rule out a large swath of the interesting parameter space of dark energy and modified gravity models or yield another breakthrough in cosmology with the detection of departures from ΛCDM. Thus, this area of cosmology is certain to yield interesting results in the near future.

SUMMARY POINTS

1. The weak and strong equivalence principles provide a means to rigorously distinguish dark energy, modified gravity, and an interacting dark sector at the theory level (**Figure 1**).

2. Quintessence dark energy is completely characterized by its equation of state $w(t)$. Dark energy physics beyond a canonical scalar field can be probed by searching for an inconsistency between geometry of the expanding universe as indicated by $H(z)$ and growth of the density fluctuation $\delta(k, z)$ (**Figure 5**).
3. At the phenomenological level, modified gravity can be probed by searching for fifth forces via comparison of dynamics with lensing or targeting screening effects (**Figure 3**). These signatures are very difficult to mimic with dark energy.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Alex Barreira, Dragan Huterer, Eiichiro Komatsu, Eric Linder, and Filippo Vernizzi for comments and Paolo Creminelli for permission to use **Figure 2**. A.J. was supported in part by the Kavli Institute for Cosmological Physics at the University of Chicago through grant NSF PHY-1125897, an endowment from the Kavli Foundation and its founder Fred Kavli, and by the Robert R. McCormick Postdoctoral Fellowship. L.L. was supported by a Science and Technology Facilities Council (STFC) Consolidated Grant for Astronomy and Astrophysics at the University of Edinburgh and a Postdoc.Mobility Grant from the Swiss National Science Foundation (SNSF) (grant 161058). F.S. acknowledges support from the Marie Curie Career Integration Grant (FP7-PEOPLE-2013-CIG) "FundPhysicsAndLSS."

LITERATURE CITED

1. Riess AG, et al. *Astron. J.* 116:1009 (1998)
2. Perlmutter S, et al. *Astrophys. J.* 517:565 (1999)
3. Weinberg S. *Rev. Mod. Phys.* 61:1 (1989)
4. Martin J. *C. R. Phys.* 13:566 (2012)
5. Frieman J, Turner M, Huterer D. *Annu. Rev. Astron. Astrophys.* 46:385 (2008)
6. Clifton T, Ferreira PG, Padilla A, Skordis C. *Phys. Rep.* 513:1 (2012)
7. Wetterich C. *Nucl. Phys. B* 302:668 (1988)
8. Peebles PJE, Ratra B. *Astrophys. J. Lett.* 325:L17 (1988)
9. Ratra B, Peebles PJE. *Phys. Rev. D* 37:3406 (1988)
10. Caldwell RR, Dave R, Steinhardt PJ. *Phys. Rev. Lett.* 80:1582(1998)
11. Caldwell RR, Linder EV. *Phys. Rev. Lett.* 95:141301 (2005)
12. Wetterich C. *Astron. Astrophys.* 301:321 (1995)
13. Copeland EJ, Liddle AR, Wands D. *Phys. Rev. D* 57:4686 (1998)
14. Ferreira PG, Joyce M. *Phys. Rev. Lett.* 79:4740 (1997)
15. Zlatev I, Wang LM, Steinhardt PJ. *Phys. Rev. Lett.* 82:896 (1999)
16. Sandvik H, Tegmark M, Zaldarriaga M, Waga I. *Phys. Rev. D* 62:023511 (2000)
17. Chiba T, Okabe T, Yamaguchi M. *Phys. Rev. D* 62:023511 (2000)
18. Armendáriz-Picón, Mukhanov VF, Steinhardt PJ. *Phys. Rev. D* 63:103510 (2001)
19. Cheung C, et al. *J. High Energy Phys.* 03:014 (2008)
20. Creminelli P, D'Amico G, Norena J, Vernizzi F. *J. Cosmol. Astropart. Phys.* 0902:018 (2009)
21. Gubitosi G, Piazza F, Vernizzi F. *J. Cosmol. Astropart. Phys.* 1302:032 (2013)

22. Bloomfield JK, Flanagan, Park M, Watson S. *J. Cosmol. Astropart. Phys.* 1308:010 (2013)
23. Gleyzes J, Langlois D, Piazza F, Vernizzi F. *J. Cosmol. Astropart. Phys.* 1308:025 (2013)
24. Bellini E, Sawicki I. *J. Cosmol. Astropart. Phys.* 7:050 (2014)
25. Lue A, Scoccimarro R, Starkman G. *Phys. Rev. D* 69:044005 (2004)
26. Brans C, Dicke RH. *Phys. Rev.* 124:925 (1961)
27. Will CM. *Living Rev. Relativ.* 17:4 (2014)
28. Khoury J, Weltman A. *Phys. Rev. Lett.* 93:171104 (2004)
29. Khoury J, Weltman A. *Phys. Rev. D* 69:044026 (2004)
30. Hinterbichler K, Khoury J. *Phys. Rev. Lett.* 104:231301 (2010)
31. Capozziello S. *Int. J. Mod. Phys. D* 11:483 (2002)
32. Capozziello S, Carloni S, Troisi A. *Recent Res. Dev. Astron. Astrophys.* 1:625 (2003)
33. Carroll SM, Duvvuri V, Trodden M, Turner MS. *Phys. Rev. D* 70:043528 (2004)
34. Erickcek AL, Smith TL, Kamionkowski M. *Phys. Rev. D* 74:121501 (2006)
35. Barrow JD, Cotsakis S. *Phys. Lett. B* 214:515 (1988)
36. Chiba T. *Phys. Lett. B* 575:1 (2003)
37. Joyce A, Jain B, Khoury J, Trodden M. *Phys. Rep.* 568:1 (2015)
38. Hu W, Sawicki I. *Phys. Rev. D* 76:064004 (2007)
39. Starobinsky AA. *J. Exp. Theor. Phys. Lett.* 86:157 (2007)
40. De Felice A, Tsujikawa S. *Living Rev. Relativ.* 13:3 (2010)
41. Woodard RP. *Scholarpedia* 10(8):32243 (2015)
42. Deffayet C, Gao X, Steer DA, Zahariade G. *Phys. Rev. D* 84:064039 (2011)
43. Carroll SM, Hoffman M, Trodden M. *Phys. Rev. D* 68:023509 (2003)
44. Cline JM, Jeon S, Moore GD. *Phys. Rev. D* 70:043543 (2004)
45. Horndeski GW. *Int. J. Theor. Phys.* 10:363 (1974)
46. Kobayashi T, Yamaguchi M, Yokoyama J. *Prog. Theor. Phys.* 126:511 (2011)
47. Nicolis A, Rattazzi R, Trincherini E. *Phys. Rev. D* 79:064036 (2009)
48. de Rham C, Gabadadze G. *Phys. Rev. D* 82:044020 (2010)
49. Dvali GR, Gabadadze G, Porrati M. *Phys. Lett. B* 485:208 (2000)
50. Luty MA, Porrati M, Rattazzi R. *J. High Energy Phys.* 09:029 (2003)
51. de Rham C. *C. R. Phys.* 13:666 (2012)
52. Langlois D, Noui K. *J. Cosmol. Astropart. Phys.* 1602:034 (2016)
53. Fierz M, Pauli W. *Proc. R. Soc. A* 173:211 (1939)
54. de Rham C, Gabadadze G, Tolley AJ. *Phys. Rev. Lett.* 106:231101 (2011)
55. Arkani-Hamed N, Georgi H, Schwartz MD. *Ann. Phys.* 305:96 (2003)
56. Creminelli P, Nicolis A, Papucci M, Trincherini E. *J. High Energy Phys.* 9:003 (2005)
57. Boulware DG, Deser S. *Phys. Rev. D* 6:3368 (1972)
58. Hassan SF, Rosen RA. *Phys. Rev. Lett.* 108:041101 (2012)
59. Hinterbichler K. *Rev. Mod. Phys.* 84:671 (2012)
60. de Rham C. *Living Rev. Relativ.* 17:7 (2014)
61. Deffayet C, Dvali GR, Gabadadze G. *Phys. Rev. D* 65:044023 (2002)
62. Deffayet C. *Phys. Lett. B* 502:199 (2001)
63. Nicolis A, Rattazzi R. *J. High Energy Phys.* 6:059 (2004)
64. Lue A. *Phys. Rep.* 423:1 (2006)
65. Dvali G, Gabadadze G, Shifman M. *Phys. Rev. D* 67:044020 (2003)
66. Arkani-Hamed N, Dimopoulos S, Dvali G, Gabadadze G. arXiv:hep-th/0209227 (2002)
67. Dvali G, Hofmann S, Khoury J. *Phys. Rev. D* 76:084006 (2007)
68. Weinberg S. *Phys. Rev. B* 138:988 (1965)
69. Feynman R, Morinigo F, Wagner W, Hatfield B. *Feynman Lectures on Gravitation*. Boulder, CO: Westview (1996)
70. Pietroni M. *Phys. Rev. D* 72:043535 (2005)
71. Olive KA, Pospelov M. *Phys. Rev. D* 77:043524 (2008)
72. Damour T, Polyakov AM. *Nucl. Phys. B* 423:532 (1994)
73. Brax P, van de Bruck C, Davis AC, Shaw D. *Phys. Rev. D* 82:063519 (2010)

74. de Rham C, Ribeiro RH. *J. Cosmol. Astropart. Phys.* 1411:016 (2014)
75. Babichev E, Deffayet C, Ziour R. *Int. J. Mod. Phys. D* 18:2147 (2009)
76. Milgrom M. *Astrophys. J.* 270:365 (1983)
77. Babichev E, Deffayet C, Esposito-Farese G. *Phys. Rev. D* 84:061502 (2011)
78. Khoury J. *Phys. Rev. D* 91:024022 (2015)
79. Vainshtein AI. *Phys. Lett. B* 39:393 (1972)
80. de Rham C, Tolley AJ, Wesley DH. *Phys. Rev. D* 87:044025 (2013)
81. Chu YZ, Trodden M. *Phys. Rev. D* 87:024011 (2013)
82. Gralla SE. *Phys. Rev. D* 81:084060 (2010)
83. Will CM. *Theory and Experiment in Gravitational Physics*. Cambridge, UK: Cambridge Univ. Press (1993)
84. Silvestri A, Trodden M. *Rep. Prog. Phys.* 72:096901 (2009)
85. Bekenstein JD. *Phys. Rev. D* 51:6608 (1995)
86. Hui L, Nicolis A, Stubbs C. *Phys. Rev. D* 80:104002 (2009)
87. Hiramatsu T, Hu W, Koyama K, Schmidt F. *Phys. Rev. D* 87:063525 (2013)
88. Cooray A, Sheth RK. *Phys. Rep.* 372:1 (2002)
89. Schmidt F, Lima MV, Oyaizu H, Hu W. *Phys. Rev. D* 79:083518 (2009)
90. Tinker JL, et al. *Astrophys. J.* 688:709 (2008)
91. Bond JR, Cole S, Efstathiou G, Kaiser N. *Astrophys. J.* 379:440 (1991)
92. Press WH, Schechter P. *Astrophys. J.* 187:425 (1974)
93. Tinker JL, et al. *Astrophys. J.* 688:709 (2008)
94. Mo HJ, White SDM. *Mon. Not. R. Astron. Soc.* 282:347 (1996)
95. Navarro JF, Frenk CS, White SDM. *Astrophys. J.* 490:493 (1997)
96. Creminelli P, et al. *J. Cosmol. Astropart. Phys.* 3:27 (2010)
97. de Putter R, Huterer D, Linder EV. 81:103513 (2010)
98. Armendáriz-Picón C. *J. Cosmol. Astropart. Phys.* 7:7 (2004)
99. Wei H, Cai RG. *Phys. Rev. D* 73:083002 (2006)
100. Koivisto T, Mota DF. *Phys. Rev. D* 73:083502 (2006)
101. Mota DF, Kristiansen JR, Koivisto T, Groeneboom NE. *Mon. Not. R. Astron. Soc.* 382:793 (2007)
102. Hu W. *Astrophys. J.* 506:485 (1998)
103. Farrar GR, Peebles PJE. *Astrophys. J.* 604:1 (2004)
104. Amendola L, Baldi M, Wetterich C. *Phys. Rev. D* 78:023015 (2008)
105. Baldi M, Pettorino V, Robbers G, Springel V. *Mon. Not. R. Astron. Soc.* 403:1684 (2010)
106. Bertschinger E. *Astrophys. J.* 648:797 (2006)
107. Hu W, Sawicki I. *Phys. Rev. D* 76:104043 (2007)
108. Gleyzes J, Langlois D, Mancarella M, Vernizzi F. *J. Cosmol. Astropart. Phys.* 8:054 (2015)
109. Silvestri A, Pogosian L, Buniy RV. *Phys. Rev. D* 87:104015 (2013)
110. Bertschinger E, Zukin P. *Phys. Rev. D* 78:024015 (2008)
111. Baker T, Ferreira PG, Skordis C. *Phys. Rev. D* 87:024015 (2013)
112. Wang J, Hui L, Khoury J. *Phys. Rev. Lett.* 109:241301 (2012)
113. Oyaizu H. *Phys. Rev. D.* 78:123523 (2008)
114. Schmidt F. *Phys. Rev. D.* 81:103002 (2010)
115. Schmidt F. *Phys. Rev. D.* 80:123003 (2009)
116. Saltas ID, Sawicki I, Amendola L, Kunz M. *Phys. Rev. Lett.* 113:191101 (2014)
117. Riess AG, et al. *Astrophys. J.* 730:119 (2011); Riess AG, et al. Erratum. *Astrophys. J.* 732:129 (2011)
118. Planck Collab. arXiv:1502.01589 [astro-ph] (2015)
119. Lombriser L, et al. *Phys. Rev. D* 85:102001 (2012)
120. Lombriser L, Slosar A, Seljak U, Hu W. *Phys. Rev. D* 85:124038 (2012)
121. Hu B, Liguori M, Bartolo N, Matarrese S. *Phys. Rev. D* 88:123514 (2013)
122. Reyes R, et al. *Nature* 464:256 (2010)
123. Hojjati A, et al. *Phys. Rev. D* 93:043531 (2016)
124. Schmidt F, Vikhlinin A, Hu W. *Phys. Rev. D* 80:083505 (2009)
125. Cataneo M, et al. *Phys. Rev. D* 92:044009 (2015)
126. Terukina A, et al. *J. Cosmol. Astropart. Phys.* 1404:013 (2014)

127. Jain B, Vikram V, Sakstein J. *Astrophys. J.* 779:39 (2013)
128. Lombriser L, Simpson F, Mead A. *Phys. Rev. Lett.* 114:251101 (2015)
129. Sachs RK, Wolfe AM. *Astrophys. J.* 147:73 (1967)
130. Sunyaev RA, Zel'dovich YB. *Mon. Not. R. Astron. Soc.* 190:413 (1980)
131. Sunyaev RA, Zel'dovich YB. *Annu. Rev. Astron. Astrophys.* 18:537 (1980)
132. Kaiser N. *Mon. Not. R. Astron. Soc.* 227:1 (1987)
133. Lam TY, Nishimichi T, Schmidt F, Takada M. *Phys. Rev. Lett.* 109:051301 (2012)
134. Zu Y, et al. *Mon. Not. R. Astron. Soc.* 445:1885 (2014)
135. Abbott P, et al. *Phys. Rev. Lett.* 116:061102 (2016)
136. LIGO Sci. Collab., Virgo Collab. arXiv:1602.03841 [gr-qc] (2016)
137. Jimenez JB, Piazza F, Velten H. *Phys. Rev. Lett.* 116:061101 (2016)
138. Chevallier M, Polarski D. *Int. J. Mod. Phys. D* 10:213 (2001)
139. Linder EV. *Phys. Rev. Lett.* 90:091301 (2003)
140. Linder EV. *Phys. Rev. D* 72:043529 (2005)
141. Mortonson MJ, Hu W, Huterer D. *Phys. Rev. D* 79:023004 (2009)
142. Ishak M, Upadhye A, Spergel DN. *Phys. Rev. D* 74:043513 (2006)
143. Ruiz EJ, Huterer D. *Phys. Rev. D.* 91:063009 (2015)
144. Zhao GB, Pogosian L, Silvestri A, Zylberberg J. *Phys. Rev. Lett.* 103:241301 (2009)
145. Baldi M, et al. *Mon. Not. R. Astron. Soc.* 440:75 (2014)
146. Puchwein E, Baldi M, Springel V. *Mon. Not. R. Astron. Soc.* 436:348 (2013)
147. Lombriser L, Taylor A. *J. Cosmol. Astropart. Phys.* 1603:031 (2016)
148. Zhang P, Liguori M, Bean R, Dodelson S. *Phys. Rev. Lett.* 99:141302 (2007)
149. Lombriser L, Hu W, Fang W, Seljak U. *Phys. Rev. D* 80:063536 (2009)
150. Raccanelli A, et al. *Mon. Not. R. Astron. Soc.* 436:89 (2013)
151. Xu L. *J. Cosmol. Astropart. Phys.* 1402:048 (2014)
152. Wyman M, Khoury J. *Phys. Rev. D* 82:044032 (2010)
153. Barreira A, Li B, Baugh C, Pascoli S. *J. Cosmol. Astropart. Phys.* 1408:059 (2014)
154. Barreira A, et al. *J. Cosmol. Astropart. Phys.* 8:028 (2015)

Annual Review of Nuclear and Particle Science

Volume 66, 2016

Contents

Errata

An online log of corrections to *Annual Review of Nuclear and Particle Science* articles may be found at http://www.annualreviews.org/errata/nucl