The Dark Matter Double Bind: Astrophysical Aspects of the Evidential Warrant for General Relativity*

William L. Vanderburgh†‡

The dark matter problem in astrophysics exposes an underappreciated weakness in the evidential warrant for General Relativity (GR). The "dark matter double bind" entails that GR gets no differential evidential support from dynamical phenomena occurring at scales larger than our solar system, as compared to members of a significant class of rival gravitation theories. These rivals are each empirically indistinguishable from GR for phenomena taking place at solar system scales, but make predictions that may differ radically from GR's at larger scales. Thus the typical confidence in the universal applicability of GR is insufficiently warranted in the present evidential context.

1. Introduction. This paper discusses the evidential warrant for the General Theory of Relativity—or rather, a lack therein. The existence of this weakness in the warrant for General Relativity (hereafter, GR) is typically not recognized. The issue arises in connection with the dark matter problem in astrophysics. The dark matter problem is an empirical discrepancy between independent methods of determining the masses of large scale dynamical structures such as galaxies and clusters of galaxies. Although the dark matter problem is philosophically interesting in its own right (see Vanderburgh 2001 and in preparation), this paper focuses on an empirical difficulty for testing gravitation theories that is raised by the dark matter problem, a difficulty originating in something I call the "dark matter double bind."

---

*Received September 2002; revised May 2003.

†To contact the author write to Department of Philosophy, Campus Box 74, Wichita State University, 1845 North Fairmount Street, Wichita, KS 67260-0074; e-mail: william.vanderburgh@wichita.edu.

‡I am very grateful to the following people for helpful comments on earlier versions of this paper: Virginia Trimble, Chris Smeenk, Kathleen Okruhlik, Bill Harper, Andrew Wayne, Bob Feleppa, and David Soles.

Philosophy of Science, 70 (October 2003) pp. 812–832. 0031-8248/2003/7004-0009$10.00

The plan of the paper is as follows. The remainder of this section provides some context for the overall discussion. Section 2 describes some of the features of the dark matter problem relevant to the evidential warrant of gravitation theories. Section 3 discusses two gravitation theories that have been proposed as alternatives to the existence of dark matter—though neither seems likely to be successful, considering them illustrates what is involved in attempts to establish the applicability of a gravitational law to galaxies and clusters of galaxies. Section 4 gives the central argument, based on the dark matter double bind, the conclusion of which is that the available dynamical evidence provides no empirical grounds for preferring GR over members of a class of rival gravitation theories as the theory applicable to the dynamics of galaxies and clusters. The concluding remarks in Section 5 include an answer to a possible objection to the approach taken in this paper.

To be clear from the outset, I am not challenging the available tests of GR. GR has proven to be highly empirically successful in the realms in which it has been tested, and these tests do provide strong epistemic support for the theory. I am arguing that the available tests of GR do not provide sufficient evidential warrant for the confidence in the *universal* applicability of GR typical amongst physicists and philosophers of science. GR is usually taken to yield empirically adequate predictions not just for systems of the types on which detailed tests have actually been carried out, but for all others as well—including large scale dynamical structures such as galaxies and clusters of galaxies, and even the universe as a whole. Even when it is acknowledged that GR must strictly speaking be false since it is incompatible with quantum mechanics, it is nevertheless assumed that the eventual theory of quantum gravity will make predictions empirically equivalent to those of GR for all phenomena above the Planck scale. However, this assumption is insufficiently warranted in the present evidential context: it turns out that there are even reasons to doubt the empirical adequacy of GR at galactic and greater scales. So although I am not challenging the available tests of GR, I am challenging the usual interpretation of the evidential import of those tests. I am, moreover, insisting on being careful about what counts as an epistemically significant "test" of a gravitation theory.

Though I discuss alternatives to GR, I am not advocating that some particular alternative is superior to GR. In fact, I think that GR is likely the best available option given the totality of the evidential and other considerations that could at present be brought to bear. But I am not concerned here to make that case. I am concerned, rather, to establish that the dynamical evidence available for deciding which dynamical theory governs systems of galactic and greater scale does not by itself warrant a high degree of confidence in GR's applicability to those systems.

If this assessment of the evidential situation is correct, the only route left for adjudicating between competing gravitation theories at galactic and greater scales will be to invoke methodological criteria of theory choice (such as simplicity). Although the final section remarks briefly on this topic, a full discussion is beyond the scope of this paper. As I indicate, the details of the case make applying methodological criteria more difficult than usual. Note that even if it is possible to formulate methodological arguments warranting a preference for some particular gravitation theory at galactic and greater scales, the resulting warrant will be very different in both kind and strength from the warrant GR has at solar system scales.

This raises a key point, namely the fact that all the available *tests* of GR involve interactions taking place over distances corresponding to the scale of the solar system or shorter. These tests include: the deflection of starlight by the gravitational field of the Sun; the motions of bodies within our solar system (notably including the motion of Mercury); the Shapiro time delay for signals passing near the Sun; the gravitational redshifting of light; the "frame dragging" effect for objects in orbit about other massive bodies; the formation and properties of black holes;[1] and some others. (See Will 1993 for details of the tests of GR; see also Harper and DiSalle 1996. Glymour 1980 and Earman 1992 both contain philosophical analyses of how this evidence provides epistemic support for GR.) One other success of GR is its correct prediction of the decrease of the periods of binary pulsars.[2] Binary pulsars, though very far from us, nevertheless have orbits smaller than our solar system. In order to include binary pulsars among the successful tests of GR, I use the phrase "stellar system scale tests." For perspective, note that the radius of a stellar system is much less than a light-year (even Pluto is only a few light-hours from the Sun), but the

---

[1] The masses involved in black hole formation from stellar collapse (above the Oppenheimer-Volkov limit of 2 to 3 solar masses for the formation of neutron stars) are greater than the total mass present in our solar system, but even stars very much more massive than our Sun have radii smaller than the orbital radius of Jupiter. Black holes have event horizons that are small, astronomically speaking, even in the case of the supermassive black holes found at the cores of many galaxies. (The equation is $R = 2GM/c^2$: a 100 million solar mass black hole would have an event horizon comparable in size to the orbit of Mars.) Black hole accretion disks and polar jets can extend thousands of light-years, but scientists are only beginning to establish the consistency of observations of these phenomena with GR, and are able to do so only by making significant assumptions about physical details of the systems in question. Observations of black holes and related phenomena are thus presently unable to yield tests that could ground a decision between competing gravitation theories at large scales.

[2] For discussions of binary pulsars as tests of GR, see Stoeger 1985b and van Stratten et al. 2001. The gravitational waves GR takes to be emitted by binary pulsars (and thus to explain the observed energy loss) are too weak to be detectable in the current generation of gravity wave detectors. (See, for example, http://www.ligo.caltech.edu/.)

gravitational interactions of stars within galaxies take place over up to tens of thousands of light-years. For galaxies within clusters, the scale of interaction is up to millions of light-years. The radius of the observable universe is about 14 billion light-years. Astronomically speaking, then, the above are (extremely) short scale tests, and probe only the tiniest fraction of the distances relevant to gravitation.

GR is also used to analyze the "gravitational lensing" of distant background light sources by foreground masses. Although the interactions involved in "micro-lensing" by stellar-mass and smaller foreground objects could perhaps be considered as taking place at stellar system scales, lensing by galaxies and clusters certainly cannot. But I argue in Section 4 that the gravitational lensing observations nevertheless do not warrant a preference for GR over its rivals.

The primary reason for the usual confidence in the universal applicability of GR is no doubt the fact that it is the most explanatorily powerful and empirically successful theory of gravity yet developed. It is natural to reason that the empirical successes of GR with regard to the phenomena for which it has been tested provide grounds for applying the theory to all gravitational phenomena whatsoever. To generalize in this way is to emulate Newton's own procedure in his argument for Universal Gravitation (hereafter, UG). Broadly speaking, Newton showed that multiple, diverse phenomena near the surface of the Earth and within the solar system yield agreeing high precision measurements of fundamental parameters of UG (such as the power law of the force of gravity), and argued that the theory should therefore be taken to apply to every pair of bodies in the universe. (Newton's argument is, naturally, more complex: see Harper 1997 and 2002.) An argument analogous to Newton's is at least implicit in current accounts of the evidential warrant for and scope of GR (see Harper and DiSalle 1996). The "Parameterized Post-Newtonian" testing framework, for example, leads to choices between competing gravitation theories on the basis of their respective levels of empirical adequacy with regard to values of theoretical parameters measured from various (stellar system scale) phenomena. (See Will 1993; and Earman 1992, 177ff.) The theory that "wins" in this process is then generalized to cover other phenomena for which it has not been tested.

The confidence in GR's universal applicability arising from the impressive successes of the solar system tests together with the argument for its universalization seems to spawn a dismissive attitude toward alternative theories. But the universalization of a theory assumes that it is at least consistent with phenomena for which detailed tests have not been carried out. As we know, Newton's argument for the universalization of UG ultimately failed, and it did so in part because the theory (taken together with appropriate background information) was discovered to be inconsistent with Mercury's motion. Replacing UG with GR allowed for the correct prediction of Mercury's motion (see Earman and Janssen 1993). UG turns out to be the weak field, low velocity limit of GR: the two theories are observationally indistinguishable for phenomena taking place in weak gravitational fields at relative velocities significantly slower than the speed of light. The predictions of the two theories diverge (and UG's are incorrect) for strong field and/or high velocity phenomena. From the point of view of UG, Mercury's discrepant perihelial precession is a phenomenon incompatible with the theory taken together with accepted background information—including, crucially, information about the mass distribution in the solar system. From the point of view of GR, Mercury's motion shows that UG is not universally applicable; UG is, rather, empirically adequate only for a restricted class of phenomena.

There is, as I discuss below, an empirical discrepancy of a corresponding sort for GR. This discrepancy casts doubt on the universal applicability of GR, and thus also undermines dismissive attitudes toward rival theories. GR is potentially consistent with the empirical evidence at galactic and greater scales, but only provided that substantive and otherwise unsupported background hypotheses are introduced (for example, about the mass distributions in systems at those scales). The same can be said, however, for an indefinite number of alternatives to GR.

Given that the motions within galaxies and clusters are much slower than the velocity of light and the gravitational fields are weak because of the distances involved, strictly speaking it is the Newtonian limit of GR that will be tested by the dynamical phenomena of interest here. The test of GR provided by these phenomena is therefore indirect to the extent that it depends on the assumption that GR reproduces the Newtonian predictions under the relevant circumstances. In the strictest sense, then, the astrophysical evidence undercuts the warrant for GR only insofar as GR reduces to Newtonian gravity in the appropriate limit for galaxy-scale and larger mass distributions. (As a referee noted, this distinction is important for several reasons, among them the fact that no one has produced galaxy models by 'patching together' Swarzschild-type solutions representing distinct stars and then shown that a Newtonian description holds in the limit.) To put it another way, the evidence may show that the correct weak-field, low-velocity limit of gravity at galactic and greater scales is not the Newtonian limit but something else.

2. Dark Matter. This paper is concerned with what may be called the *dynamical* dark matter problem, as opposed to the *cosmological* dark matter problem. Briefly, the cosmological dark matter problem is that the observed matter density in the universe is only twenty to forty percent of the theoretically expected density; to make up the difference, theorists proposed the existence of a huge quantity of unobserved matter—“dark” because, if it exists, it has no detectable electromagnetic signature. The interest in and importance of cosmological dark matter has waned in light of recent observations that the Hubble expansion is accelerating. The unknown (and misleadingly named) “dark energy” driving the acceleration appears sufficient to provide the extra effective mass (by the equivalence of mass and energy) without the need for cosmological dark matter. In any case, cosmological dark matter will not concern us further here.

The dynamical dark matter problem arises from a radical discrepancy between independent and seemingly unobjectionable methods of determining the masses of galaxies, clusters of galaxies and other large scale dynamical structures. The first method of determining the masses of such systems uses their observed motions, in concert with the Newtonian limit of GR, to calculate the gravitational mass required in order to maintain the system’s visible morphology given its internal motions. I will refer to the quantity thus measured as “dynamical mass.”

A typical case of measuring dynamical mass involves first determining the radius of a spiral galaxy from its angular size and distance. Then, from more detailed spectroscopic work, the galaxy’s rate of rotation (calculated from the Doppler shift of light) at various radii is measured. The result is a “rotation curve,” a plot of velocity along the line of sight versus distance from the galactic center. The observed rotation curves for spiral galaxies are decidedly “non-Keplerian” (by analogy with Kepler’s third law): instead of having a maximum near the center that falls off asymptotically to zero as radius increases, the rotation velocities are constant or even rise slightly. The non-Keplerian rotation is by itself evidence either for a large quantity of non-luminous matter in galaxies, or for a non-Newtonian dynamical law, even independent of the numerical *measure* of mass from the dynamics, since there is no way for the distribution of luminous mass plus the Newtonian limit to produce the observed rotation. (Parallel results are obtained for elliptical galaxies and clusters of galaxies, in which there is no uniform sense of rotation, by obtaining velocity dispersions.)

From data about velocities of rotation it is possible to calculate the total mass interior to a given radius at which the orbital velocity has been determined. Calculating dynamical mass in this way requires invoking some specific gravitation theory. The Virial Theorem (derived from Newtonian dynamics) is the most common way to calculate dynamical mass: $m = r\langle v^2\rangle/\alpha G$, where $\langle v^2\rangle$ is the average of the squares of the velocities at a given radius r, G is the gravitational constant, and α is a constant whose value depends on the mass distribution but which is usually of order unity (Tayler 1991, 194).

A second method of determining the masses of large scale astronomical structures is related to the Hertzpruung-Russell diagram, quantum mechanical theories of stellar evolution, dynamical studies of binary star systems, and empirical studies of the amount of mass associated with a given quantity of light for relatively nearby systems. Together these things establish expected "mass-to-light ratios" for distant systems of analogous kinds. Thus from an observed flux of light (at all wavelengths) the total mass that ought to be present in that system can be calculated. This "visible" or "luminous" mass includes, by inference from observed systems and reasonable theoretical assumptions, contributions to the total mass from gas, dust, dim stars and other objects that emit too little light to be detectable over astronomical distances. (The details of calculations of luminous mass are too complex to be fruitfully discussed here: see Vanderburgh 2001, 95–99 for an introduction.)

The strikingly odd thing is that despite the fact that the dynamical and luminous mass measures both have strong theoretical and observational bases and thus should be independently reliable, they nevertheless radically disagree. The dynamical masses of galaxies and clusters of galaxies are 10 to 100 times greater than their visible masses! This is *the astrophysical dynamical discrepancy,* often called the (dynamical) dark matter problem. The size of the discrepancy between the dynamical mass and the luminous mass depends on the kind of system in question—generally, the larger the system, the greater the difference between the two measures.

The dynamical discrepancy has two classes of possible solutions, each corresponding to taking as well-founded the assumptions of one rather than the other of these two methods of determining the masses of large scale dynamical structures. The first possible class of solutions (the one that most scientists favor) takes the assumptions of the dynamical mass measures to be correct. This implies that the dynamical discrepancy arises because, for some reason, the vast majority of the matter that makes up these systems (90–99% by mass) is not detected. The fact that this matter has so far gone undetected may indicate that it is in an entirely new and unexpected form—again "dark" because it neither absorbs, scatters nor emits discernible electromagnetic radiation. Some new type of matter is implied because almost every possible configuration of types of matter known to exist ought to be detectable given current technological capabilities and the fact that there would have to be so much of it. For example, if it were a cloud of gas or dust of sufficient mass, it would emit or scatter light in a distinctive way that is not observed; if it were a swarm of cold, jupiter-mass objects, the rate of gravitational microlensing of background light sources would be higher than is observed; and so on.

A huge number of diverse dark matter candidates have been proposed since the discovery of this problem in the 1930s, and especially since the "marriage" of particle physics and astrophysics in the mid-1970s. (For an overview of the history of dark matter, see Trimble 1990.) The candidates range from new, otherwise unknown fundamental particles, to failed or burned out stars in the haloes of galaxies, to supermassive black holes at the cores of galaxies. As Trimble has remarked (1993, 153, echoing Martin Rees), the range of masses of proposed "smallest units" of dark matter—from 10^{-38} grams for new fundamental particles to 10^{+39} grams for supermassive black holes—is a measure of our ignorance about what the dark matter might be.

There clearly must be a contribution to the dynamical mass from merely dim rather than "dark" baryons—for example brown dwarfs, small black holes, clouds of cold gas, etc.—and there is even still some possibility (a small one if the Big Bang nucleosynthesis limits are taken seriously) that the excess mass in galaxies and clusters is entirely baryonic. But the best evidence now available seems to indicate that the majority of the dark matter must be nonbaryonic. The most popular nonbaryonic candidates are fundamental particles that interact with ordinary matter only through the weak nuclear force (including neutrinos, axions, and certain supersymmetric particles). However, almost every candidate so far proposed—baryonic and nonbaryonic—can be ruled out, at least as the whole solution, on the basis of some conflict with known facts; none of the rest of the candidates have any significant confirming evidence. (See Trimble 1987 and Vanderburgh 2001.) It is nearly certain that neutrinos (for example) are a contributor to the overall dynamical mass, perhaps comparable in quantity to the luminous mass: neutrinos are weakly interacting, are predicted to have been produced in huge numbers in the Big Bang, and have been empirically shown to have non-zero mass. Nevertheless, various empirical considerations also indicate that neutrinos cannot be the whole story. Similarly, most of the other dark matter candidates so far proposed either cannot entirely resolve the dynamical discrepancies at all scales and in all types of systems by themselves or, as in the case of the supersymmetric particle candidates, are not even known to exist. It may be that there are multiple components to the dark matter, or that the correct matter candidate has yet to be described.

The second possible class of solutions to the dynamical discrepancy involves accepting the visible mass as a correct estimate of the total mass actually present in astronomical systems. A few workers in the field have over the years opted to explore this possibility, often on the grounds that dark matter is suspiciously ether-like in its apparently ad hoc origin, in its persistent (perhaps even principled) unobservability, and in its supposed ubiquitousness. (See, for example, Mannheim 1994.) But this option is unpopular among physical scientists—in part because it involves asserting that GR is not the correct theory of gravity. However, as I argue below, alternatives to GR are viable in the present evidential context and should be considered seriously.

3. Alternatives to General Relativity. Several gravitation theories have been offered in attempts to resolve the dynamical discrepancy without invoking dark matter; I mention two recent examples here. The point is to show that such solutions are possible in the current evidential and theoretical context (although these two particular attempts probably fail), and to consider some of the issues involved in trying to empirically determine what gravitational theory governs the dynamics of galaxies and clusters.

The first of the two theories I will mention is Milgrom's "Modification of Newtonian Dynamics"—hereafter, MOND (see Milgrom 1983, 1994). MOND was constructed in a curve-fitting solution to the dynamics of spiral galaxies on the assumption that the luminous mass is the only mass present. (This is analogous to Hall's 1894 determination of the gravitational power law from Mercury's motion on the assumption that there is no unknown matter near the Sun. For a discussion of Hall's and others' attempts to resolve Mercury's anomalous motion and the testing of GR, see Earman and Janssen 1993.) Thus MOND yields correct predictions for the dynamics of some galaxies without the need for dark matter. Now, there is good reason to expect that MOND must be false: whatever theory of gravitation is correct, it will be special-relativistic, and MOND is not. Astronomers do, however, just use Newtonian UG rather than GR to analyze the motions of galaxies and other dynamical structures. The gravitational fields (outside galactic centers at least) are weak and the velocities are very much lower than the speed of light,[3] so relativistic effects ought not have any important contribution to the overall dynamics. This explains why MOND has received some attention: it is potentially the weak field, low velocity ("quasi-Newtonian") limit *for large scale interactions* of whatever turns out to be the correct relativistic theory of gravity. (MOND causal stories involve either a new power law for gravity at great distances, or a distance-dependent concept of inertia. See, for example, Milgrom 1994, 1986, and 1983; Sanders 1996 and 1999.)

The second alternative gravitation theory I will mention was proposed by Mannheim and Kazanas (1989) and subsequently developed and defended by Mannheim (1996, 1994, 1993, 1992). Mannheim and Kazanas (1989) describes a new exact vacuum solution for a spherically symmetric mass distribution. (The theory was prefigured by Weyl, so they call it "Weyl gravity.") Mannheim and Kazanas were originally interested in the project of unifying the fundamental forces—because the other forces are conformally invariant, they hoped making gravity conformally invariant

---

[3] Reid et al. (1999) measure the solar system's motion around the galactic center to be 220km/s. This value is a typical for the motions within galaxies. Even the motions of galaxies in clusters are well below the speed of light.

would make unification easier. Their unification project was unsuccessful, but they happened to notice that the theory of gravity constructed in the attempt included an extra linear term. This extra linear term is so small that its effects become noticeable only on the scale of galaxies, which means that Weyl gravity can take over all of the stellar system scale successes of General Relativity—the two theories are empirically indistinguishable at stellar system scales. At larger scales, though, the contribution of the linear term does become important: the predictions of Weyl gravity differ considerably from those of GR at galactic scales, and the predictions continue to diverge as the separation between the gravitating bodies increases. So, given the same amount and distribution of mass, Weyl gravity predicts that the force of gravity should be significantly stronger at galactic scales than GR predicts. Weyl gravity reproduces fairly accurately the rotation curves for spirals and the velocity distributions for some clusters without the need for dark matter. Thus in the same way that Newton's UG turned out to be GR's weak field, low velocity limit, GR could turn out to be the stellar system scale limit of some other relativistic gravitation theory such as Weyl gravity.

Despite the initial appearance of empirical adequacy, however, Weyl gravity suffers serious (though perhaps not fatal) difficulties. Edery and Paranjape (1998) developed for the first time Weyl gravity's predictions for standard cases of gravitational lensing. They showed that Weyl gravity requires an even greater amount of dark matter than does GR in order to account for the observed cases of gravitational lensing of background sources by foreground galaxies and clusters. This is not definitive evidence against Weyl gravity—nothing excludes having a non-GR gravitation theory *and* a lot of dark matter—but it certainly removes the main motivation for considering rivals to GR, namely the promise of being able to do without dark matter. Mark Walker (1994) has also argued that Weyl gravity is incompatible with observations of cosmic scale weak gravitational lensing. (For more on MOND and Weyl gravity see Vanderburgh 2001.)

Even if Weyl gravity itself is untenable, some such (as yet unarticulated) rival gravitation theory could possibly account for all the dynamical evidence without the need for dark matter. The general issue of the relative epistemic warrant from the astrophysical evidence for competing gravitation theories remains even if some specific rivals are eliminated. The strong empirical constraints obtained through the stellar system scale tests mean that any viable theory must make predictions that are empirically indistinguishable from those of GR for short scale interactions. But the stellar system scale tests by themselves do not establish constraints on the gravitational action at larger scales. It would therefore be ideal to acquire dynamical evidence at those scales that would provide differential empirical support for one or some rivals over the others. However, such evidence is currently unavailable, and as I argue in the next section, it seems unlikely that it will become available.

4. The Dark Matter Double Bind. GR is usually taken to apply not merely to all dynamical systems but even to the universe as a whole. As in the case of Newton's argument to UG, for GR an inductive generalization is made from a set of locally obtained pieces of evidence, where these pieces of evidence are consistent with each other and ideally involve independent agreeing measures of theoretical parameters from several distinct phenomena. Given the epistemic strength of the independent measures of the parameters of UG obtained via Reasoning from Phenomena (Harper 2002), Newton's extension of the principle of mutual gravitation to all bodies whatsoever is the step of the argument on which the inductive risk is focused (Smith 2002, 160). As is now known, Newton's bet here failed. Generalizing the inverse square law to all gravitational phenomena whatsoever merely on the basis of (what are now understood to be) weak field, low velocity tests, turned out to be a mistake. In the same way, it is epistemically risky to extend GR to all dynamical systems regardless of scale merely on the basis of stellar system scale tests. This extension might nevertheless be unworrisome were it not for the astrophysical dynamical discrepancy, one possible cause of which is that GR does not apply at galactic and greater scales. Simply put, GR plus the assumption that the visible matter is all the matter present in large scale dynamical systems leads to predictions that are incompatible with the observations.

While the focus here is on the evidence for GR's applicability to galaxies and clusters, it is also worth pointing out that there are also difficulties in trying to establish GR's applicability to the cosmos as a whole. Ellis has argued (1999, 1985) that cosmology not only in fact relies on untested assumptions, but *must* rely on substantive *untestable* assumptions (e.g., the Cosmological Principle—the universe is homogeneous and isotropic on large scales) in order to make it possible to acquire information about the universe as a whole. Ellis argues that the large scale spacetime structure of the universe cannot be uniquely determined by purely observational tests. It follows, Ellis points out, that various hypotheses regarding large scale structure remain viable in the face of any cosmological evidence that might be obtained.

A related fact, relevant to the present discussion, is that some of the empirically viable alternative cosmological hypotheses will be ones based on gravitation theories other than GR. If it is impossible to determine the large scale structure of the universe by purely empirical means, it follows that it is impossible to establish on empirical grounds alone what law of gravity holds for the universe as a whole: one would need to know the large scale structure before one could begin to assess whether a given gravitation theory is even consistent with that structure.

These considerations cast into doubt Will’s claim (1993, 310–319) that cosmology has been a testing ground for gravitation theories since the 1920s. It is true that various cosmological observations (for example, the Hubble recession and the cosmic microwave background) have been *taken* as confirming that the universe satisfies the Friedman-Robertson-Walker spacetime model, and therefore GR (because the Friedman-Robertson-Walker model is a solution to the GR field equations). But if Ellis’s arguments are correct, this supposed confirmation is illusory or weak. Certainly, it does not reach the level of precision and power achieved by the stellar system scale tests of GR; the supposed “confirming evidence” in the cosmological case really amounts to no more than showing that GR is consistent with the available cosmological observations given plausible but rather strong and evidentially un- or under-supported assumptions. Given different (equally unsupported) assumptions, the cosmological observations would be consistent with universe models based on alternative gravitation theories. If there were some way to confirm one set of assumptions over the others, progress could perhaps be made toward deciding what the correct theory of the large scale structure of the universe is, and this would make possible the confirmation of a theory of gravitation at cosmic scales relative to its rivals. There seems to be no way to do this, and so the cosmological arena appears to offer no opportunity for evidentially distinguishing GR from its rivals.

A version of Ellis’s argument may or may not stand up to closer scrutiny. (One hurdle is that Ellis pitches his claims in terms of a naïve verificationism; I contend that the conclusion will hold once the argument is re-cast in more appropriate epistemic terms, but will not make the case here.) In any event the main argument of this paper does not depend on it. There is a much more serious problem that indicates at the very least a current lack of evidential warrant for GR as compared to actual and potential rivals at galactic and greater scales, and that perhaps even indicates the impossibility of being able to test GR against rivals at those scales.

The problem in question is “the dark matter double bind,” which goes as follows. Consider in general what must be shown in order to establish that some dynamical law is empirically adequate with respect to a given system. Minimally, this would require showing that the law’s predictions agree with the observed motions. In order to derive a prediction of the motions, however, the law must be conjoined with hypotheses about the number and distribution of bodies in the system, their masses, and their instantaneous velocities. Thus, in order to check whether or not GR’s predictions of the motions within a given spiral galaxy, say, agree with the observed motions, one would first need to import background information about the mass distribution within the galaxy. But the astrophysical dynamical discrepancy raises doubts about exactly this information. The mass distribution could be inferred from the observed motions plus some dynamical law, if it were known which law applied to the system in question—but that is exactly what is to be determined. Thus the dark matter double bind: *In order to evaluate the empirical adequacy of any gravitation theory at galactic and greater scales, the mass distribution in dynamical systems at those scales must first be known—but because of the astrophysical dynamical discrepancy the mass distribution is not known. In order to infer the mass distribution from the observed motions, a gravitational law must be assumed—but such a law cannot legitimately be assumed, since the very thing at issue is which gravitational law ought to be taken to apply at galactic and greater scales.*

It is true that given one from among the rival gravitation theories, the observed motions allow one to infer a mass distribution. Taking the Newtonian limit of GR as given leads to the standard inference that there is about one hundred times more dark matter than visible matter in galaxies and clusters, and that the dark matter must be distributed in a spheroid "halo" extending to several times the visible radii of these systems. But this result can be trusted only to the extent that there are grounds to be sure that GR applies. A contrary conclusion about the mass distribution could be derived from an alternative gravitational assumption, one that would be equally empirically warranted by the available dynamical evidence.

Note that since assessing even the basic consistency of a gravitation theory with the dynamical phenomena at galactic and greater scales is impossible because of the dark matter double bind, it is certainly impossible to conduct the sorts of detailed, robust and epistemically significant tests of GR that are possible at stellar system scales. The reason the stellar system scale tests do not suffer a similar problem is that there exists at these scales a systematic interrelation of agreeing and precise measurements-from-phenomena of key theoretical parameters. For example, the planets, asteroids, comets, and artificial satellites all give agreeing measures of the mass of the Sun. Likewise, the system of mutual perturbations between the planets provides multiple, agreeing measures of the masses of each of the planets; similarly for those planets with moons. The fact that it is possible to give a consistent assignment of masses to the known bodies in the solar system, one which is backed up by multiple agreeing measurements (both of the masses and of the parameters of the gravitation theory), is grounds for inferring that there is no invisible, unaccounted-for mass in the solar system. With this in place, individual solar system phenomena (the orbit of Mercury, say) can be employed to make detailed, precise comparisons of rival gravitation theories. No such agreeing measures are available for large scale dynamical systems. (An apparent case of agreeing measures of galactic masses from dynamics and from gravitational lensing is discussed below.)

The stellar system scale tests provide strong constraints on the predictions of gravitation theories at those scales; however, they provide no empirical reason to prefer any one theory from the class of potential theories that are empirically indistinguishable from GR at stellar system scales but whose predictions differ from those of GR (and which require correspondingly more or less dark matter) at galactic and greater scales. The dark matter double bind indicates that there is not currently, and perhaps cannot be, an empirical basis on which to decide among these rival gravitation theories. The evidential status of GR is thus considerably weaker than is usually supposed.

Is there any empirical way to avoid the pessimistic conclusion of the dark matter double bind? Might it be possible, for example, to find independent grounds for a particular hypothesis about the mass distribution in spiral galaxies, and then use that distribution as the basis for an empirical comparison of gravitation theories at that scale? Unfortunately the prospects for this seem dim. Deriving the mass distribution in galaxies from a theory of galaxy formation and evolution would supply independent grounds for comparing mass distributions inferred from the dynamics by rival gravitation theories, and hence for judging the relative empirical success of those rival theories, but any theory of galactic evolution will have to assume a gravitational law to begin with, so it is difficult to see how this approach could succeed. Alternatively, if some dark matter particle is eventually detected it might then be possible to infer a galactic distribution of that particle from features of the pattern of detection or from characteristics of the particle itself. But until this sort of information becomes available—and dark matter particles have been remarkably persistently undetected despite almost a quarter century of concerted efforts to detect them—there are no evidential grounds for preferring GR over rivals, or even for thinking that GR is consistent with the dynamics of large scale systems (rather than being falsified by them). Even were such evidence to become available, it would be so indirect that it would likely not license a definitive choice of gravitation theory anyway.

As mentioned earlier, the gravitational lensing of distant galaxies or quasars by foreground galaxies or clusters could in principle provide significant evidence in favor of a particular gravitation theory over its rivals at scales greater than those of stellar systems since, unlike the dynamical measures in which the Newtonian limit is sufficient, the inferences from observations of gravitational lensing depend on the properly relativistic parts of GR. Unfortunately, the hope of using lensing in this way is also dashed by the existence of the dynamical discrepancy. If it were known in advance what the mass distribution in the lensing body was, then it would be possible to compare the predictions of various gravitation theories against the lensing observations, and this would provide fairly strong constraints on the gravitation theories. But the existence of the astrophysical dynamical discrepancy means that the masses of these structures are in doubt.

In practice, when analyzing observed cases of gravitational lensing, physicists assume that GR applies and use it to infer from features of a lensing situation such things as the mass of the lens. Interestingly, the masses of spiral galaxies inferred in this way turn out to be of the same order of magnitude as the dynamical masses for spirals with similar diameters and luminosities. Galaxies that serve as lenses in this way are generally too far away, and hence too dim, for detailed spectrographic rotation curves to be obtained; as far as I am aware, no single galaxy has had its mass determined from both lensing and dynamics. However, many clusters of galaxies have been analyzed by both methods, and some have additionally had their masses corroborated by the X-ray emissions from haloes of hot diffuse gas surrounding them. X-ray emissions from gas haloes can be turned into measurements of the total mass of the system through the assumptions that the gas will emit X-rays only if continually heated, and that the source of the heating is the gravitational potential of the system. The margins of error are naturally not small, and the results are model-dependent in that the amount of mass inferred from the potential (itself inferred from the X-ray emissions) depends on which gravitational law is taken to be operative.[4]

This coincidence of mass measures is certainly suggestive, but it is by no means definitive. The agreement is only rough, and the same lensing and X-ray observations can be used in calculations of the masses of the lensing bodies via alternative theories of gravitation. Nothing rules out finding a coincidence of dynamical, X-ray, and lensing masses when assuming a different theory of gravity (though, as mentioned above, there is in fact no such coincidence in the case of Weyl gravity). Because of the dark matter double bind there is no independent check on the masses of astrophysical systems, and so gravitational lensing cannot provide definitive evidence in favor of a particular gravitation theory over its rivals. The coincidence of the gravitational lensing, X-ray, and dynamical mass measures, rough though it may be, is nevertheless the best available empirical reason for thinking that

---

4. Buote et al. (2002) notes that in a few galaxies the orientation of the cloud of X-ray gas is different from that of the luminous matter, and they argue that this implies there must be a halo of dark mass surrounding the luminous matter. The shape of the gravitational potential heating the gas must be the same as that of the mass distribution, whatever gravitational law holds. If robust, these results show that there is dark matter. Note, however, that the dark matter double bind still stands, which means this result still does not tell us which gravitational law governs large scale dynamics.

GR applies to galaxies and clusters. It must be emphasized, however, that any epistemic support GR gets from lensing observations derives merely from an order-of-magnitude agreement of masses that itself depends on a fairly loose analogical inference about correlations between dynamical masses and luminosities within spiral galaxy types. This is a much weaker kind of support than GR is usually taken to have, and it is certainly nothing like the kind of empirical support GR receives from the stellar system scale tests. Given such lax empirical standards, it seems quite likely that other possible theories will also be able to claim this sort of agreement with the evidence. This is not to say that GR should be abandoned—GR is, after all, the best theory so far available—but it is to say that a more circumspect assessment of GR's epistemic status is appropriate.

5. Concluding Remarks. One way to challenge the approach taken in this paper would be to argue that no significant implications follow from the mere possibility of the existence of rivals to an entrenched theory. Newton's Fourth Rule of Reasoning would seem to support this line.

> In experimental philosophy, propositions gathered from phenomena by induction should be considered either exactly or very nearly true notwithstanding any contrary hypotheses, until yet other phenomena make such propositions either more exact or liable to exceptions. [Newton comments:] This rule should be followed so that arguments based on induction may not be nullified by hypotheses. (Newton [1726] 1999, 796)

In Newton's sense, GR would count as a "proposition gathered from phenomena by induction," namely via the stellar system scale tests. It would at first seem that according to Rule Four GR ought therefore to be considered to apply universally, despite any rival theories that might be imagined, even though the theory has not been gathered from phenomena at galactic and greater scales. Rule Four would also thereby seem to provide grounds for rejecting merely possible alternatives to GR, "so that arguments based on induction may not be nullified by hypotheses." A second look, however, pulls in the opposite direction.

Methodological principles of theory conservatism, including Rule Four, are legitimately invoked only under conditions that are not satisfied in the present case. Such principles are meant to justify the extension of a theory to a new class of phenomena when, first of all, the theory cannot be or has not been tested with respect to those phenomena and when, second, there are reasonable grounds for thinking that those phenomena are relevantly analogous to other phenomena through which the theory has been successfully tested. Such principles are meant, in other words, to protect a theory from attack when there are no positive empirical reasons to doubt its adequacy. But in the case of the dynamics of galaxies and larger dynamical structures the grounds for asserting an analogy to solar system dynamics are undercut by the existence of the dynamical discrepancy. Given the astrophysical evidence, it cannot be pretended that GR has been gathered from phenomena by induction at galactic and greater scales. It has been so gathered only in an extremely tiny fraction of the relevant distances. At large scales GR actually contradicts the observations under the most natural initial assumption about the mass distribution at those scales: the dynamical discrepancy thus could be precisely one of Newton's "yet other phenomena" requiring revision of the theory of gravity. Rule Four licenses protecting an entrenched theory *until* such potentially falsifying phenomena come to light.

This means, I take it, that it is legitimate in the present evidential context to explore all kinds of solutions to the dynamical discrepancy—including matter solutions that assume GR in their background, and gravity solutions that hypothesize successors to GR—and that none of the possible solutions can be rejected out of hand. If my interpretation of Newton's Fourth Rule as it applies to this case is incorrect, then the best available grounds for extending GR to galactic and greater scales will be merely methodological. And a merely methodological justification is a much weaker sort of warrant than GR is typically thought to have.

Note that, as in the case of Mercury's anomalous precession, the dynamical discrepancy challenges only the *inductive extension* or *universalization* of the theory to new domains on the basis of evidence from other domains; it does not in any way weaken that evidence itself. One of the interesting and powerful things about Newtonian methodology, as Smith (2002) makes clear, is its ability to turn discrepancies between theory and observation into data for constructing a more precise successor theory. The astrophysical dynamical discrepancy could end up functioning this way for GR. That is, it could be evidence for a revision of the theory of gravity in which GR comes to be seen as an approximation to its successor, namely the approximation for stellar system scale phenomena. But we will need to obtain new and independent information about the overall mass distributions in galaxies and larger systems before this process can be carried out in the required detail. The very same discrepancy, after all, might be evidence for a "successor theory" of mass distribution if the Newtonian limit of GR is correct for galaxies and clusters.

Presumably some particular combination of mass distribution and gravitation theory will turn out to be empirically adequate with respect to the dynamical evidence at all scales. But at this point there is no clear picture of what the future solution to the astrophysical dynamical discrepancy will look like—the available evidence merely delineates a rather large portion of the space of possible combinations of gravitational laws and mass distributions. In the present evidential context there is no option but to consider "merely possible" theories in the attempt to solve the astrophysical dynamical discrepancy, and to attempt to establish which law of gravitation applies at galactic and greater scales. I contend, furthermore, that assessing "merely possible" theories will increase our understanding of the evidential status of already articulated theories, as well as increasing our understanding of the empirical and other constraints on future theories.

As Earman has remarked in a quite different context, "If a belief in the General Theory of Relativity is to be rational, it must be based on a systematic exploration of alternatives that have yet to be invented" (Earman 1992, 182). Earman is here referring to the exploration of the space of possible gravitation theories via the eliminative methods of the Parameterized Post-Newtonian testing framework (and possible extensions of that framework). Earman argues that without a systematic assessment of the rivals to a given theory—merely possible rivals as well as actually articulated ones—"there typically is no rational basis for the assignment of a high degree of confidence to some particular theory in the field" (1992, 182). The point stands in light of the considerations adduced here, which show that the space of alternative theories is even larger than might have been thought, since it includes theories which make predictions radically different from those of GR at galactic and greater scales. Earman remarks (1992, 180) that with respect to cosmological observations there is as yet no analogue of the Parameterized Post-Newtonian scheme with which to winnow the class of viable theories; the present paper shows that the same is true in the regime of distances intermediate between individual stellar systems and cosmological structure.

If the present account is correct, a purely empirical resolution of the astrophysical dynamical discrepancy will be impossible unless the evidential situation changes. Perhaps, though, it will be possible to find nonempirical grounds on which to prefer some particular gravitation theory at the scales in question. Methodological criteria of theory choice could be invoked to this end. An important point to notice, however, is that no such methodological account has yet been offered in detail.

I leave for another occasion a full analysis of possible methodological principles and arguments that could be invoked for or against various solutions to the dynamical dark matter problem, and thus for or against specific gravitation theories. It is, however, worth mentioning some factors that would need to be taken into account in any such methodological solution. For example, a comparison of the relative simplicity of rival gravitation theories will be difficult not just because of the standard problems of determining what simplicity is and why it is methodologically significant. Invoking relative simplicity will be even more difficult than usual because the judgment in this case is not to be made between rival gravitation theories alone, but between theoretical groups each consisting of a gravitational law plus its respective theory of dark matter. Both parts are required in order to account for large scale dynamics. Since there is as yet no adequate characterization of dark matter or of the forms of the rival gravitation theories, comparison of gravity-plus-dark-matter theoretical groups in terms of simplicity cannot even get off the ground. Moreover, this approach requires being able to compare very different kinds of simplicity. For example, it must be possible to say whether a theory that maintains GR but proposes a large quantity of undetectable matter is simpler than another in which, say, the law of gravitation has a more baroque mathematical structure but requires less dark matter—or fewer kinds of dark matter, or simpler kinds of dark matter, etc. Similar considerations will apply for other sorts of methodological criteria. It may be that the empirical and theoretical situation is currently so underdeveloped that not even methodological preferences among the possible rival theories can be adequately formulated. This is not to say that it will ultimately be impossible to find a methodological principle that could succeed in warranting a preference for one particular gravitation theory over its rivals. It is to say, however, that finding (and justifying) such a principle will be difficult.

In conclusion, the existence of the dark matter double bind entails that the available dynamical evidence cannot license distinguishing GR from its rivals at galactic and greater scales. This is of no consequence for stellar system scale interactions, where there is solid evidence that GR saves the phenomena better than all rivals so far evaluated. But if we are concerned to find out about galaxies and clusters, to study their long term evolution, to theorize about the large scale structure and evolution of the universe as a whole, or to draw philosophical implications about the nature of space and time from physical theories, it ought to give us pause to consider that it is not known whether GR or some other substantially different gravitation theory applies above the scale of stellar systems. GR might well be universally applicable, but in the present evidential context that has not been established. It is possible that GR is simply the short scale limit of some other relativistic gravitation theory. The only available evidential constraint on potential successor theories is that they be able to account for the stellar system scale phenomena as well as or better than GR. Exploring the relations between merely possible rival theories in this context helps to illuminate the kinds of evidential and methodological considerations that will need to be brought to bear in order to reach a solution to the dark matter problem, and to decide which theory of gravity applies at galactic and greater scales.

REFERENCES

Buote, David A., Tesla E. Jeltema, Claude R. Canizares, and Gordon P. Garmire (2002), “Chandra Evidence for a Flattened, Triaxial Dark Matter Halo in the Elliptical Galaxy NGC 720”, *Astrophysical Journal* 577: 183–196.

Cohen, I. Bernard, and George E. Smith (eds.) (2002), *The Cambridge Companion to Newton*. Cambridge: Cambridge University Press.

Earman, John (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge: MIT Press.

Earman, John, and Michel Janssen (1993), “Einstein’s Explanation of the Motion of Mercury’s Perihelion”, in John Earman, Michel Janssen and John D. Norton (eds.), *The Attraction of Gravitation*. Boston: Birkhauser, 129–172.

Edery, A., and M. B. Paranjape (1998), “Classical Tests for Weyl Gravity: Deflection of Light and Time Delay”, *Physical Review* D 58: 024011.

Ellis, G.F.R. (1985), “Observational Cosmology After Kristian and Sachs”, in Stoeger (1985a), 475–86.

——— (1999), “The Different Nature of Cosmology”, *Astronomy and Geophysics* (August 1999): 4.20–4.23.

Glymour, Clark (1980), *Theory and Evidence*. Princeton, NJ: Princeton University Press.

Harper, William (1997), “Isaac Newton on Empirical Success and Scientific Method”, in John Earman and John D. Norton (eds.), *The Cosmos of Science: Essays of Exploration*. Pittsburgh: University of Pittsburgh Press, 55–86.

——— (2002), “Newton’s Argument for Universal Gravitation”, in Cohen and Smith (2002), 174–201.

Harper, William, and Robert DiSalle (1996), “Inferences from Phenomena in Gravitational Physics”, *Philosophy of Science* 63 (Proceedings): S46–54.

Mannheim, Philip D. (1992), “Conformal Gravity and the Flatness Problem”, *The Astrophysical Journal* 391: 429–432.

——— (1993), “Linear Potentials and Galactic Rotation Curves”, *The Astrophysical Journal* 419: 150–154.

——— (1994), “Open Questions in Classical Gravity”, *Foundations of Physics* 24: 487–511.

——— (1996), “Local and Global Gravity”, *Foundations of Physics* 26: 1683–1709.

Mannheim, Philip D., and Demosthenes Kazanas (1989), “Exact Vacuum Solution to Conformal Weyl Gravity and Galactic Rotation Curves”, *The Astrophysical Journal* 342: 635–638.

——— (1994), “Newtonian Limit of Conformal Gravity and the Lack of Necessity of the Second Order Poisson Equation”, *General Relativity and Gravitation* 26.4: 337–361.

Milgrom, Mordehai (1983), “A Modification of the Newtonian Dynamics as a Possible Alternative to the Hidden Mass Hypothesis”, *Astronomical Journal* 270: 365–70.

——— (1986), “Solutions for the Modified Newtonian Dynamics Field Equations”, *Astrophysical Journal* 302: 617–25.

——— (1994), “Dynamics with a Non-Standard Inertia-Acceleration Relation: An Alternative to Dark Matter in Galactic Systems”, *Annals of Physics* 229: 384–415.

Newton, Isaac (1999), *The Principia: Mathematical Principles of Natural Philosophy.* Translated by I. Bernard Cohen and Anne Whitman. Berkeley: University of California Press.

Reid, M. J., A. C. S. Readhead, R. C. Vermeulen, and R. N. Treuhaft (1999), “The Proper Motions of Sgr A*: I. First VLBA Results”, <http://www.arxiv.org/PS_cache/astro-ph/pdf/9905/9905075.pdf>.

Sanders, R. H. (1996), “The Published Extended Rotation Curves of Spiral Galaxies: Confrontation with Modified Dynamics”, *Astrophysical Journal* 473: 117–29.

——— (1999), “The Virial Discrepancy in Clusters of Galaxies in the Context of Modified Newtonian Dynamics”, *Astrophysical Journal Letters* 512: L23–L26.

Smith, George (2002), “The Methodology of the *Principia*”, in Cohen and Smith (2002), 138–173.

Stoeger, William R. (ed.) (1985a), *Theory and Observational Limits in Cosmology: Proceedings of the Vatican Observatory Conference*. Vatican City: Specola Vaticana

——— (1985b), "The Binary Pulsar PSR 1913+16 and its Role in Testing Theories of Gravity", in Stoeger (1985a), 497–509.

Tayler, Roger J. (1991), *The Hidden Universe*. Chirchester, UK: Ellis Horwood Limited.

Trimble, Virginia (1987), "Existence and Nature of Dark Matter in the Universe", *Annual Review of Astronomy and Astrophysics* 25: 425–472.

——— (1990), "History of Dark Matter in the Universe (1922–1974)", in B. Bertotti, R. Balbinot, and S. Bergia (eds.), *Modern Cosmology in Retrospect*. Cambridge: Cambridge University Press, 355–362.

——— (1993), "Dark Matter", in Noriss S. Hetherington (ed.), *Encyclopedia of Cosmology: Historical, Philosophical and Scientific Foundations of Modern Cosmology.* New York: Garland, 148–58.

Vanderburgh, William L. (2001), Dark matters in contemporary astrophysics: A case study in theory choice and evidential reasoning, Ph.D. diss., University of Western Ontario.

——— (in preparation), "Astrophysical Dark Matter: An Introduction to its History and Philosophical Significance".

van Stratten, W., M. Bailes, M. Britton, S. R. Kulkarni, S. B. Anderson, R. N. Manchester, and J. Sarkissian (2001), "A Test of General Relativity from the Three-Dimensional Orbital Geometry of a Binary Pulsar", *Nature* 412: 158–160.

Walker, Mark (1994), "Lensing in Alternative Gravity", *Astrophysical Journal* 430: 463–466.

Will, Clifford M. (1993), *Theory and Experiment in Gravitation Physics,* rev. ed. Cambridge: Cambridge University Press.