

The Methodological Value of Coincidences: Further Remarks on Dark Matter and the Astrophysical Warrant for General Relativity

William L. Vanderburgh[†]

Four techniques for measuring the masses of galaxies and larger astrophysical systems from their dynamics are discussed. Their apparent agreement is sometimes invoked as warrant for postulating huge quantities of ‘dark matter’ as the best solution to “the dynamical discrepancy,” the disparity between the amount of mass visible in large scale astrophysical systems and the amount calculated from dynamics. This paper argues that the agreement, though suggestive, is not definitive. The coincident measurements remain the best reason for preferring dark matter over revisions to General Relativity for solving the dynamical discrepancy, but the preference is only weakly warranted.

1. Introduction. This paper is a follow-up to Vanderburgh 2003, which explores the evidential warrant for accepting General Relativity as the theory of gravity applicable to phenomena at the scale of galaxies and larger structures. That paper describes the “dynamical discrepancy” for galaxies, clusters of galaxies and other large scale systems—popularly known as the “dark matter problem”—and argues that in light of the dynamical discrepancy, the warrant for GR at galactic and greater scales is relatively weak. The present paper addresses some additional considerations, particularly the issue of whether the apparent agreement between four different kinds of measurements of the masses of galaxies and larger structures constitutes a good evidential or methodological basis for preferring General Relativity over a particular class of rival gravitation theories. The conclusion is that these coincident measurements do provide

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

Philosophy of Science, 72 (December 2005) pp. 1324–1335. 0031-8248/2005/7205-0055\$10.00
Copyright 2005 by the Philosophy of Science Association. All rights reserved.

some grounds, albeit weak and defeasible grounds, for thinking that GR is better than its rivals at galactic and greater scales.

General Relativity (hereafter GR) makes predictions that agree to high precision with all of the available evidence regarding interactions that take place over scales corresponding to roughly the size of our solar system (what I will call “stellar system scales”). But in fact GR also has a much stronger kind of confirmation at stellar system scales: the observed phenomena can be used to measure the parameters of the theory. These measurements from phenomena are the basis for epistemically robust theory comparisons via the Parametrized Post-Newtonian (or PPN) framework. The result of applying the PPN framework is that, at stellar system scales at least, GR has a very high degree of empirical support; it is clearly better than every rival gravitation theory so far articulated. (See Harper and DiSalle 1996, Will 1993, and Earman 1992, 177ff.) On the basis of these stellar system scale successes, GR is inductively extended to cover all (non-quantum) phenomena. This is analogous to the pattern of reasoning used by Newton to establish Universal Gravitation, wherein diverse phenomena—including pendulums, the orbit of the Moon, the orbits of the planets, the orbits of the Jovian moons, and so on—are used to make independent and agreeing measurements that the power law of the force of gravity is inverse square, and then the inverse square law is inductively extended to cover all phenomena whatsoever.

For interactions taking place at stellar system scales, then, GR is highly confirmed. For interactions taking place over galactic and greater scales,¹ however, the situation is quite different. The observed motions within galaxies, clusters of galaxies and other large scale dynamical systems are actually inconsistent with the predictions of General Relativity given the amount and distribution of mass observed in those systems.

The dynamical discrepancy for galaxies and larger structures is analogous to the discrepancy discovered in Uranus’ orbit in the early 1800s. The observed motions of Uranus were found to be inconsistent with the predictions of Universal Gravity given the then-known distribution of mass in the solar system. Two possible solutions were available: either modify Universal Gravity, or posit the existence of previously unknown mass. The latter, of course, was the kind of solution pursued independently

1. Binary pulsars, the decreasing periods of which provide an important test of GR, are included under stellar system scale phenomena. As Vanderburgh 2003, 814–815, describes, the available tests of GR are all for interactions taking place over (much) less than one light-year, whereas galaxies have radii in the range 10^4 to 10^5 light-years or larger, and the distance between the galaxies in clusters averages 10^6 light-years or more. The available tests of GR thus probe only the tiniest fraction of the relevant distances.

by Adams and by Leverrier; in 1846 LeVerrier's prediction of the geocentric position of the unknown mass was good enough to lead to the telescopic discovery of Neptune. (See Standage 2000.) The case of Mercury's excess perihelion precession is similarly analogous to the astrophysical dynamical discrepancy. For Mercury, however, the solution turned out to be a new theory of gravity, namely GR. (See Earman and Janssen 1993.) Note that solutions of both types were tried for both problems.

Just as for Uranus and Mercury, there are two classes of possible solutions to the astrophysical dynamical discrepancy. The members of the first class, the *matter solutions*, postulate the existence of about 100 times more mass than is visible in astrophysical systems. The distribution of this mass, on the assumption that it exists, is fairly easy to determine from the observed motions. It is more difficult to say what this matter is. It is called "dark" matter because it neither emits nor absorbs detectable electromagnetic radiation at any wavelength. Dark matter has so far eluded direct detection despite thirty years of active searching (the dynamical discrepancy itself was first discovered in the late 1920s, but it wasn't taken seriously by the astronomical community until the mid 1970s). A plethora of dark matter candidates have been proposed, ranging from otherwise unknown fundamental particles to black holes. Many of the candidates have been eliminated because they were found to conflict with empirical or theoretical considerations; others have been shown to be unable to resolve the entire discrepancy. The matter candidates that remain viable have little to no positive empirical support. Just about the strongest claim that can be made is that it is not impossible, so far as we can tell at present, that the remaining candidates could resolve the dynamical discrepancy.

The members of the second class of possible solutions, the *gravity solutions*, postulate no unseen matter and instead modify the action of gravity at large scales. There is no empirical reason to think that a matter solution is more probable than a gravity solution, or vice versa, just as in the Uranus and Mercury cases it was impossible to tell in advance which type of solution would ultimately succeed. The empirical constraints on theories of gravity offered as solutions to the dynamical discrepancy are surprisingly weak. Obviously, because of the epistemically robust solar system tests of GR, gravity solutions must be empirically equivalent to GR at stellar system scales. But at larger scales, their predictions may diverge, even radically, from GR's. One possibility is that GR will turn out to be the "stellar system scale limit" of some successor relativistic gravitation theory, in the same way that Newton's Universal Gravity turned out to be the "low velocity, weak field limit" of GR.

How can we decide between GR and the potential gravity solutions to

the dynamical discrepancy? It would be ideal to construct a theory testing and comparison framework analogous to the PPN formalism for this purpose. Unfortunately, the following difficulty, which I call “the dark matter double bind,” seems to preclude the possibility of constructing such a testing framework:

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. (Vanderburgh 2003, 824)

2. Coincident Measurements. Harper and DiSalle 1996 argue convincingly that Newtonian methodological ideals inform testing in current gravitation physics. An important part of Newton’s use of Reasoning from Phenomena in the argument for Universal Gravitation is that diverse phenomena yield precisely agreeing measurements of parameters of the theory. The coincidence of these measurements lends strength both to the unification of apparently diverse phenomena under a single gravitational law, and to the inductive extension of that law to all possible cases.² The methodological value of such coincident measurements, then, is that the coincidence is the foundation for stronger arguments in favor of a theory than would be possible without it.

There is a set of measurements of the masses of galaxies and larger structures that appear to provide independent coincident results. The question then arises whether the coincidence of these measurements does for GR at galactic and greater scales what the coincident measures at stellar system scales do. The “luminous mass” of a galaxy or larger system is obtained by measuring its total luminosity at all wavelengths and then comparing that result to a “mass-to-light ratio” derived from a combination of empirical and theoretical considerations. The “dynamical mass” is calculated by one of four methods: rotation curves, velocity dispersions, X-ray temperatures and gravitational lensing. Different techniques are

2. The unification argument is that since the same force law governs each of the phenomena, therefore the very same force—gravity—is at work in each case. The inductive extension or generalization of the results of reasoning from phenomena says that the law found for all phenomena studied so far should be taken to be the law governing all phenomena whatsoever.

used for different types of systems. But no matter which techniques are used, a discrepancy is always found between the dynamical mass and the luminous mass; moreover, the four techniques yield apparently agreeing measurements of the masses of large scale astrophysical systems. Since the techniques seem to be independent of one another, the coincidence of their results is taken to provide grounds for thinking that the astrophysical dynamical discrepancy will have a matter solution rather than a gravity solution. However, as I argue below, the apparent coincidence of the measures is not as evidentially or methodologically significant as it is sometimes thought to be. That said, these measurements taken together are still the best available reason for preferring matter solutions over gravity solutions, and thus for thinking that GR rather than some rival gravitational theory correctly describes the action of gravity at galactic and greater scales. Let me now discuss each of these mass measurement techniques in more detail.

3. Details of the Four Mass Measurement Techniques. First, for spiral galaxies, which have a well-defined plane and sense of rotation, a “rotation curve” can be taken, and from the rotation curve an overall mass distribution for the system can be inferred. By the well-known relationship between an object’s velocity along the line of sight and the Doppler shifting of its spectrum, the observed “redshifts” of stars and clouds of gas give the component of rotation along the line of sight at a given radius from the galactic center. A graph of redshift against radius yields the rotation curve. The mass interior to any given radius can be calculated from the rotational velocity of objects orbiting at that radius. The principle is the same as measuring the mass of the Sun from the radius and speed of the orbit of a planet around it. The observed rotation curves for spiral galaxies show that the absolute value of the rotation at any given radius is much higher than predicted given the luminous mass and the Newtonian limit of GR. Even more importantly, instead of eventually falling off asymptotically to zero, observations of gas clouds show that the rotation velocity of spirals actually remains flat or even *rises* out to several times the radius of the visible disk of stars. (See Figures 1a through 1d.)

On the assumption that the Newtonian limit of GR correctly describes the action of gravity at galactic and greater scales, the only way to account for the observed rotation curves is to conjecture the existence of a spherical “halo” of invisible matter surrounding every spiral, where the halo extends to several times the disk radius and contains many times the luminous mass. The alternative to this extravagant excess of invisible mass of unknown type is to propose a new account of gravity at galactic and greater scales.

Parallel results are found for elliptical galaxies and clusters (in which

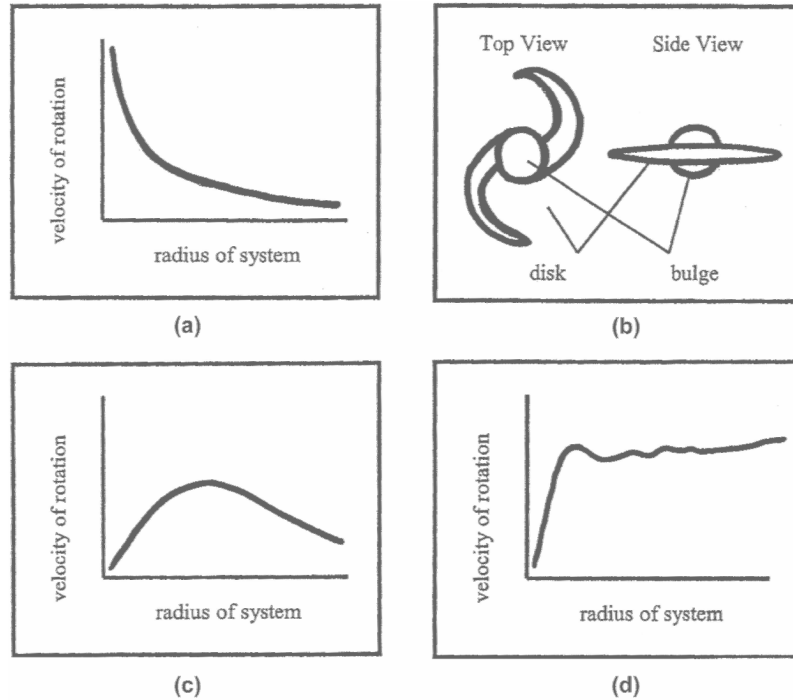


Figure 1. (a) Solar system rotation curve; (b) typical spiral galaxy; (c) Newtonian expectation for spiral galaxy rotation; (d) observed spiral galaxy rotation.

the internal motions are essentially random) by the second mass measurement method, velocity dispersions. Velocity dispersions are obtained spectroscopically as well. The collection of relative redshifts of the stars in elliptical galaxies, or of the galaxies in clusters, gives information about the motions within those systems. The Virial Theorem, originally developed in thermodynamics from principles of Newtonian mechanics, can then be applied to determine the gravitational potential needed to produce the observed velocities, and this yields a value for the total mass of the system. The Virial Theorem is $m = r\langle v^2 \rangle / \alpha G$, 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). This is the most common way of calculating the dynamical masses of large scale astrophysical systems.

Observations reveal X-ray emissions from clouds of diffuse gas enveloping many galaxies and clusters of galaxies. This leads to the third mass

measurement technique, which operates on the assumption that the gravitation potential of the system is the only plausible mechanism for continually heating the gas, as is required for the continual emission of X-rays. The intensity and spectrum of the X-ray emissions determines the amount of heating needed, and this in turn is converted into a value for the gravitational potential, and hence the mass, of the system in question. As in the case of rotation curves and velocity dispersions, the masses found by the X-ray method are roughly 100 times greater than the luminous masses of the system types in question.

The three previous measurements techniques are in practice performed using only Newtonian physics—the systems in question satisfy the weak-field, low-velocity limit, and thus no relativistic contributions are expected. In contrast, the fourth mass measurement technique, gravitational lensing, depends on the specifically relativistic parts of GR. In rare instances, a foreground object (a galaxy or cluster) is observed to lie along the line of sight to a background object (a galaxy, cluster or quasar). The gravitational field of the foreground object deflects the light of the background object. From the appearance of the image of the background object, the mass of the foreground object can be estimated. Gravitational lensing is well understood theoretically: given that GR is correct, the image patterns that will be produced by different magnitudes and configurations of mass in the “lens” (the foreground object whose gravity deflects the background light) given different distances and alignments between observer, lens and background object, can be predicted. Conversely, from an observed image pattern, the mass of the lens can be inferred.

With each of these four methods, a general trend has been found wherein larger systems tend to have a higher ratio of dynamical mass to luminous mass. This is true both within and across system types: larger spirals have a larger dynamical discrepancy than smaller spirals, and clusters have a larger discrepancy than individual galaxies.³ The methods, moreover, agree with each other about the masses of systems of any given type and given parameters. The overall results are summarized in Table 1. As is typical, the results there are reported as mass-to-luminosity (M/L) ratios.⁴ The Sun (an average star) is defined as having $M/L = 1.0$.

3. A likely explanation of this trend is that in the early universe locations where dark matter was more concentrated attracted more baryonic matter, which later became the luminous parts of galaxies and clusters. Or it could be that something else such as initial density fluctuations seed galaxy formation, and that baryonic matter and dark matter are both preferentially attracted to the larger seeds.

4. Astronomers use M/L ratios instead of absolute masses because they are more accurate than absolute masses, and are stable even when error-prone data such as the true distance to a system is revised (whereas the calculations of the absolute visible and dynamical masses are sensitive to such changes).

TABLE 1.

Object: Size Scale	Methods	M/L	Ω -Contribution
Binary stars, star clusters (AU to a few pc)	Orbit velocities; velocity dispersions; stellar structure models	.5–3	~.003
Galactic disks and nuclei (1–10 kpc)	Stellar velocity dispersions; rotation curves	3–10	.003–.01
Binary galaxies; small groups (.1–2 Mpc)	Velocity differences; X-ray temperatures; orbit modelling	20–50	.02–.05
Rich clusters (1–10 Mpc)	Virial theorem; X-ray temperatures; kinematic models	50–200	.05–.20
Superclusters (~100 Mpc)	Virial theorem; models of Virgocentric in-fall, kinematic evolution, and mergers	100–400	.10–.40

Because massive stars are not only much brighter (roughly $L \propto M^3$) than little ones but also much rarer (N proportional to about M^{-2} over most of the range 0.3–60 [solar masses]), a typical stellar population will also have M/L near 1. Values of 0.5–3.0 are, in fact, observed for star clusters of varying ages. The mass in gas is less than or, at most, equal to the mass in stars in all common varieties of galaxies and clusters of galaxies. *An object with M/L much greater than 3 must, then, be regarded as containing significant dark matter.* (Trimble 1993, 151; italics added)

In order to avoid begging the question in favor of matter solutions, Trimble should have said that an object with an M/L ratio much greater than 3 must be regarded as having a significant dynamical discrepancy. Table 1, with a few small changes, is taken from Trimble 1993, 150. The rightmost column lists the fraction of the total mass density of the universe (Ω) contributed by each class of objects; each Ω value includes the contributions listed above it (the total dynamical mass contribution to Ω is just the value on the last line, not the sum of the column). (“AU” means “astronomical unit,” the average distance between the Earth and the Sun, which is about 93 million miles or 8.3 light-minutes; “pc” means “parsec,” 3.26 light-years; “kpc” and “Mpc” are thousands and millions of parsecs respectively.)

Advocates of gravity solutions⁵ to the dynamical discrepancy might try to discount the agreement between the four different mass measurement techniques discussed here as providing evidence for the existence of dark matter by claiming (as I once heard Philip Mannheim claim) that the measurements must all be wrong by the same amount. It is hard to see how this could plausibly be the case, however. A better way to challenge the evidential or methodological value of these measurements is to show that the agreement between them is much less close than it first appears to be. In this way it becomes obvious that there is plenty of room for gravity solutions to the dynamical discrepancy.

The main point is that often it is impossible to apply more than one method to any one system. Rotation curves, for example, are possible for spirals but not ellipticals, and the velocity dispersion technique applies to ellipticals but not to spirals. There are relatively few cases of gravitational lensing known. The available gravitational lensing results do find that the masses of spirals that act as lenses are of the same order of magnitude as is typically found by rotation curves for other spirals with similar diameters and luminosities. Similarly, where the lens is an elliptical galaxy the mass derived from the lensed image agrees to within an order of magnitude with the typical masses found by velocity dispersions for ellipticals whose parameters are similar to those of the lens in question. But (so far as I am aware) no single galaxy has had its mass determined both through gravitational lensing and through either rotation curves or velocity dispersions. (Systems that act as lenses are normally extremely distant, and hence too dim for the detailed spectroscopic work necessary to construct a rotation curve or velocity dispersion.) This means that the “agreement” between rotation curves or velocity dispersions and gravitational lensing is based on an analogical argument. All there is, then, is *order of magnitude* agreement for *analogous* systems, rather than precise agreement in individual cases. This is suggestive, but it is surely not definitive. It is unlikely, moreover, that the agreement can be made much better, even with further observations and improved techniques. Fairly large margins of error are present in lensing calculations because of the need to make assumptions that cannot be definitively checked observationally—including assumptions about the diameter of the lens, the shape

5. Vanderburgh (2003, 820) mentions two such attempted gravity solutions: Mannheim’s “Weyl gravity” was there said to be non-viable for empirical reasons, and Milgrom’s “MOND” (Modification of Newtonian Dynamics) was said to be non-viable in part because it was explicitly non-relativistic. The latter complaint has now been eliminated: Bekenstein (2004) has proposed a relativistic version of MOND which he calls TeVeS. Evaluating TeVeS is beyond the scope of the present paper.

of its overall mass distribution, its distance from us, and its distance from the background object.

It is true that many galaxies and clusters have their masses estimated both from velocity dispersions and from X-ray temperatures, and that the results essentially agree. But the margins of error are not small in either kind of measurement. Even if the margins of error can be made smaller, two other possible objections can be raised. The first is that X-ray temperatures and velocity dispersions are based on the same Newtonian principles, and hence it could be argued that they are not really independent methods. The second and more telling objection is that if an alternative gravitational law were assumed in these techniques, agreeing measurements of a different amount of mass might well be found. This is to say that the mass results from velocity dispersions and X-ray temperatures are model dependent. Thus the fact that their results agree in a given case could indicate that the correct value for the dynamical mass has been found, or it could be construed as an agreeing measurement of the parameters of an alternative theory of gravitation. In short, by itself the agreement between velocity dispersion and X-ray mass measurements only confirms the overall size and character of the dynamical discrepancy itself—it does not give preferential support to the hypothesis that hidden mass is the cause of the discrepancy.

As a final note, let me mention a recent observational study. Buote et al. 2002 reported that in a few galaxies the orientation of the cloud of X-ray gas is different from the orientation of the luminous matter. This is an interesting result because it implies that there must really be a halo of dark mass surrounding the luminous matter. On theoretical grounds it is clear that the shape of the gravitational potential that heats the gas must be the same as the shape of the overall mass distribution. So the fact that the X-ray-emitting gas cloud is oriented differently than the luminous mass indicates that there must exist a distribution of hidden mass in the galaxy, that it dominates the luminous mass, and that it is oriented in the same pattern as the X-ray cloud. If robust, this result shows that there really is dark matter. Note, however, that this can be true even if the correct theory of gravity for galaxies and larger structures is not GR. That is, even these observations are perfectly consistent with alternative theories of gravity: they do not prove that the dark matter that is needed to heat the X-ray emitting gas is the entire solution to the dynamical discrepancy. These observations leave open a complex solution involving both dark matter and a new law of gravitation. That is an ugly solution that no one wants, but the universe has defeated our expectations more than once, and the question I am interested in here is whether there are evidential or methodological reasons to prefer a “non-ugly” solution: it seems that such reasons are few and relatively weak.

4. Conclusion. The upshot of all this is that the four methods for measuring dynamical mass at galactic and greater scales do not really yield closely agreeing measurements of the masses of large scale astrophysical systems. The agreement we get from these techniques is somewhat loose and could be merely coincidental. And whatever agreement is present, it certainly does not have the same epistemically robust character as the multiple, agreeing and precise measurements of the parameters of GR that are obtained from solar system phenomena via the PPN formalism.

It cannot be denied, however, that the four methods discussed here do give roughly agreeing results. The roughness of the agreement must be taken into account, as should the fact that what is being measured is the mass of these systems *on the assumption that GR applies to these systems*. Strictly speaking, these techniques give us agreeing measures (insofar as they are agreeing) of the value of the total forces produced by the compound of “mass distribution plus gravitational law”. The techniques do not tell us the relative contributions of those two components: the measured values could result from just the visible matter plus a new law of gravity at large scales, or from the Newtonian limit of GR plus dark matter, or from some combination of new matter and new gravity. That said, my intuitions are the same as those of most physicists: the available evidence seems more likely to yield to a matter solution than a gravity solution, which is to say that on balance it seems more likely that GR is the correct theory of gravitational interactions at galactic and greater scales. That position is not strongly warranted, however, and a gravity solution to the astrophysical dynamical discrepancy remains open.

REFERENCES

- Bekenstein, Jacob D. (2004), “Relativistic Gravitation Theory for the MOND Paradigm”, *Physical Review D* 70: 083509–083543.
- 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.
- Earman, John (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: 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, 130–131.
- Harper, William, and Robert DiSalle (1996), “Inferences from Phenomena in Gravitational Physics”, *Philosophy of Science* 63: S46–S54.
- Standage, Tom (2000), *The Neptune File: A Story of Astronomical Rivalry and the Pioneers of Planet Hunting*. New York: Walker.
- Taylor, Roger J. (1991), *The Hidden Universe*. Chichester, U.K.: Ellis Horwood.
- Trimble, Virginia (1993), “Dark Matter”, in Noriss S. Hetherington (ed.), *Encyclopedia of Cosmology: Historical, Philosophical and Scientific Foundations of Modern Cosmology*. New York: Garland, 148–158.

- Vanderburgh, William L. (2003), "The Dark Matter Double Bind: Astrophysical Aspects of the Evidential Warrant for General Relativity", *Philosophy of Science* 70: 812–832.
- Will, Clifford M. (1993), *Theory and Experiment in Gravitation Physics*. Cambridge: Cambridge University Press.