# Chapter 3
# Effective Theories of Classical Gravity

**Abstract** If the concepts underlying Effective Theory were appreciated from the earliest days of Newtonian gravity, Le Verrier's announcement in 1845 of the anomalous perihelion precession of Mercury would have been no surprise. Furthermore, the size of the effect could have been anticipated through "naturalness" arguments well before the definitive computation in General Relativity. Thus, we have an illustration of how Effective Theory concepts can guide us in extending our knowledge to "new physics", and not just in how to reduce larger theories to restricted (e.g., lower energy) domains.

## 3.1 Introduction

The purpose of these lectures is to introduce the concepts of Effective Theories to students of Philosophy, Mathematics and Physics who have a shared interest in the philosophy and history of physics. The concept I wish to discuss, Effective Theory, is a thoroughly modern notion; nevertheless, I wish to illustrate it with a very old and intuitively accessible problem in physics: Mercury's anomalous perihelion precession.

Le Verrier announced in 1845 a small discrepancy in the precession rate of Mercury's perihelion compared to Newton's theory, even after taking into account all the disturbing influences throughout the solar system such as the effect of other planets' orbits.[1] This came as a surprise, and more or less nobody believed at the time that it was the fault of Newton, but rather the fault of observers who had not seen the other celestial bodies that must surely be perturbing Mercury's orbit. Historically, that is the beginning of the problem. Le Verrier believed that an as-yet unobserved mass distribution inside the orbit of Mercury was the source

---

---

[1] In 1859 Le Verrier gave a number for this advance: 35 arcseconds per century (Le Verrier 1859). It was later reevaluated by S. Newcomb (Newcomb 1882), who determined the correct value of 43 arcseconds per century.

of the discrepancy. He and others advocated the existence, for example, of a new small planet ("Vulcan" as it was sometimes called) that would be observed when astronomers developed the instruments necessary to find it (Roseveare 1982). Such was not the case. By the 1890's it became clear to most that new large-scale object(s) was not the explanation (Oppenheim 1920), despite some ill-fated protestations otherwise (Poor 1921). The resolution of the problem came with Einstein's General Relativity, which predicted precisely the 43" of arc per century observed, and the case was closed.

However, I want to argue that anticipation of the "problem" could have occurred much before Le Verrier. What prevented scientists from anticipating Mercury's perihelion precession was not lack of mathematical skill, or lack of experimental abilities. It was solely due to not having the right mindset. Unlike perhaps in decades and centuries gone by, no competent scientist should retain an unfailing commitment to any theory. All theories are incomplete, even given that some theories are better than others. The code phrase of this mindset is Effective Theories. The concept is a powerful one that has born much fruit in theories of particle physics, condensed matter systems, and even cosmology.

These notes are meant to be a somewhat pedagogical and technical exposition of the Mercury problem and the application of Effective Theory ideas to the problem. In some parts of this lecture I will follow an "alternative history" path with the scientists Alice and Bob who vaguely understand the importance of Effective Theories and who will devise a theory that can accomodate the perihelion precession rate well before Einstein's General Relativity comes along, and may even be able to predict roughly the numerical rate of the precision and make predictions for other planets through "naturalness" arguments. The latter could have been possible after diligent reflections on the philosophical challenges of Newton's theory. I will compute the General Relativity rate at the end, in order to show how elegantly it comes out of that more complete theory, and to show that it matches the Effective Theory "predictions" by Bob and Alice. And finally I will conclude with some more remarks on the meaning of the results.

## 3.2 Orbits in Newton's Theory

To remind some students who have not seen celestial mechanics for some time, we begin with the computation of particle orbits in Newton's gravity. The reader familiar with these basics should feel free to skim the section only for definitions and conventions that I will use later.

We know that the orbits predicted by Newton's law of gravity are respected quite well by the planets, and so any change in the equations of motion for the orbits will need to be small perturbations. In Newton's gravity, a test particle with mass $m$ orbits around a particle of mass $M \gg m$ according to the equations of motion derived from

the lagrangian

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 + \frac{\alpha}{r} \tag{3.1}$$

where $\alpha = GMm$ with $G$ being Newton's constant. $M$ represents the sun's mass in this lecture, and $m$ the planet's mass, most often Mercury. It is appropriate to assume that $M \gg m$ such that any correction is negligible due to the difference of $m$ from the reduced mass $\mu = Mm/(M + m)$, which is technically the precise mass one should use in the kinetic energy term. Lagrange's equations of motion are

$$m\ddot{\mathbf{r}} = -\frac{\alpha}{r^2}\hat{\mathbf{r}} \tag{3.2}$$

where $\hat{\mathbf{r}}$ is the unit vector in the $\mathbf{r}$ direction.

### 3.2.1 Orbital Solution

The lagrangian is rotationally invariant, and so the motion of the particle is most conveniently evaluated by casting the vector equation of motion into the two polar component equations

$$m(\ddot{r} - r\dot{\phi}^2) = -\frac{\alpha}{r^2} \quad \text{(radial equation)} \tag{3.3}$$

$$m(2\dot{r}\dot{\phi} + r\ddot{\phi}) = 0 \quad \text{(angular equation)}. \tag{3.4}$$

The second equation is equivalent to

$$\frac{d}{dt}(mr^2\dot{\phi}) = 0 \tag{3.5}$$

which implies that $mr^2\dot{\phi}$ is a constant in time. At the apogee (furthest) or perigee (closest) point of the orbit the radius vector $\hat{\mathbf{r}}$ is exactly perpendicular to the angular vector $\hat{\phi}$ and the magnitude of the angular momentum vector $\ell = \mathbf{r} \times \mathbf{p}$, where $\mathbf{p} = mr\dot{\phi}\hat{\phi}$, is exactly $mr^2\dot{\phi}$. Since angular momentum is conserved and $mr^2\dot{\phi}$ is conserved, if they are equal at one point they are equal at all points in the orbit. Thus, the constant value inside the time derivative Eq. (3.5) is none other than angular momentum: $\ell = mr^2\dot{\phi}$. This also proves that the motion is in a plane. Since angular momentum is a conserved vector quantity, the direction must also be preserved which is only possible if $\mathbf{p}$ perpetually lies in the same plane as $\mathbf{r}$. This justifies our evaluation of a three-dimensional problem in terms of just two variables $(r, \phi)$ in the plane of motion.

Let us now solve the radial differential equation to obtain an exact solution of the orbit for particle $m$. By rewriting $r \equiv 1/u$, recasting all time derivatives as $d/dt \to \dot{\phi} d/d\phi$ when possible, and recognizing that $\dot{\phi} = l/r^2 m$ from conservation of angular momentum, one finds that the governing differential equation of motion is

$$\frac{d^2 u}{d\phi^2} + u = \frac{\alpha m}{\ell^2}. \tag{3.6}$$

Interestingly, this equation takes the form of the harmonic oscillator equation. The solution is

$$u(\phi) = u_0 \cos \phi + \frac{\alpha m}{\ell^2}, \tag{3.7}$$

where $u_0$ is a constant that is not determined by the theory but the particular circumstances (i.e., initial conditions) of the system. In terms of the more direct variable $r$, the solution is

$$r(\phi) = \frac{\rho}{1 + e \cos \phi}, \quad \text{where } \rho = \frac{\ell^2}{\alpha m}, \text{ and } e = u_0 \rho, \tag{3.8}$$

and it is assumed that $\phi = 0$ is at perigee. $\rho$ is sometimes called the lactus rectum of the orbit.

The constant $e$ is called the eccentricity with which one can classify an orbit as circular ($e = 0$), elliptical ($0 < e < 1$), parabolic ($e = 1$), or hyperbolic ($e > 1$). Focusing on the $0 \le e < 1$ case of elliptical or circular orbits, we find that

$$r_{\min} = \frac{\rho}{1 + e}, \quad \text{and } r_{\max} = \frac{\rho}{1 - e}. \tag{3.9}$$

The relation between the semimajor axis $a$ of the elliptical orbit and the other variables is given by

$$a = \frac{r_{\min} + r_{\max}}{2} \quad \text{which implies} \quad \rho = a(1 - e^2). \tag{3.10}$$

### 3.2.2 The Hamiltonian and $V_{eff}$ Description

An alternative way to approach the problem is to compute the Hamiltonian and consider the orbit from the perspective of a one-dimensional effective potential for radial motion. I provide the very basics of this to remind the students of the formalism which is used by some papers relevant to the perihelion precession. We first expand

the lagrangian in terms of radial and angular coordinates starting from the identity

$$\dot{\mathbf{r}}^2 = \dot{r}^2 + r^2 \sin^2 \theta \dot{\phi}^2 + r^2 \dot{\theta}^2$$
$$\implies \dot{r}^2 + r^2 \dot{\phi}^2 \text{ (valid in the } \sin \theta = 1 \text{ fixed orbital plane)} \qquad (3.11)$$

The Hamiltonian is constructed as

$$H = \sum_i \dot{q}_i \, p_i - L \qquad (3.12)$$

using the momentum factors

$$p_r = \frac{\partial L}{\partial \dot{r}} = m\dot{r}, \text{ and } p_\phi = \frac{\partial L}{\partial \dot{\phi}} = mr^2 \dot{\phi}, \qquad (3.13)$$

which implies

$$H = \frac{p_r^2}{2m} + \frac{p_\phi^2}{2mr^2} - \frac{\alpha}{r}. \qquad (3.14)$$

The Hamiltonian is independent of $\phi$, which implies from Hamilton's equations of motion,

$$\dot{p} = -\frac{\partial H}{\partial q}, \text{ and } \dot{q} = \frac{\partial H}{\partial p}, \qquad (3.15)$$

that $\dot{p}_\phi$ is a conserved quantity:

$$\dot{p}_\phi = \frac{\partial H}{\partial \phi} = 0 \implies p_\phi = \text{const.} \qquad (3.16)$$

This of course is just a restatement of the conservation of angular momentum

$$\ell = p_\phi = mr^2 \dot{\phi}. \qquad (3.17)$$

We can substitute Eq. 3.17 back into Eq. 3.14, which gives a one-dimensional Hamiltonian as promised:

$$H = \frac{p_r^2}{2m} + \frac{\ell^2}{2mr^2} - \frac{\alpha}{r}. \qquad (3.18)$$

The Hamiltonian is a constant of the motion—the energy of the system—and it is useful sometimes to consider the dynamics of particle motion from this considera-tion where $E \equiv H = T + V$ and $T$ and $V$ are the kinetic and potential energies respectively. Here the potential for the one dimensional motion is often called the

effective potential and is given by

$$V_{eff}(r) = \frac{\ell^2}{2mr^2} - \frac{\alpha}{r}. \tag{3.19}$$

It is this potential that governs the radial potential with the first term pushing the particle away from the origin and the second term attracting the particle to the origin. The balance giving orbital motion between two turning points of zero radial kinetic energy, the apogee and perigee.

## 3.3 Perihelion Precessions from Perturbations

From the previous section we know that the orbit from Newton's simple $1/r^2$ force law is

$$u(\phi) = \frac{1}{r(\phi)} = \frac{1}{\rho}(1 + e\cos\phi). \tag{3.20}$$

This obviously does not allow any advancement of the perihelion. The minimum is where $du/d\phi = 0$, which implies $\sin\phi = 0$ and therefore $\phi = 0, 2\pi, 4\pi, \ldots$ mark the successive perihelions. The discovery of the anomalous perihelion precession of Mercury, if it can be established, would signal the end of the Newtonian era and initiate the search for a better theory. As the reader is no doubt aware, perihelion precessions exist for every planet's orbit (see Table 3.1), but for the present let us continue on our theoretical discussion.

### 3.3.1  $1/r^2$ Correction to the Central Potential

Let us look at how the orbits change if we add a $1/r^2$ correction to the potential of the gravitational interaction lagrangian. Let us call this Bob's theory with lagrangian

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 + \frac{\alpha}{r}\left(1 + \frac{R_{\text{bob}}}{r}\right) \tag{3.21}$$

where $\alpha = GMm$, with $G$ being Newton's constant, $M$ is the mass of the sun, and $m$ is the mass of the planet under consideration. This new law requires the introduction of a new fundamental length scale $R_{\text{bob}}$, which is a priori unknown. However, we do know, as will be shown below, that it leads to a perihelion precession of the orbits governed by this law.

Lagrange's equation of motion for this theory is

$$\text{radial} : m(\ddot{r} - r\dot{\phi}^2) = -\frac{\alpha}{r^2}\left(1 + \frac{2R_{\text{bob}}}{r}\right) \tag{3.22}$$

$$\text{angular} : m(2\dot{r}\dot{\phi} + r\ddot{\phi}) = 0 \tag{3.23}$$

The angular equation yields conservation of angular moment $\ell = mr^2\dot{\phi} = \text{const}$ just as before. Using this, we can write the radial differential equation as

$$\frac{d^2u}{d\phi^2} + u = \frac{\alpha m}{\ell^2}(1 + 2R_{\text{bob}}u) \tag{3.24}$$

This can be rewritten as

$$\frac{d^2u}{d\phi^2} + \left(1 - \frac{2R_{\text{bob}}}{\rho}\right)u = \frac{1}{\rho}, \quad \text{where} \ \rho = \frac{\ell^2}{\alpha m}. \tag{3.25}$$

The general solution to this equation, assuming perihelion is placed at $\phi = 0$, is

$$u(\phi) = u_0 \cos\left(\phi\sqrt{1 - \frac{2R_{\text{bob}}}{\rho}}\right) + \frac{1}{\rho - 2R_{\text{bob}}}, \tag{3.26}$$

or, written differently,

$$u(\phi) = \left(\frac{1}{\rho - 2R_{\text{bob}}}\right)\left[e\cos\left(\phi\sqrt{1 - \frac{2R_{\text{bob}}}{\rho}}\right) + 1\right] \tag{3.27}$$

where $e = u_0(\rho - 2R)$.

The $u(\phi)$ solution describes the motion of a precessing ellipse. The first perihelion by definition is at $\phi = 0$ and the second perihelion occurs when

$$\phi\sqrt{1 - \frac{2R_{\text{bob}}}{\rho}} = 2\pi \implies \phi = \frac{2\pi}{\sqrt{1 - \frac{2R_{\text{bob}}}{\rho}}} = 2\pi + 2\pi\frac{R_{\text{bob}}}{\rho} \tag{3.28}$$

The small perihelion advance is the deviation of $\phi$ from $2\pi$ and is $\delta = 2\pi R_{\text{bob}}/\rho$.

Given our previous computations, we are now able to evaluate the relationship between the extra length scale $R_{\text{bob}}$ and the perihelion advance of an orbit. In one case, if we have made a measurement of the perihelion advance, we can derive what value $R_{\text{bob}}$ must be to reproduce that value

$$R_{\text{bob}} = (1.16\,\text{km})\left(\frac{\delta/T_{\text{orbit}}}{\text{arcsec}\cdot\text{century}^{-1}}\right)\left(\frac{\rho}{1\,\text{au}}\right)\left(\frac{T_{\text{orbit}}}{1\,\text{year}}\right) \tag{3.29}$$

where $\rho$ is related to the common parameters of the semimajor axis $a = (r_{\text{min}} + r_{\text{max}})/2$ and eccentricity $e$ by the relation $\rho = a(1 - e^2)$.

**Table 3.1** Data for planetary orbits

| Planet | $T_{orbit}$ (years) | $e$ | $a$ (au) | $\rho$ (au) | $r_{min}$ (au) | $r_{max}$ (au) |
|---|---|---|---|---|---|---|
| Mercury | 0.241 | 0.206 | 0.387 | 0.371 | 0.307 | 0.467 |
| Venus | 0.615 | 0.007 | 0.723 | 0.723 | 0.718 | 0.728 |
| Earth | 1.000 | 0.017 | 1.000 | 1.000 | 0.983 | 1.017 |
| Mars | 1.881 | 0.093 | 1.524 | 1.511 | 1.382 | 1.666 |
| Jupiter | 11.86 | 0.048 | 5.203 | 5.191 | 4.953 | 5.453 |
| Saturn | 29.46 | 0.056 | 9.539 | 9.509 | 9.005 | 10.07 |
| Uranus | 84.02 | 0.047 | 19.19 | 19.15 | 18.29 | 20.09 |
| Neptune | 164.8 | 0.009 | 30.06 | 30.06 | 29.79 | 30.33 |
| Pluto | 247.7 | 0.249 | 39.46 | 37.01 | 29.63 | 49.29 |

$T_{orbit}$ is the time for one full revolution in earth years, $e$ is the eccentricity of the orbit, $a$ is the semimajor axis in astronomical units (1 au $= 1.496 \times 10^{11}$ m), $\rho = \ell^2/GMm^2 = a(1 - e^2)$ is the orbital latus rectum in astronomical units (and is independent of $m$ ultimately), $r_{min} = a(1 - e)$ is the distance of perigee in astronomical units, and $r_{max} = a(1 + e)$ is the distance of apogee in astronomical units

On the other hand, if we have a theory for what $R_{bob}$ should be, we can make a prediction for the perihelion advance in units of arc seconds per century:

$$\frac{\delta}{T_{orbit}} = \frac{2\pi R_{bob}}{\rho T_{orbit}} = (0.866 \, \text{arcsec} \cdot \text{century}^{-1}) \left( \frac{1 \, \text{au}}{\rho} \right) \left( \frac{\text{years}}{T_{orbit}} \right) \left( \frac{R_{bob}}{1 \, \text{km}} \right) \quad (3.30)$$

### 3.3.2  $1/r^3$ Correction to the Central Potential

Alice's theory has a $1/r^3$ correction to the potential

$$L_{\text{alice}} = \frac{1}{2}m\dot{\mathbf{r}}^2 + \frac{\alpha}{r}\left( 1 + \frac{R_{\text{alice}}^2}{r^2} \right), \quad (3.31)$$

which gives a $1/r^4$ correction to the gravitational force law. Lagrange's equations for her theory are

$$\text{radial}: m(\ddot{r} - r\dot{\phi}^2) = -\frac{\alpha}{r^2}\left( 1 + \frac{3R_{\text{alice}}^2}{r^2} \right) \quad (3.32)$$

$$\text{angular}: m(2\dot{r}\dot{\phi} + r\ddot{\phi}) = 0 \quad (3.33)$$

Here again the angular momentum $\ell = mr^2\dot{\phi}$ is conserved from the angular equation, and the radial equation becomes

$$\frac{d^2u}{d\phi^2} + u = \frac{\alpha m}{\ell^2}(1 + 3R_{\text{alice}}^2 u^2). \tag{3.34}$$

We'll solve this equation employing techniques of perturbation theory. We treat the last term of Eq. 3.34 as a small perturbation and solve first the equation

$$\frac{d^2u}{d\phi^2} + u = \frac{\alpha m}{\ell^2} \tag{3.35}$$

which is just the standard Newtonian orbit solution

$$u_N(\phi) = \frac{1}{\rho}(1 + e\cos\phi), \quad \text{where } e = u_0\rho \tag{3.36}$$

where the subscript $N$ refers to the Newtonian solution, $u_0$ is an initial condition constant and $\rho = \ell^2/\alpha m$ is the usual value.

The next step is to now substitute $u \to u_N + \delta u$ into Eq. 3.34 where we only keep one order in perturbation theory. Since $u_N$ part of this expression cancels the usual part of the differential equation from Newton's law, we are left with a differential equation for the perturbation $\delta u$:

$$\frac{d^2\delta u}{d\phi^2} + \delta u = \frac{3}{\rho}R_{\text{alice}}^2 u_N^2(\phi) \tag{3.37}$$

$$= \frac{3}{\rho^3}R_{\text{alice}}^2(1 + 2e\cos\phi + e^2\cos^2\phi) \tag{3.38}$$

To obtain the complete solution we need to solve for $\delta u$. The theory of ordinary differential equations tells us that all we need is *any* particular solution, and here is one:

$$\delta u = \frac{3}{\rho^3}R_{\text{alice}}^2\left(1 + e\phi\sin\phi + \frac{e^2}{3}\cos 2\phi + e^2\sin^2\phi\right) \tag{3.39}$$

The perihelians of the orbit can be obtained by solving for $\phi$ in

$$\rho\frac{du}{d\phi} = -e\sin\phi + 3\frac{R_{\text{alice}}^2}{\rho^2}$$
$$\times\left(e\sin\phi + e\phi\cos\phi - \frac{2}{3}e^2\sin 2\phi + 2e^2\sin\phi\cos\phi\right) = 0 \tag{3.40}$$

The existence of the $\phi\cos\phi$ term in this equation, which came from the $\phi\sin\phi$ term in $\delta u$, is causing the perihelion on the next cycle to shift away from $2\pi$. Defining $\phi = 2\pi + \delta$ we can solve for $\delta$ in the perturbative expansion:

$$\delta = 6\pi\frac{R_{\text{alice}}^2}{\rho^2}. \tag{3.41}$$

Given our previous computations, we are now able to evaluate the relationship between the extra length scale $R_{\text{alice}}$ and the perihelion advance of an orbit. In one case, if we have made a measurement of the perihelion advance, we can derive what value $R_{\text{alice}}$ must be to reproduce that value

$$R_{\text{alice}} = (7.58 \times 10^6 \, \text{m}) \left(\frac{\rho}{\text{au}}\right) \left(\frac{T_{\text{orbit}}}{\text{years}}\right)^{1/2} \left(\frac{\delta/T_{\text{orbit}}}{\text{arcsec/century}}\right)^{1/2} \quad (3.42)$$

On the other hand, if we have a theory for what $R_{\text{alice}}$ should be, we can make a prediction for the perihelion advance in units of arc seconds per century:

$$\frac{\delta}{T} = \frac{6\pi R_{\text{alice}}^2}{\rho^2 T_{\text{orbit}}} = (1.74 \, \text{arcsec} \cdot \text{century}^{-1}) \left(\frac{R_{\text{alice}}}{10^7 \, \text{m}}\right)^2 \left(\frac{1 \, \text{au}}{\rho}\right)^2 \left(\frac{1 \, \text{yr}}{T_{\text{orbit}}}\right). \quad (3.43)$$

## 3.4 Philosophical Challenges to Newton's Theory

We pause here to describe some foundational questions that Newton's theory faced. There are three main philosophical problems: (1) What is the nature of absolute time and space, and is it necessary to invoke it? (2) Why should the gravitational mass be equal to the inertial mass in the equations of motion? And (3) how does nature enable action at a distance responses?

Regarding Absolute Space and Time, Newton sets forth his ideas in the first Scholium of *Principia*. Almost immediately upon the publication of his book, Newton faced criticism from noted physicists and mathematicians. The most famous adversary regarding this was Leibnitz, who claimed that the only thing that need be talked about, and which ultimately defines space and time, is the relative motions of objects (relativism). Appeals to absolutes make no sense. Newton's friend Samuel Clarke argued vociferously for the absolute viewpoint (substantivalism). Their correspondences are famous, and illuminating in the history of science. Over time these discussions progressed from what some might think is word quibbling to important physics principles emphasized by Mach and Einstein to name just two. Pedantic rigor of thinking can lead to the thought processes that generate significantly better theories, and this philosophical problem is arguably an illustration of that.

The second problem, why is the gravitational mass equal to the inertial mass in my mind is the problem that should have kept everyone sleepless for those many centuries when there was not an answer. Newton's theory has nothing to say on the matter, except *well, there it is*. These masses are two separate beasts, and why they should be the same? The resolution of this issue is one of the core motivating principles behind General Relativity, which succeeds in giving a deeper explanation for this curious equality.

The third philosophical problem is sometimes called the problem of action at a distance. There are two aspects of action at a distance. The first is why should two

bodies far removed from each other with nothing in between them feel gravitational attraction. Should not there be some "touching" or medium that carries the gravity force from one body to the other? This action at a distance occurs between particles separated by a large vacuum of nothing. This is hard to take. Even Newton was disturbed by it, especially the latter aspect. In 1693 he wrote his friend Richard Bentley "It is inconceivable that inanimate brute matter should, without the mediation of something else which is not material, operate upon and affect other matter without mutual contact ...." (Thayer 1953).

The second aspect of the problem, which is related to the first, is how can two bodies far removed from each other in space instantaneously feel the effect of another's gravitational force. Newton's theory implicitly assumes that all particles feel all other particles' gravitational attraction strength by the exact separations of those particles at each moment of time. If a particle moves just a little, everybody knows about it instantly and the resolution of forces are adjusted instantly. To Newton and others, action at a distance was intolerable, but the Newtonian system was the best thing going, and it had tremendous practical value, so it was not to be abandoned despite its flaws.

The issue of instantaneity was noted from the start, and Laplace touched upon it in his highly influential *Traité de Mécanique Céleste*, published from 1797 to 1825. He stated that instantaneous propagation did not appear convincing,[2] and noted that Bernoulli had suspicions as well. But Laplace knew that if the propagation were indeed finite it would have to be extraordinarily fast, and even suggested, incorrectly as it turns out, that some observations imply that it is eight million times that of light. Laplace briefly brought up the possibility of modifying the inverse square law based on this potential objection but ultimately dismissed it, stating that the simplicity of Newton's theory authorizes us to think of it as a rigorous law of nature.[3]

Nevertheless, the philosophical challenges to Newton's theory are enough to realize that it was not a complete theory. As we say often in physics today, there must be "physics beyond the Standard Model". How might signal of "new physics" show up beyond Newton's theory? Let us consider, for example, the disturbing underlying assumption of action at a distance. As we implied above, there are two different issues with action at a distance. There is the aspect of reaching across the mediumless vacuum, and there is the aspect of instantaneous transmission of information to all particles in the universe when one particle moves.

Transforming our theory from reaching across the vacuum action at a distance to action by local contact is the subject of the theory of fields. Particles source fields that permeate spacetime, and other particles experience those fields. Thus, action at a distance is replaced by particle-field interactions in this classical point of

---

[2] "La propagation instantanée qu'ils supposaient à cette force me parut peu vraisemblable" (Laplace 1805).

[3] "En général, on verra dans le cours de cet ouvrage que la loi de la gravitation réciproque au carré des distances représente avec une extrême précision toutes les inégalités observées des mouvement célestes: cet accord, joint à la simplicité de cette loi, nous autorise à penser qu'elle est rigoureusement celle de la nature" (Laplace 1805).

view. The emanating field propagates at finite velocity, which is incorporated self-consistently into modern field theories, retaining causality and introducing the more acceptable action by local contact.

We do not need to fast forward all the way to the field theories of today to ask how Newton's theory can be pressured experimentally by applying our philosophical worries. The most obvious way one should have thought to do it is by testing the instantaneous aspect of action at a distance. If one doubts that it is to be rigorously upheld, then we should expect that a quick movement of a body in a mechanical system might yield unexpected results since it might be significantly displaced from its original position by the time the other bodies "get word" of its flight, and it becomes ambiguous to know what direction and magnitude of force should be applied at all times. Thus, at some sufficiently high speed we might expect to see something unusual—something unplanned for in the Newtonian world. The trouble is, we do not know a priori what speed this breakdown would occur, and we certainly do not know what new description would be applicable.

In circumstances like this, it is often best to write down effective theories that satisfy the symmetries of your worldview and do precision measurements to find deviations. The pattern of deviations or the values of couplings in the effective theory can lead to new insight when explained by a deeper theory. Bob's $1/r^2$ correction theory and Alice's $1/r^3$ correction theory to the gravity potential in the preceding sections do precisely that. They are Galilean invariant, and satisfy all the symmetries cherished even then: rotational invariance and translation invariance.

We apply this approach of writing down corrections to planetary motion because this is our greatest hope to find cracks in the old classical world view. Since no cherished symmetries are violated by the additional terms we have found before, we may even expect to find breakdowns of Newton's theory by the orbits of the planets, especially since they are accessible and moving faster with respect to each other and the sun than any laboratory system that could have possible been created on the earth at the time. Precision measurements of fast planetary motions thus had good reason to be the first place to find breakdown of Newton's theory. No planet moves faster than Mercury. Indeed, it is Mercury where the first fissures arise, as we shall describe in the following sections.

## 3.5  Effective Theories

It is my contention that the concepts of Effective Theories, if understood and held by the early Newtonian scientists, would have led to a prediction that there must *necessarily* be an anomalous perihelion precession of Mercury and other planets, and that even the order of magnitude could have been guessed well before Le Verrier's announcement in 1859. There was no barrier to adopting these ideas in Newton's day, as it requires no new special experimental knowledge, nor knowledge of Einstein's relativity, but rather a more mature approach to how we think about the

laws of nature. In order to present this viewpoint, I shall first give a précis of the modern notions of Effective Theories.

At its core, the term Effective Theory is short for a body of evidence that has led us to understand that "everything depends on everything else" may be true in principle but certainly not true in practice. In a restricted domain, the theory manifests symmetries and properties that provide the ability to calculate observables without the requirement of making reference to features outside the domain. A simple example of this is that we can compute the trajectory of a football to any practical precision without needing to know the location of Uranus. The effects of Uranus on the trajectory are suppressed by a relative factor of $\frac{m_e r_e^2}{m_U d_U^2} \sim 3 \times 10^{-14}$, where $r_e$ is the radius of the earth, $d_U$ is the distance from Uranus, and $m_e$ ($m_U$) is the mass of the earth (Uranus). This is much too small to take into account for any practical need. The diminishing effect of Uranus as $d_U \to \infty$ is the principle of decoupling, which is at the core of Effective Theory utility and is the central reason why science works and we are able to compute and predict observables.

A central concept of Effective Theory is the recognition that a full theory with heavy and light degrees of freedom can be written at low energies in terms of just light degrees of freedom after "integrating out" the heavy states or "coarse graining" over the small scales. We use "heavy" and "light" abstractly here, as it could refer to masses, momenta, velocity, etc. The chiral lagrangian of QCD, the Fermi theory of electroweak interactions, the Landau-Ginzburg theory of superconductivity (Polchinski 1992) can all be recognized as an Effective Theory of a more fundamental theory.

This top-down approach to understanding Effective Theories can give us a multitude of theoretical insights into the nature of simplified low-energy theories. It is this top-down approach that is traditionally how the power of Effective Theory concepts is demonstrated in particle physics (Cohen 1993; Rothstein 2003), fluid mechanics (Delgado-Bucalioni et al. 2005), material science (Abrams 2005), and essentially any other field that has a separation of scales. However, when considering theories from bottom up, the concepts we learn from Effective Theories can help us deduce modifications and additions to our present theories that can be tested by experiment. Success then can lead to motivations for inducing a more fundamental theory that reproduces the Effective Theory when restricted to its domain. It is this direction in theory analysis that I emphasize here for our present purposes.

The insight that I would like to focus on, which I believe is the most powerful one when it comes to divining additions and modifications to theories, is the role that symmetries and naturalness play in the construction of the "complete" Effective Theory. A symmetry is a recognition that something (a triangle, an equation, etc.) stays the same even if you make a closed set of transformations (i.e., group operations) on that object (rotations by 180°, interchange of $x$ and $y$ variables, etc.). All of our fundamental theories have inherent recognized symmetries in them. We cannot proceed without these recognitions in the Effective Theory, because even the names we give to objects are merely shorthand notation for their symmetry properties (e.g., electrons

are spin-1/2 representations of the Lorentz Group with additional gauge symmetry representation labels).

One of the principle consequences of the Effective Theory approach to establishing natural law is that all possible interactions (or "terms") consistent with the recognized symmetries of the Effective Theory are generically expected. There may or may not be additional terms that violate the symmetries, but terms that do not violate the symmetries must be included. In the realm of Effective Theories within quantum field theory, Weinberg, reflecting on the last three decades of work on the subject, has made the equivalent point that an Effective Theory may be considered self-consistent and not sick "as long as every term allowed by symmetries is included" (Weinberg 2009).

In short, the precise form of a theory or law is not what is to be taken most seriously—it is the recognized symmetries. Upon sorting out the symmetries, the Effective Theory is to be developed with all possible terms consistent with the symmetry, and then qualitative expectations for experiment can be presented. What remains is measurement and pinning down the actual values of the coefficients to each symmetry preserving interaction term.

### 3.5.1  Application to Newton's Gravitation

Newton's law of gravitation is that the force between two bodies of masses $m$ and $M$ is inversely proportional to the square of the distance between them, with the proportionality constant being Newton's constant $G$:

$$F(r) = \frac{GMm}{r^2}, \text{ or } V(r) = \frac{GMm}{r} \tag{3.44}$$

where $V(r)$ is the potential. In Book 3 of Principia, Newton states categorically that the inverse square law is "proved with the greatest exactness from the fact that the aphelia are at rest" and that "the slightest departure from the ratio of the square would necessarily result in a noticeable motion of the apsides...." (Newton 1999). Thus, the theory was created and solidified as a proposition to the world.

Newton's inverse-square law was so sacrosanct that few would ever doubt it. Immanuel Kant in 1747 used the inviability of the inverse-square law to derive that space had three dimensions. This is due to what we would say today is the conservation of gravitational flux lines emanating from a point mass through the surface of a sphere of arbitrary radius. God could have chosen a different gravity law, Kant says, and the number of spatial dimensions then would have had to be different.[4]

---

[4] "Zweitens, dass das Ganze, was daher entspringt, vermöge dieses Gesetzes [inverse-square law] die Eigenschaft der dreifachen Dimension habe; drittens, dass dieses Gesetz willkürlich sei, und da Gott dafür ein anderes, zum Exempel des umgekehrten dreifachen Verhältnisses [i.e., inverse-cube law], hätte wählen können; dass endlich viertens aus einem andern Gesetze auch eine Ausdehnung von andern Eigenschaften und Absmessungen geflossen wäre" (Sect. §10 in Kant 1747).

This rigid adherence to "god-given" specific law is ultimately incorrect reasoning, and it is in conflict with modern views of Effective Theories.

The modern sensibility says that we should focus more on the symmetries, and then refashion the complete Effective Theory using them. What are the symmetries of the Newtonian world? The symmetries are that the laws of physics cannot be affected by one's orientation in space, by one's location in space, nor by one's location in time. The laws must be invariant to any transformation of rotation, spatial translation, or time translation. These symmetry properties go under the name of Galilean invariance. As a side comment, the Lorentz invariance of Einstein's special relativity asymptotes to Galilean invariance in the low velocity limit (i.e., when $v \ll c$).

The interaction term of Eq. 3.44 is merely one term in an infinite number of terms that could be written down that are completely consistent with Galilean invariance. An Effective Theory approach would be to introduce them all and investigate the consequences. There is no meaningful symmetry that demands only the inverse square law interaction. Assured of this, one example would be to embellish Newton's law by

$$V_{ET}(r) = \frac{GMm}{r}\left[1 + \sum_{n=1}^{\infty} \lambda_n \left(\frac{r_0}{r}\right)^n\right] + \cdots \tag{3.45}$$

where $r_0$ is some dimensionful Effective Theory length scale and $\lambda_n$ are dimensionless coefficients, which together with $r_0$ can be found by performing precise experiments. We should note that there are an infinite variety of other terms that could be added, including $r^j$ and $\dot{r}^k$ interactions, but we streamline the argument by looking only at one class of corrections that decouple as $r \to \infty$.

### 3.5.2 Inevitable Perihelion Precession

An extremely important conclusion can already be presented from the rules of Effective Theories. Any deviation from the pure inverse square law will lead to a perihelion precession of the planets, and as the constructed Effective Theory demands additions to the inverse square law there will be an anomalous perihelion precession of the planets. On the other hand, we know that the inverse square law is approximately correct and thus we have added terms that decouple as $r \gg r_0$. The perihelion precession of Mercury is very small, and so we expect that $r_0$ should be much less than the orbital radius of Mercury around the sun. In that case, we are justified in looking at the first-order corrected potential, which we can write as ($\lambda_1 r_0 \to R$):

$$V_1(r) = \frac{GMm}{r}\left(1 + \frac{R}{r}\right). \tag{3.46}$$

By these arguments of Effective Theory, an anomalous perihelion precession of Mercury is inevitable. It is only a question of what value does $R$ take, which then sets

the numerical value of the precession. In the subsequent sections we discuss some arguments for what $R$ might be, from the vantage point of pre-special relativity and pre-general relativity days, and make rough quantitative predictions for the precession rate.

Up to this point we have argued that the focus should have been more on the symmetries of the gravitational theory rather than the concretization of the theory. A more complete Effective Theory for Newtonian gravity would have been accepted and one would have fully expected anomalous perihelion precessions of the planets. A potential similar in form to Eq. 3.46 would have been put forward, and the task of theoretically divining or experimentally measuring $R$ would have been the consuming activity.

## 3.6 Mercury's Anomalous Perihelion Precession

Let us imagine that Bob and Alice are two physicists who are working in the post Le Verrier and pre Einstein era. They are smitten by the Newtonian worldview. They do not wish to do radical things to explain this perihelion precession. They are well-versed in the concepts of Galilean Invariance, Hamilton's Principle, and have an inkling of the ideas of effective theories. Naturally, they want to describe this precession through a Galilean invariant effective theory of gravity. Bob announces that he wishes to add a $1/r^2$ correction to the lagrangian. Not wanted to follow in Bob's footsteps, Alice declares that the force law should be even powers of $1/r^2$ and so her first correction to the lagrangian is $1/r^3$. The two lagrangians are

$$L_{\text{bob}} = \frac{1}{2}m\dot{\mathbf{r}}^2 + \frac{\alpha}{r}\left(1 + \frac{R_{\text{bob}}}{r}\right) \tag{3.47}$$

$$L_{\text{alice}} = \frac{1}{2}m\dot{\mathbf{r}}^2 + \frac{\alpha}{r}\left(1 + \frac{R_{\text{alice}}^2}{r^2}\right) \tag{3.48}$$

where $\alpha = GMm$, with $G$ being Newton's constant, $M$ is the mass of the sun, and $m$ is the mass of the planet under consideration. These are the two lagrangians of Bob and Alice that we studied in a previous lecture. These new laws of Bob and Alice require the introduction of a new fundamental length scale $R_i$. They do not know what that length scale is, but they have hopes that the new data will pin it down for them.

Before we look more closely at Bob and Alice's theories, we should remark again that in the classical history of gravity, there were early attempts to explain anomalies by changing Newton's laws, even in the manner of Alice and Bob. Such theories go under the name of "Clairaut laws". Clairaut proposed in 1745 that Newton's law should be corrected by a $1/r^4$ force term in order to explain some thought-to-be anomalies in the movement of the lunar perigee. However, he found in the end there was not a discrepancy, which buried such laws deeper into the dustbin of history.

**Table 3.2** Anomalous perihelion precession rates of the planets compared to expectations from Newton's law of gravity and taking into account all other sources of precession (effects of other planets orbits, etc.) (Duncombe 1956)

| Planet | $\delta/T$ (arcsec/century) |
| --- | --- |
| Mercury | $43.11 \pm 0.45$ |
| Venus | $8.4 \pm 4.8$ |
| Earth | $5.0 \pm 1.2$ |

More modern references test gravity (including precession rates) through the parameters of the so-called parametrized post-Newtonian (PPN) approach (Will 2005)

Newcomb commented in 1882 that such laws were "out of the question" because they disrupted the gravitational strength so wildly at very close distances where the correction term would come to dominate (Newcomb 1882). As late as 1910 Newcomb, the world's leader on this issue, was stating that all the data up to that point "... seems to preclude the possibility of any deviation from that law [Newton's inverse-square law]" and that Mercury's perihelion advance is best explained by "the hypothesis of Seeliger" (Newcomb 1910), which was a zodiacal light theory that contained intra-Mercurial distributions of orbital matter minimally disruptive to all other astronomical observations except Mercury's perihelion advance (see, e.g., Chap. 4 of (Roseveare 1982)).

Bob and Alice's theory are a return to the Clairaut law in some ways. In the next few subsections we merely state the effect they would have on planetary orbits. After a discussion of Effective Theories and how they apply to this problem, we shall proceed with a somewhat fanciful alternative history of how deviations from Newton's laws could have been explained and interpreted from the point of view of Effective Theories after the anomaly was announced by Le Verrier. But it should be kept in mind, and will be emphasized again in the concluding section, that these theories could have been anticipated, and perhaps even should have been anticipated, before Le Verrier's announcement.

### 3.6.1 Analyzing Bob's $1/r^2$ Correction Theory

From Eq. 3.30 we can compute in Bob's theory that it is necessary that $R_{\text{bob}} = 4.4$ km if Mercury is to have the measured 43 s of arc per century in its perihelion precession. Given this value of $R_{\text{bob}}$, Bob can make predictions for the perihelion advance of other planets. Using Eq. 3.28 he finds $\delta/T_{\text{orbit}} = 8.6''$ of arc per century for Venus's perihelion precession and $3.8''$ for the earth. These predicted values compare favorably to the measurements for Venus and Earth presented in Table 3.2. The predictions are well within the errors, and Bob is pleased because he has found a way to explain the anomaly while yet retaining Galilean invariance as a fundamental symmetry of spacetime. He has done this through the means of a simple expansion correction to Newton's law of gravity. Nothing radical was done.

Despite the successes, Bob is not totally satisfied. He wants to know if he can argue for this new length constant in nature $R_{bob}$. It's a very strange distance 4.4 km. He wonders how he can formulate this distance from all the invariants swirling around him. It should not depend on the mass of each planet, he reasons, because we have just shown that one value of $R_{bob}$ appears to work universally well for all planets. The other options we have to build a length scale are from Newton's constant $G$, the mass of the Sun $M$ and angular momentum. Bob fails to find any natural combination that will give 4.4 km.

Before giving up he recalls that his intuition has told him that there is some characteristic high speed such that Newton's simple laws become strained (see Sect. 3.4). He does not know what that speed value is, and his new law is just as much action at a distance as the old one, but he carries on by giving this new speed a name, $v_{bob}$. With this new undetermined speed in hand he realizes immediately that he can form a new length scale $GM/v_{bob}^2$. Can this be the origin of $R_{bob}$? What value must $v_{bob}$ be to recover $R_{bob} = 4.4$ km? A simple calculation yields

$$v_{bob} = \sqrt{\frac{GM}{R_{bob}}} = 1.7 \times 10^8 \text{ m/s}. \qquad (3.49)$$

This quantity $v_{bob}$ that Bob has derived is a very curious number! His colleagues down the hall have been working on the theory of electromagnetic phenomenon and a speed very close to that keeps showing up in their equations, $c = 3.0 \times 10^8$ m/s. This is the propagation speed of light. He decides this cannot be a coincidence, but he is not sure what to make of it. He decides to define a new scale based on these thoughts, the "sun's electro-gravity scale" $R_{EG} \equiv GM/c^2$. $R_{bob}$ can now be written in terms of this definite scale $R_{bob} = \lambda_{bob} R_{EG}$. It is very curious that the data fits very well if $\lambda_{bob} = 3$ is an integer. He writes on a piece of paper his new theory of gravity

$$L_{bob} = \frac{1}{2} m \dot{\mathbf{r}}^2 + \frac{GMm}{r} \left( 1 + 3 \frac{GM/c^2}{r} \right), \qquad (3.50)$$

and he is pleased with its simplicity, elegance and symmetry. He does not know how the speed of light $c$ crept in, but he is satisfied since his lagrangian looks "natural" given that there are no really big or really small numbers populating it. Furthermore, he knows that if he must construct a new length scale with a speed, the "natural" next known threshold of speed is the speed of light, and so this correction is "natural" to explore. He feels he is on to something big.

Bob finds another interesting connection with this scale $GM/c^2$. He recognizes that there is a small radius $R_E$ of a infinitesimal (i.e., radius less than $R_E$) spherical body of mass $M$ for which an object going the speed of light would not be able to escape. This light-speed trapping radius is a curiosity: if light were corpuscular in any sense, as Newton and others thought it might be, then we could see no light

emanating from within the radius $R_E$ of the massive body. This sets a mystery scale to gravity that requires further scrutiny and may be a length scale associated with changes in gravity. The computation of this scale is simple in the Newtonian world, and is

$$R_E = \frac{2GM}{c^2} \quad \text{(light non} - \text{escape radius)} \tag{3.51}$$

This is only a factor of two different than the value of $R$ he has derived from the perihelion precession rate. It should be noted that $R_E$ is the precisely the Schwarzschild radius derived in General Relativity, which is a well-known special scale for spherically symmetric objects for more reasons than just what was stated above (Schwarzschild 1916; Wald 1984). Furthermore, it should be recalled that the speed of light was being quantitatively estimated (Rømer 1676) even before Newton's *Principia*, and by 1729 it was known to within a few percent (Bradley 1729), and so this scale had precise meaning from the very beginning days of Newtonian gravity.

Despite these interesting connections, Bob gets nervous looking over his equations. Equation 3.27 seems to indicate that if $\rho < 2R_{\text{bob}} = 6GM/c^2$, the orbits do not make sense anymore, as the equations formally say $r < 0$ which is nonsensical. He relaxes briefly when he realizes that $2R_{\text{bob}}$ is only 9 km, which is well below the orbital radius of any planet, and furthermore it is even below the radius of the sun, which is $7 \times 10^5$ km. Thus, there is no danger that some small object rotating around the sun would have no chance to be described by Bob's theory, since it would be inside the sun.

Nevertheless, he is still a bit uncomfortable. Nowhere in his derivation was the radius of the sun ever required. In principle, all that mass of the sun could have been at one infinitesimal point for all the equations knew. Nevermind how to pack all that mass in with a radius less than 9 km, it is a possibility in principle that such a tightly packed object exists, and if it did, there is no way his theory could describe close-by orbits with characteristic orbital latus rectum size $\rho < 9$ km. He knows his theory cannot be the end all of all the theories anyway due to not knowing why $c$ crept into his equations, despite that being the natural next "speed scale" to consider, but now he is even more discomfited because he can imagine configurations where his theory just cannot even give an answer. But that is for another day. He has succeeding in explaining the precessions of Mercury, Venus and Earth and that is enough for a day's work. And that is what Effective Theories do. They explain the day's work—Bob clearly has made progress—but there is more to be learned and understood. Effective Theory practitioners understand that all possible questions cannot be resolved instantly, and that there are necessarily deeper effective theories to come.

### 3.6.2 Analyzing Alice's $1/r^3$ Correction Theory

Alice now wishes to make definite her lagrangian with $1/r^3$ potential corrections by specifying the value of $R_{\text{alice}}$ from Mercury's anomalous perihelion precession and then predicting what the other precession rates are. Upon fitting Mercury data she finds $R_{\text{alice}} = 9.04 \times 10^5$ m. Using that fixed value for all planets she then predicts $\delta/T_{\text{orbit}} = 4.43''$ of arc per century for Venus and $1.4''$ of arc per century for the Earth. The Venus result is nearly $2\sigma$ off compared to the measurement, and the Earth result is about $3\sigma$ off of the measurement (see Table 3.2). Alice has a choice now. She can say her theory predicts that further refined measurements of the precession rates will yield smaller central values of the precession rates for Venus and Earth in concert with her theory. Or, she can take the $3\sigma$ discrepancy seriously and attempt to modify her theory.

Alice makes the right choice and seeks to modify the theory. She computes what $R_{\text{alice}}$ needs to be for each planetary case to precisely hit the measured values. She finds

$$R_{\text{alice}}^{\text{mercury}} = 90 \times 10^7 \, \text{m}, \quad R_{\text{alice}}^{\text{venus}} = 1.3 \times 10^7 \, \text{m}, \quad R_{\text{alice}}^{\text{earth}} = 1.5 \times 10^7 \, \text{m}. \quad (3.52)$$

Similar to Bob, she begins to think about how these length scales can be identified with all the quantities that she has available to her in the problem: $M$, $m_{\text{planet}}$, and $\ell$. She cannot come to a satisfactory answer. These constants alone are not enough to form the length scales of Eq. 3.52.

However, in Alice's trials she notices something interesting. The $R_{\text{alice}}$ lengths are proportional to angular momentum divided by mass of the planet, $R_{\text{alice}}^i \propto \ell_i/m_i$, with the same proportionality constant. This constant has the dimensions of an inverse velocity. She decides to call it $v_{\text{alice}}$ and solves for its value:

$$R_{\text{alice}}^i = v_{\text{alice}}^{-1} \frac{\ell_i}{m_i} \implies v_{\text{alice}} = \frac{\ell_i}{m_i R_{\text{alice}}^i} = 3.0 \times 10^8 \, \text{m/s} \qquad (3.53)$$

Alice also has colleagues that work on electromagnetism and she recognizes this value as exactly the speed of light, $v_{\text{alice}} = c$. How did that happen? She does not know, but she is surely excited about the result, as she too recognises that $c$ is the next fundamental "speed threshold" and so is a "natural" value in the Effective Theory correction. She has explained all the planetary precession data. She writes down on a piece of paper her new theory of gravity,

$$L_{\text{alice}} = \frac{1}{2} m \dot{\mathbf{r}}^2 + \frac{GMm}{r} \left( 1 + \frac{1}{c^2} \frac{\ell^2/m^2}{r^2} \right). \qquad (3.54)$$

which like Bob's theory possesses symmetry and has a measure of elegance and simplicity.

As she reflects on her theory she realizes that since angular momentum is $\ell \sim mrv$, where $v$ is the velocity of the planet orbiting the sun, the second term inside the parenthesis can be thought of as an $m$-independent $v^2/c^2$ correction to the Newtonian gravitational potential. Thus, she believes that she will be the first to show that the simple inverse-square law of Newton is corrected by factors of $v^2/c^2$. As the speed of the planet gets closer to the speed of light, Newton's theory begins to crack. So far the basic assumptions of spacetime symmetries—Galilean Invariance—are not breaking down, just the simple form of Newton's theory of gravity. Despite these successes of her theory, she remains slightly dissatisfied with one aspect. How can she convince herself, much less others, that her theory is better than Bob's? Surely one or the other or some combination of these corrections are required by nature, she reasons, but can they be determined from deeper theory principles? The answer is yes, and Einstein's General Relativity is that theory.

### 3.6.3 Gerber's "Utterly Worthless" Theory

Before going to Einstein's General Relativity, let us comment briefly on velocity dependent approaches to augmenting Newton's law. Manipulations of the Newtonian potential were initiated in earnest well after Laplace's work with the goal of rigorously incorporating finite speed effects of gravity. The most straightforward approaches failed. However, Paul Gerber proposed in 1898 (Gerber 1898) a velocity dependent potential correction that correctly accounted for Mercury's perihelion precession:

$$V(r, v) = -\frac{M}{r} \left(1 - \frac{v}{c}\right)^{-2} \tag{3.55}$$

where $c = 3 \times 10^8$ m/s is the speed of light, and $v$ is the velocity of Mercury in the Sun-Mercury center of mass system.

Gerber's theory captured the attention of many due to its combined simplicity and effectiveness in accommodating Mercury's anomalous perihelion precession rate. For example, Mach wrote, "Only Paul Gerber [reference to 1898 paper] studying the motion of Mercury's perihelion ... did find that the speed of propagation of gravitation is the same as the speed of light" (Mach 1901). He was attacked for not giving good reasons for his theory—a topic we shall take up below—but he did provide a simple theory that worked. It was also a "natural" theory due to its utilization of $c$ as the next fundamental speed scale of the theory.

Seventeen years after Gerber's potential, the question of Mercury's perihelion precession was resolved powerfully by Einstein's GR (Wald 1984). At low velocities the first-order correction to gravitational attraction of Gerber's theory matches the first-order correction of Einstein's theory. However, Einstein's approach had coherent principles and unassailable logic, and thoughts about Gerber's theory quickly faded away.

Despite the success in accommodating Mercury's perihelion precession, Gerber was roundly criticized for his theory. The strength of the reaction that Gerber faced seems harsh for somebody who actually did write down a simple theory of no free parameters with the speed of light in it that worked. It is as though the deep thinkers at the time knew there was something appealing about Gerber's work, but could not quite put their finger on it, and so harshly criticized it as a community building exercise to dismiss that kind of apparently principle-less approach to physics.

Einstein, commenting on Gerber's theory well after he had developed his own theory of General Relativity summarized the attitudes well: "But specialists in the field agree not only that Gerber's derivation is thoroughly incorrect, but that the formula cannot even be obtained as a consequence of Gerber's leading assumptions. Mr. Gerber's paper is therefore *utterly worthless*" (Capria 1999) (italics are mine). This appears to be an overly strong dismissal of Gerber's simple theory that gained so much attention.

Pauli, in his famous Encyclopedia article on Relativity said,

> Recently, an earlier attempt by P. Gerber has been discussed which tries to explain the perihelion advance of Mercury with the help of the finite velocity of propagation of gravitation, but which must be considered *completely unsuccessfully* from a theoretical point of view. For while it leads admittedly to the correct formula—though on the basis of false deductions—it must be stressed that, even so, only the numerical factor was new. (Paul 1981) (italics mine)

Whatever can be said of Gerber and his theory and the faulty logic behind his theory, it was not "utterly worthless" or "completely unsuccessful". I believe it was a crude attempt at effective theory analysis. It was something he may have intuited but was unsuccessful in articulating well due to the mindset and style of physics of the day. Back then, no term was allowed to augment a theory without it being derived first from a deeper principle. The standard rigor of the day was that laws were exact by argument and deduction, and any deviations or changes must be accounted for by a replacing new principle.

An excellent example of this prevailing attitude is provided by Max Born in his book on Einstein's theory of relativity (Born 1924). He describes briefly the case of Mercury's anomalous perihelion precession and then goes on to harangue all those people before Einstein who generated ad hoc solutions to the problem:

> Changes in the laws [Newton's laws] have been proposed, but they have been invented quite arbitrarily and can be tested by no other facts, and their correctness is not proved by accounting for the motion of Mercury's perihelion. If Newton's theory really requires a refinement we must demand that it emanate, without the introduction of arbitrary constants, from a principle that is superior to the existing doctrine in generality and intrinsic probability. Einstein was the first to succeed in doing this.

This attitude is partially in conflict with our understanding of Effective Theories today. The introduction of arbitrary constants is a key step in the construction of Effective Theories, and the role of experiment is to pin those down. If anything, the ad hoc inventors of changes in Newton's law were too sheepish about introducing arbitrary parameters, and instead got tangled up with incoherent "deep reasons" for their particular laws.

Effective Theory is an intermediate step between an old regime (e.g., Newton's laws) and a new regime (e.g., Einstein's General Relativity), and this intermediate step necessarily has "arbitrary couplings" and does not "emanate from a principle that is superior to the existing doctrine". Instead, it says that the existing doctrine should be taken to its utmost seriousness (e.g., Galilean invariance) and data should fit the parameters of all allowed interactions, and perhaps a deeper new theory can come along later to explain the relations among those parameters.

Although Gerber's theory was not worthless, it is not as valuable as Einstein's General Relativity. Alice and Bob's effective theories would not have been worthless either had they written it down much earlier. They would have been an intermediate stepping stone from one principled theory to the next that would have predicted the existence of Mercury's perihelion precession and motivated earlier discovery of the phenomena.

## 3.7 Perturbation from General Relativity

We have talked about Einstein's General Relativity being the deeper theory that explains Mercury's perihelion precession. It is worthwhile in these lectures to go through that computation to see how it comes about.

We wish to compute the trajectory of a particle subject to a central, radially symmetric gravitating source in the general approach followed, for example, by (Hartl 2003). The metric applicable for this computation is the Schwarzschild metric:

$$ds^2 = -\eta(r)c^2dt^2 + \frac{dr^2}{\eta(r)} + r^2d\theta^2 + r^2\sin^2\theta d\phi^2 \tag{3.56}$$

where

$$\eta(r) = 1 - \frac{2GM}{c^2r} = 1 - \frac{r_s}{r}, \quad \text{where } r_s \equiv 2GM/c^2 \tag{3.57}$$

The quantity $r_s$ is the Schwarzschild radius. This defines the metric tensor to be

$$g_{\alpha\beta} = \text{diag}(-\eta(r), \eta(r)^{-1}, r^2, r^2\sin^2\theta) \tag{3.58}$$

in the $(t, r, \theta, \phi)$ basis. Note that the signature of the metric (asymptotically weak field far away) in normal rectilinear coordinates is $g^{\alpha\beta} = \text{diag}(-1, 1, 1, 1)$.

The Schwarzschild metric is unperturbed by making shifts in the time direction and by making shifts in the angular direction $\phi$. These define Killing vectors $\xi_{\text{time}}^\lambda = (1, 0, 0, 0)$ and $\xi_{\text{rot}}^\lambda = (0, 0, 0, 1)$. The nice property of a Killing vector is that when dotted into the four-velocity vector $dx^\alpha/d\tau$ the result must be constant along the geodesic motion:

$$\xi^\lambda \frac{dx_\lambda}{d\tau} = g_{\alpha\beta}\xi^\alpha \frac{dx^\beta}{d\tau} = \text{const.} \tag{3.59}$$

Applying this theorem to the Schwarzschild metric gives

$$g_{\alpha\beta}\xi^\alpha_{\text{time}} \frac{dx^\beta}{d\tau} = \eta(r)\frac{dt}{d\tau} = c_1 \tag{3.60}$$

$$g_{\alpha\beta}\xi^\alpha_{\text{rot}} \frac{dx^\beta}{d\tau} = r^2 \sin^2\theta \frac{d\phi}{d\tau} = c_2 \tag{3.61}$$

where $c_1$ and $c_2$ are mere constants. We know that independence of time implies conservation of energy, and we also know that independence of rotation implies conservation of angular momentum. Thus, we know that $c_1$ is some function of energy, and we know that $c_2$ is some function of angular momentum as we usually define the quantities. However, at this stage we do not know the precise correspondence, so it is prudent to just carry the constants $c_1$ and $c_2$ with us until the precise relations become obvious.

From Eq. 3.61 we solve for $dt/d\tau = c_1/\eta(r)$ and $d\phi/d\tau = c_2/r^2 \sin^2\theta$. Now, we should simplify this all by taking the orbit in the $\theta = \pi/2$ plane and so $d\phi/d\tau = c_2/r^2$. Please note, conservation laws have given us this, and this is where deep physics lies. Now, let's expand out the defining equation of the four-velocity

$$g_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = -1, \quad \text{which gives} \tag{3.62}$$

$$-\eta(r)\left(\frac{dt}{d\tau}\right)^2 + \frac{1}{\eta(r)}\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\phi}{d\tau}\right)^2 = -1 \tag{3.63}$$

for the Schwarzschild metric. Substituting the values of $d\phi/d\tau$ and $dt/d\tau$ that we obtained above from the Killing equations, we find

$$-\frac{c_1^2}{\eta(r)} + \frac{1}{\eta(r)}\left(\frac{dr}{d\tau}\right)^2 + \frac{c_2^2}{r^2} = -1 \tag{3.64}$$

After carrying out some algebra one finds

$$\frac{mc^2}{2}(c_1^2 - 1) = \frac{1}{2}mc^2\left(\frac{dr}{d\tau}\right)^2 - \frac{GMm}{r} + \frac{mc^2 c_2^2}{2r^2} - \frac{GMmc_2^2}{r^3} \tag{3.65}$$

The form of Eq. 3.65 is very suggestive of our equation for energy of a particle in an orbit, and the correspondence becomes precise if we make the identifications

$$\frac{mc^2}{2}(c_1^2 - 1) \equiv E \quad \text{and} \quad c_2^2 \equiv \frac{\ell^2}{m^2 c^2}. \tag{3.66}$$

We also can identify $\tau = ct$ in the non-relativistic limit. It turns out that this substitution is acceptable for the problem at hand as long as $\dot{r} \ll r\dot{\phi}$, which is generally the situation for low eccentricity orbits, and certainly the case for the planetary orbits of our solar system. Making these identifications the energy equation becomes

$$E = \frac{1}{2}m\left(\frac{dr}{dt}\right)^2 + \frac{\ell^2}{2mr^2} - \frac{GMm}{r}\left(1 + \frac{\ell^2/m^2c^2}{r^2}\right). \qquad (3.67)$$

This is the energy equation for a particle in Newtonian gravity except for the small shift in the effective potential

$$\Delta V_{eff}(r) = -\frac{GMm}{r}\left(\frac{\ell^2/m^2c^2}{r^2}\right) \qquad (3.68)$$

which is precisely the same correction to Newton's theory we derived earlier from Alice's effective theory approach to explain Mercury's precesion in Eq. 3.54.

There are multiple ways to derive the correction to Newton's gravity law for the particular problem of perihelion precessions. In our derivation, we found Alice's theory correction. This is also the result derived in General Relativity by many other authors (see e.g., Schutz 1990; Goldstein et al. 2002; Hartl 2003). However, another approach to the General Relativity derivation gives Bob's theory, and that has been demonstrated by a set of different authors (see e.g., Paul 1981; Landau and Lifshitz 1975; Iwasaki 1971; Donoghue 2009). These two theories, if treated as god-given complete theories, are not equivalent. However, they are equivalent results for this problem as all approximations and culling of the General Relativity terms have been carried out with the sole purpose of finding the perihelion precession. In the end, the precession rate angle per orbit period from either correction is the same:

$$\delta = \frac{6\pi GM/c^2}{a(1 - e^2)} \qquad (3.69)$$

Algebraically, the orbital identity

$$\ell^2 = GMm^2a(1 - e^2). \qquad (3.70)$$

is what guarantees that the two solutions predict the same anomalous perihelion precession rate. So, we see that Albert explains both Alice's theory and Bob's theory, and puts them on firmer footing.

## 3.8 Conclusions

At the beginning of these lectures we decided that Newton's law of Gravitation was very successful in describing the orbits, but that it is not the precise law that captures our most profound admiration. Rather, it is the symmetries that the theory possesses. We elevated those to the highest principles and constructed reasonable effective theories that could be expected by the data. We illustrated the results with two theories: Bob's $1/r^2$ and Alice's $1/r^3$ potential correction theories. Both theories were able to account for the perihelion rate naturally. We even made the case that philosophical challenges to Newton's world view, if taken seriously, could presage the size of Mercury's correction that was actually measured by Le Verrier. This is done with the aid of "naturalness" arguments about the speed of light being the next speed scale of nature by which to construct corrections to Newton's potential. In this way the concepts of natural effective theory have some predictive power. That power is certainly qualitative, but also to some degree quantitative.

Einstein had keen insights into the nature of space and time and developed the theory of General Relativity based on them. It describes gravity at a deeper level, and one of its first orders of business was to compute the anomalous precession rate of Mercury to see if it could account for the discrepancy between Newton's theory and measurement. The answer is yes, and we have shown that this correction matches nicely the effective theory results of Alice and Bob.

Einstein's General Relativity theory is "better" than Alice's theory or Bob's theory for two reasons. First, it gives a deeper principles understanding of the correction with no additional free parameters. This deeper understanding is nothing other than further assumptions on spacetime symmetries that panned out. Second, Einstein's theory is a more complete theory of gravity that makes additional predictions (such as bending of light, and binary pulsar spin-down) that are confirmed by data. Alice or Bob's theory clearly cannot match the riches of General Relativity and so cannot be considered as fundamental as Einstein's.

Despite Bob and Alice's theory coming up short, the general lesson remains. Newton's theory was an effective theory, which is in some aspects superceded in success by Bob and Alice's effective theory, and Bob and Alice's effective theories are superceded in success by Einstein's General Relativity. The obvious next question is whether Einstein's General Relativity theory can be succeeded in success by another theory. A deeper theory that perhaps could be explained as effective theory expansion of Einstein's theory for the purposes of solving some lower energy precision measurement problem. There is little doubt that is the case (Donoghue 1994).

Finally, one of the most profound shifts in our thinking over the decades, illustrated well by the Perihelion precession example, is that it is really no longer appropriate to speak of "the correct theory." There is no correct theory. Our tasks are to improve theories via the effective theory approach, to seek deeper and simplifying assumptions that account for it, solidify those into a new theory, and then treat that new theory as an effective theory, and repeat. These steps are accomplished by

continually improving and refining observations and theory computations that enable us to choose between effective theories, followed by deducing deeper new symmetries that force its inevitability. Theories are never to be trusted—they are always "wrong" in the end—and with concerted effort we can even anticipate when and how they will break down.

The concepts of Effective Theory lead one to predict qualitatively that a perihelion precession of Mercury was *a priori* guaranteed even knowing only the experimental facts of the Newtonian era. In particular, elevating symmetries above the concretization of hypothesized law, in this case the rigid devotion to the inverse square law, is the basic ingredient that would have led unambiguously to this conclusion. The general approach to science during the Newtonian era required almost complete devotion to concrete laws and their propositional justifications, which impeded its progress toward developing theory enhancements guided by symmetries and naturalness. Gerber, a school teacher who was perhaps not as indoctrinated in this more rigid fashion, found a potential that worked yet then made unjustified arguments for why it should be true. Effective Theories give the best of both words: deep but modest justifications for theories that can anticipate data and fit the data.

We have also shown that even during the time of Newton a reasonably well supported hypothesis for the perihelion precession of Mercury could have been put forth that is close to the actual experimental result of $43''$ of arc per century. This is a clear illustration of how the ideas of Effective Theory can be utilized to extrapolate modestly beyond the rigidly set forth laws of fundamental physics.

# References

Abrams, C.F.: Effective theories for materials and macromolecules, Institute for Mathematics and Its Applications, University of Minnesota (8–11 June 2005). http://www.ima.umn.edu/matter/W6.8-11.05/

Born, M.: Einstein's Theory of Relativity. Dover Publications, New York, 1962, 1965, original German edition 1920, and original English edition (1924)

Bradley, J.: Account of a new discovered motion of the fixed Stars. Phil. Trans. **35**, 637 (1729)

Capria, M.M.: Common sense, the history, and the theory of relativity. Acta Scientiarum **21**, 779 (1999)

Cohen, A.G.: Selected topics in effective field theories for particle physics. In: Proceedings of the Theoretical Advanced Study Institute (TASI 1993), Boulder, Colorado (1993)

Delgado-Bucalioni, R. et al.: Hybrid molecular-continuum fluid models. Phil. Trans. R. Soc. A **363**, 1975 (2005)

Donoghue, J.F.: General relativity as an effective field theory: the leading quantum corrections. Phys. Rev. D **50**, 3874 (1994). [arXiv:gr-qc/9405057]

Donoghue, J.F.: PoS **EFT09**, 001 (2009). [arXiv:0909.0021 [hep-ph]]

Duncombe, R.L.: Relativity effects for the three inner planets. Astron. J. **61**, 174 (1956)

Gerber, P.: Die räumliche und zeitliche Ausbreitung der Gravitation. Zeitschrift für Mathematik und Physik **43**, 93 (1898)

Goldstein, H., Poole, C., Safko, J.: Classical Mechanics, 3rd edn. Addison-Wesley, San Francisco (2002)

Hartl, J.: Gravity: An Introduction to Einstein's General Relativity. Addison-Wesley, San Francisco (2003)

Iwasaki, Y.: Quantum theory of gravitation vs. classical theory. Prog. Theor. Phys. **46**, 1587 (1971)

Kant, I.: *Gedanken von der wahren Schätzung der lebendigen Kräfte*, M.E. Dorn, Königsberg (1747)

Landau, L.D. Lifshitz, E.M.: The Classical Theory of Fields, 4th revised English edn. Pergamon Press, Oxford (1975)

Laplace, P.S.: Traité de Mécanique Céleste. Tome III, Paris, 1797–1825 (1805)

Le Verrier, U.: Theorie du mouvement de Mercure. Ann. Observ. Imp. Paris (Mém) **5**, 1 (1859)

Mach, E.: Die Mechanik in ihrer historisch-kritisch Entwicklung dargestellt, Leipzig (1901) (As quoted in Capria 1999)

Newcomb, S.: Entry Under "Mercury" in Encyclopedia Britannica, 11th edn. Cambridge University Press, Cambridge (1910)

Newcomb, S.: Discussion and results of observations on transits of Mercury from 1677 to 1881. Astro. Pap. Am. Ephem. Naut. Alm. **1**, 367 (1882)

Newton, I.: The Principia: Mathematical Principles of Natural Philosophy, translated by Cohen I.B., Whitman, A. 3rd edn. University of California Press, Berkeley (1999) [Newton's original three editions were published in 1687, 1713, and 1726. This book is translated from Newton's 3rd edition from 1726]

Oppenheim, S.: Kritik des newtonschen Gravitationsgesetzes. In: Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen, 6.2.2: 80 (1920)

Paul, W.: Theory of Relativity. Dover Publications, New York, 1958, 1981, translated from Relativitätstheorie, Encyclopädie der mathematischen Wissenschaft, vol. V19. B.G. Teubner, Leipzig (1921)

Polchinski, J.: Effective Field Theory and the Fermi Surface (1992). arXiv:hep-th/9210046

Poor, C.L.: The motions of the planets and the relativity theory. Science **54**, 30 (1921)

Rømer, O.: Touchant le mouvement de la lumiere trouvé par M. Rømer de l'Académie Royale des Sciences. Journals des Sçavans 233–236 (1676)

Roseveare, N.T.: Mercury's Perihelion from Le Verrier to Einstein. Clarendon Press, Oxford (1982)

Rothstein, I.Z.: TASI lectures on effective field theories. In: Proceedings of the Theoretical Advanced Study Institute (TASI 2003), Boulder, Colorado (2003)

Schutz, B.F.: A First Course in General Relativity. Cambridge University Press, Cambridge (1990)

Schwarzschild, K.: Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. Klasse für Mathematik, Physik, und Technik, Berlin, p. 189 (1916)

Thayer, H.S.: Newton's Philosophy of Nature: Selections from His Writings. Haftner Press, New York (1953)

Wald, R.M.: General Relativity. University of Chicago Press, Chicago (1984)

Weinberg, S.: Effective Field Theory, Past and Future (2009). arXiv:0908.1964 [hep-th]

Will, C.M.: The Confrontation between general relativity and experiment. Living Rev. Rel. **9**, 3 (2005). [gr-qc/0510072]