Draft of May 11, 2010
PLEASE DO NOT CITE

On (Some) Explanations in Physics[1]

James Owen Weatherall[2]
Logic and Philosophy of Science
University of California, Irvine

"What do we mean here by 'explanation'? ... This whole issue, which perhaps lies between nature and sociology, seems to be a bit vague. Quite possibly, an attempt to make the word *explanation* more precise may do more harm to the field [of physics] than good."
-Robert Geroch (1978, pg. 63)

1 Introduction

Consider the following questions, any of which might be heard in the halls of a physics department.

1. Our best theory of particle physics predicts that in very high energy experiments, which probe the smallest distance scales, the electromagnetic, weak, and strong forces should have approximately the same strength. But at these same distance scales, gravitation is many orders of magnitude weaker. Why is gravity so much weaker than any of the other forces?

2. The Standard Model of particle physics makes predictions that are valid to 15 significant digits (Odom et al., 2006). But the Standard Model's predictions rely on 19 parameters that are "put in by hand" to agree with experiment; in order for the Standard Model to make accurate predictions at all, these parameters must be finely tuned. Why do these Standard Model parameters take the values they do, and is there a sense in which they are "natural" or determined by some underlying mathematical or physical principle?

3. In Newtonian physics,[3] inertial mass (the value m that appears in $\mathbf{F} = m\mathbf{a}$) is usually taken to have the same value as gravitational mass (the coupling to the gravitational field, i.e. the value m that appears in $U_{\mathcal{G}} = m\varphi_{\mathcal{G}}$, where $\varphi_{\mathcal{G}}$ is the gravitational potential and $U_{\mathcal{G}}$ is the potential energy of a particle with mass m), even though in principle the theory distinguishes these masses. This equivalence is empirical: it was

[1]Thanks to Kyle Stanford and David Malament for helpful remarks on earlier drafts of this paper.
[2]weatherj@uci.edu
[3]I will use the expression "Newtonian theory" interchangeably with "Newtonian physics." In both cases I mean Newtonian dynamics plus gravitation.

first established (in slightly different terms) by Galileo; at the end of the 19th century, it was tested with very high precision by Loránd Eötvös. Yet the correspondence seems highly suggestive. Why are inertial and gravitational mass equal in Newtonian physics?

I need not multiply examples. Each of these is a why question asked in a particular scientific context (physics). As such, I take it that they are calls for scientific explanation.[4] Indeed, they are why questions of a particularly important sort: these are the kind of questions that physicists often use to motivate their research projects. Questions 1 and 2 are open and form the basis of several major contemporary research programs[5] in high energy particle physics and quantum gravity. Question 3, meanwhile, has been settled, or at least, we now have the theoretical machinery available to provide one sort of answer to it. I claim that the kind of explanation one can now give in response to question 3 is one kind of explanation that would satisfy the physicists who ask questions 1 and 2. It may not be the only kind of explanation that physicists would ultimately deem satisfactory, though I think it is an ideal of the sort of explanation physicists have in mind when they ask questions 1 and 2. My central goal in this paper will be to examine just what kind of explanation it is.

Before proceeding, however, I should give some context to the present discussion. Over the last 20 years, since Salmon (1989) proposed a detente between the causal and unificationist accounts of explanation, the idea that some sort of pluralist account of explanation is necessary to capture the full variety of explanatory phenomena has gained considerable support. Salmon's own line was that the causal and unificationist accounts were not inconsistent. Instead, he thought they offered different kinds of understanding, corresponding to the different kinds of explanation; in some sense, he thought, *both* the causal and unificationist accounts are correct. On Salmon's view, the two accounts together offer a full account of explanation on which unificationist explanations are "top-down" and causal explanations are "bottom-up." Any given event or phenomenon can be explained in both ways. But this form

[4]Perhaps not all calls for explanation take the form of why questions, and perhaps not all why questions call for explanations. But I claim these why questions do call for explanations.

[5]I do not mean "research program" in a technical philosophical sense. I just mean that string theorists, loop quantum gravity theorists, and many particle phenomenologists (in the physicist's sense of phenomenology) take these questions to be central to their research.

of pluralism is still too limiting: it now seems that some explanations do not fit neatly into either account (see, for instance, Batterman (2002)), and moreover, that some phenomena that are easily explained using one kind of explanation do not, as Salmon suggests, have explanations of the other sort.[6]

More recently, Godfrey-Smith (2003) (following Kuhn (1977)) has suggested a different kind of pluralism, in which what counts as a good explanation can vary depending on scientific context. Explanation in biology need not be the same as explanation in physics, and explanation in either field in the early 21st century need not be the same as explanation was in, say, the 17th century. On this view, it is a mistake (as Geroch suggests above) to attempt to characterize scientific explanation in advance: what counts as a good explanation in science is evolving along with the sciences themselves. I am very sympathetic to this view. But I take it that two projects remain, even after a pluralistic, contexualist account has been accepted. The first project is to identify the working parts of contextualist pluralism: what makes a particular explanation an appropriate one in a given instance, as an answer to a particular why question? Certainly, scientific context—i.e. field of study and historical moment—determines the array of explanations that are available. But in many scientific contexts, such as contemporary physics, it seems that more than one form of explanation is common. Moreover, as Salmon suggests, many phenomena may have explanations of radically different types.[7] The second project, meanwhile, is to identify interesting explanations used in various contexts and attempt to understand their epistemic virtues (or, perhaps, lack thereof). The current paper is an example of the second kind of project.

From here I will proceed as follows. I will start by clarifying what question 3 is asking. I will then sketch what I take the answer to be.[8] This question has been answered informally

[6]I take explanations such as Fisher (1930)'s explanation of sex ratio in humans as examples of this latter sort. (See also Skyrms (1996, Ch. 1), where Fisher's work is put in perspective.) A causal explanation can explain why any individual turned out to have a particular gender, but it cannot explain why the ratio must be what it is.

[7]In my (2005), I suggest that the problem of identifying appropriate explanations is essentially pragmatic, and offer one way of solving it, at least in contemporary particle physics.

[8]The discussion in the body of the paper is precise, but informal. The technical details of the explanation are included in an appendix.

in a variety of ways since General Relativity (GR) first appeared in 1915; the answer I will present here is certainly in the spirit of these standard responses, though it precisifies a number of details about the relationship between mass in GR and Newtonian physics that are usually left vague. To my knowledge, the form of the answer I will present here is original and I take it to be of (minor) independent interest. After presenting the explanation I have in mind, I will turn to the question of whether the explanation I offer here can be understood within the rubrics of well established accounts of explanation. I will conclude that it cannot. In the remainder of the paper, I will try to articulate how the present explanation works, highlighting its distinctive features.

2 Why are inertial and gravitational mass equal in Newtonian gravitation?

As I have said, inertial mass and gravitational mass are conceptually distinct in Newtonian physics (I will distinguish them here by writing $m_{\mathcal{I}}$ for inertial mass and $m_{\mathcal{G}}$ for gravitational mass). Indeed, one would expect them to be unrelated to each other. Inertial mass is a constant of proportionality in the fundamental dynamical principles of the theory. It appears in Newton's second law, which states that $\mathbf{F} = m_{\mathcal{I}}\mathbf{a}$; momentum is defined as $\mathbf{p} = m_{\mathcal{I}}\mathbf{v}$; kinetic energy is $T = 1/2 m_{\mathcal{I}} v^2$. One can think of inertial mass as a measure of a body's tendency to accelerate under the influence of an impressed force. Inertial mass is closely related to inertial *motion*, which enters Newtonian theory via Newton's first law. The first law states that a body undergoing uniform rectilinear motion will not deviate from that motion unless acted on by an external force; inertial mass, then, determines a body's tendency to deviate from uniform rectilinear motion when acted on by an external force, whether gravitational or otherwise.

Gravitational mass, meanwhile, determines the strength of the gravitational force that a body exerts on other bodies and, conversely, is exerted on the body by other bodies. It enters the theory via Newton's law of universal gravitation, which states that given two bodies with respective gravitational masses $\overset{1}{m}_{\mathcal{G}}$ and $\overset{2}{m}_{\mathcal{G}}$, each will exert a force of

$$F_{\mathcal{G}} = \frac{G\overset{1}{m}_{\mathcal{G}}\overset{2}{m}_{\mathcal{G}}}{r^2}$$

on the other, where r is the magnitude of the distance between the bodies' centers of mass and G is Newton's constant. Equivalently, one can think of gravitational mass as a measure of a body's response to a background gravitational field. If a test body[9] with gravitational mass $m_{\mathcal{G}}$ is placed in a gravitational potential $\varphi_{\mathcal{G}}(\mathbf{r})$, then the body will have gravitational potential energy $U_{\mathcal{G}}(\mathbf{r}) = m_{\mathcal{G}}\varphi_{\mathcal{G}}(\mathbf{r})$ and will experience a force $F_{\mathcal{G}} = -m_{\mathcal{G}}\nabla\varphi_{\mathcal{G}}(\mathbf{r})$. Gravitational mass can be thought of (in modern terms) as gravitational *charge*, in analogy with classical electric charge. Indeed, the fundamental force equations have exactly the same structure. Given two bodies with electric charges $\overset{1}{q}$ and $\overset{2}{q}$, Coulomb's law gives that each will exert a force of

$$F_{\mathcal{E}} = \frac{C\overset{1}{q}\overset{2}{q}}{r^2}$$

on the other, where C is Coulomb's constant. Likewise, a test charge q in an electric potential $\varphi_{\mathcal{E}}(\mathbf{r})$ will have electrical potential energy $U_{\mathcal{E}} = q\varphi_{\mathcal{E}}(\mathbf{r})$ and will experience a force of $F_{\mathcal{E}} = -q\nabla\varphi_{\mathcal{E}}(\mathbf{r})$.

The parallel with electric force is particularly salient here. Suppose one wants to know the acceleration exhibited by a test particle of charge q and inertial mass $m_{\mathcal{I}}$ in an electric potential $\varphi_{\mathcal{E}}$. Combining Newton's second law with the force law for a test particle in an electromagnetic field, one finds that

$$\mathbf{a} = -\frac{q}{m_{\mathcal{I}}}\nabla\varphi_{\mathcal{E}}.$$

In other words, the acceleration depends on the ratio of the charge to the inertial mass of the body, both of which are freely varying, independent objects. One can find in nature bodies

[9] By test body, I mean a body that is assumed not to contribute to the gravitational field itself. In other words, when considering test bodies one neglects the "backreaction" of a body's own gravitational field.

with many different values for the ratio $q/m_{\mathcal{I}}$. Meanwhile, if one performs the identical calculation to determine the acceleration due to gravity (given a fixed gravitational potential), one likewise finds,

$$\mathbf{a} = -\frac{m_{\mathcal{G}}}{m_{\mathcal{I}}}\nabla\varphi_{\mathcal{G}}. \tag{2.1}$$

Again, the acceleration depends on the ratio of two values: the gravitational and inertial masses. Given the structural similarities between the gravitational and electric cases, one should expect to go out into the world and find bodies with a wide array of different values for the ratio $m_{\mathcal{G}}/m_{\mathcal{I}}$. After all, (a) how much a body will tend to deviate from rectilinear motion given an external force and (b) the strength of that external force should be independent quantities (as they are when considering electric force). But it turns out that when we start looking into how bodies behave in a gravitational potential, we find something quite different. Given any body at all, the ratio $m_{\mathcal{G}}/m_{\mathcal{I}}$ always takes the same value: choosing the natural units, we always find that $m_{\mathcal{G}}/m_{\mathcal{I}} = 1$.

Given this background, the *explanandum* can be stated as follows: all evidence suggests that, given any body, the gravitational and inertial masses of that body can be demonstrated to be equal, despite the fact that Newtonian theory gives no reason to expect these two masses to be related. In other words, we want to understand *why* gravitational mass and inertial mass appear to be identical in Newtonian theory. In some ways this is an unusual why question (at least with respect to standard accounts of explanation), so I want to spend some time up front focusing on its distinctive features. First off, it is a question about a general observational feature of the world, but it is expressed in the terms of a specific physical theory. In other words, the question takes the Newtonian concepts of gravitational and inertial mass for granted. One can express the observational fact without reference to the Newtonian theory—Galileo first described the phenomenon that all bodies fall at the same rate, irrespective of mass, before Newton was born—but when one does so, the question does not arise. One might, perhaps, wonder about the apparent universality of free fall in other contexts or even quite generally: after all, Galileo's results certainly conflicted with the Aristotelian expectation, so one might well have demanded an explanation for Galileo's observations in the context of *Aristotelian* physics, too. But without the background conceptual machinery of Newtonian physics, the question is different. I am interested in a specific question *about* the world, expressed *within* the Newtonian framework. It seems to me that questions 1 and 2 are similar in this regard, *mutatis mutandis*.

This first feature suggests a second feature. Although the question is posed within the Newtonian framework, and cannot be quite the same question if posed in other contexts, it explicitly cannot be answered within the Newtonian framework. As we have seen, the Newtonian concepts of gravitational and inertial mass are distinct. That the two masses always take the same value for any given body is contingent, though suggestive, within Newtonian physics. It is this contingency that forms the *explanandum* under consideration. The question would be quite different if inertial and gravitational mass were natively identified within the theory, or if the theory itself gave some indication for why their coincidence was only apparent. The reason the question arises is that the observed equivalence cannot be explained within the Newtonian framework. But this means that any appropriate answer will have to go beyond Newtonian physics. So we have a question posed in one theory that can only be adequately answered by appealing to another, presumably more general or fundamental, theory.

But what form could such an answer take? Einstein (1920) claimed that the observed equivalence between gravitational and inertial mass was an important factor in his development of GR, and that it can be taken as evidence in favor of the "general postulate of relativity." The principle that the two mass concepts should be identified, or equivalently the idea that free fall does not depend on mass, is often called the weak equivalence principle and continues to play a central role in some presentations of GR (see, for instance, Weinberg (1972, ch. 1)). From this point of view, the observed equivalence is explained by asserting that in a supervening theory (GR), no distinction is made between the two masses. Inertial and gravitational masses are simply the same. But there is something strange, and potentially misleading, about this answer to the original question. The reason for the difficulty is that a more accurate account of the situation in GR, using only concepts native to GR itself, would be to say that there is *only* inertial mass. It is not that gravitational mass is explicitly identified with inertial mass, but rather than gravitational mass has been stricken from the theory altogether. There *is no gravitational potential* in GR; nor is there a gravitational force. And so it makes no sense to ask how a body responds to a gravitational potential or how strongly it exerts a gravitational force on another body.[10] To say that gravitational and inertial masses are identified in GR is simply a confusion.[11] When one attempts to answer the original question be appealing to some supposed equivalence between the two kinds of mass in GR, one mixes terms from two theories in a way that is dubious and confusing.

These considerations suggest another response to our original question. Given that gravitational mass does not make sense in GR, one might say that the question turns out to be an error. We used to think that gravitation was a force (one might say), and that a body's gravitational mass determined the magnitude of that force exerted by and on the body. But now we know that gravitation is not a force at all and so questions about gravitational mass do not make any sense. This type of response is intended to dissolve, rather than answer, the question, by directing the questioner to a textbook on GR. But I claim that this kind of response is unsatisfying. First, the question was asked in a specific framework; to say that that framework is no longer widely accepted is irrelevant. Second, even if one accepts that GR supercedes Newtonian physics, and one accepts moreover that gravitational mass does not exist in GR, the question remains and *should have an answer*. In Newtonian physics, which everyone accepts as a predictively accurate theory, there are two different kinds of mass. As such, I can point to two kinds of roles that one might expect mass to play. Supposing that GR is correct that gravitational mass does not make sense, why does it seem

[10] This point is put clearly, for instance, by Sachs (1976), though it has not always been recognized by the physics community, as Weinberg (1972) attests.

[11] Einstein, of course, can be excused for making statements to this effect: he was building a new theory and, in Neurath's boat fashion, needed to work with what he had in order to make himself understood (and in order to understand what he was doing himself).

that inertial mass gets peeled apart into two separate things? Or equivalently, if we limit attention to gravitation on the scales at which Newtonian physics is effective, why do we find gravitational mass to be a useful concept, and moreover, why is it equal to inertial mass?[12]

I think this discussion helps to clarify both what question was originally being asked and what kind of answer would be appropriate. We have two fixed points to navigate between. First, the question is such that it cannot be answered by the Newtonian theory. New physics is necessary to answer the question. But second, it is a question that needs to be answered in the terms in which it was asked, i.e. within the Newtonian framework. As we have seen, using Newtonian terms within the framework of GR leads to inconsistencies and serious confusion, while using concepts native to GR at best allows a dissolutive response, rather than an answer, to the question. With these two points in mind, one might rephrase the question as follows. Given that we now believe GR to have superseded Newtonian theory as our best description of large-scale dynamics and gravitation, why are gravitational and inertial mass equal in Newtonian theory?[13] In this particular case, since it turns out that gravitational mass does not make sense in GR, there is an additional question hidden within the original question: if GR is right, why does Newtonian theory support two concepts of mass in the first place?

To answer these questions, one needs to show, in detail, how Newtonian theory relates to GR. One needs to show why, if GR is true, Newtonian theory is such an effective descriptor of how the world works, at least at certain distance/energy scales. One way of doing this

[12]Curiously, one can already imagine giving the same response to question 1: why should we compare gravitation to the other three forces? After all, we already know that there is no such thing as gravitational force! But this answer would be equally unsatisfying in that context, for the same reason: gravitation *is* conceived of as a force in modern particle physics, which is at least part of the difficulty in making quantum field theory and GR mesh. It is easy to imagine a similar response, too, to question 2: it may well turn out that these parameters are not important to our next class of theories. I think this serves to underline the curious character of the questions above: they are all expressed in the language of one theory, but one fully expects that the answer will come from a theory in which the terms of the question may not make any sense.

[13]Rephrasing the question in this way is only possible when one can point to the superseding theory. In the cases of questions 1 and 2, no superseding theory is known. These questions might be rephrased in terms of a future possible theory, or as a statement of a certain kind of research objective. One is looking for a new theory T that can tell us why, given that theory T supersedes the Standard Model (say), the parameters in the Standard Model take the particular values that they are observed to take.

would be to show that Newtonian theory can be reached from GR as an appropriate limit that captures (within GR) the circumstances in which Newtonian theory seems so effective.[14] Then one would try to show that if Newtonian theory is taken as the limit of GR in the appropriate way, then Newtonian theory *does* support two concepts of mass, and moreover, that for any body these two masses must be equal. In other words, one shows that gravitational mass arises in some way in the limiting process, and that the result must be equal to inertial mass.

It turns out that it is possible to perform this procedure exactly as described. (See appendix A.) There is a precise sense in which Newtonian theory is a limit of GR (Künzle, 1976; Ehlers, 1981; Malament, 1986a). It involves a two step process. One begins by considering a one-parameter family of relativistic spacetimes, parametrized by some variable λ. λ, at the present level of discussion, can be taken to reflect the inverse of the "speed of light"[15] in each of the spacetimes. By constructing this family of spacetimes carefully, one can consider the limit that the spacetimes approach as λ approaches 0 (corresponding to taking the speed of light to infinity). The result is a degenerate "classical" spacetime with many of the

[14] Many philosophers have questioned when and whether it is possible to show that an old theory reduces to a new theory (in the philosopher's parlance) or a new theory reduces to an old theory (in the language of working physicists) in the sense I have in mind here of a new theory explaining why an old theory worked. Nagel (1998), for instance, treated intertheoretic reduction as explanation (in the deductive-nomological (DN) sense) in a way that bears a rough family resemblance to what I am thinking of; Nickles (1998), meanwhile, argues that often explanatory reductions are not possible at all (at least in the DN sense). Curiously, Nickles identifies Newtonian physics and relativity theory as a prime example of a reduction relationship that is *not* an explanation in the DN sense, but rather a collection of rough intertheoretic relations. It seems to me that if *any* intertheoretic relationship deserves to be called deductive, it is the relation between Newtonian theory and GR. But I do not intend to enter a debate on intertheoretic reduction here. Instead, I want to distinguish *identifying* reduction and explanation (as Nagel does) from a more ambiguous demand that a new theory explain, at least in some sense, why our old theories succeeded. One way of cashing this requirement out is to say that a new theory cannot make predictions that are inconsistent with the successes of previous theories (in which case the new theory will at least reveal regularities captured by the old theory). At least in physics, if a new theory is inconsistent with the success of an old theory, it is perceived as a major difficulty for the new theory. One can see this tension, for instance, in the relation between quantum mechanics and Newtonian physics. The apparent inability to recover Newtonian theory as some limit from quantum mechanics, though not a barrier to the acceptance of quantum mechanics, remains a central problem for working physicists. In any case, in the present example such an explanation *is* possible, and I maintain that parallel explanations are demanded by each of the questions asked in the introduction.

[15] There is an abuse of language, here. Really, λ parametrizes something about the metrical structure of the spacetime, specifically how wide the light cones are at a point. But the widths of the light cones indicates how null vectors, which are the possible tangent vectors for the wordlines traversed by light, relate to the timelike and spacelike vectors in the spacetime. Hence wider lightcones indicate "faster" light.

characteristic features of Newtonian physics: space is always flat and Euclidean; there is a unique sense of space at a time and absolute simultaneity holds; the spacetime is Galilean relativistic, which means that (1) measurements of elapsed time and the distance between simultaneous events will be the same for all observers, irrespective of their motion and (2) there is no absolute standard of rest. But we have not yet recovered Newtonian physics. Rather, we have reached an intermediate point between GR and Newtonian physics, which I will call geometrized Newtonian theory.[16,17] Geometrized Newtonian theory is classical in the sense that the spacetimes it permits are classical (in the sense just described). But in geometrized Newtonian theory, gravitation is still geometrical: the geometry of spacetime is curved, with curvature determined by the distribution of matter in the universe, and gravitational effects are manifestations of the resulting geometry. In this sense geometrized Newtonian theory is like GR, though with a different spacetime structure and with different fundamental equations. Importantly, since gravitation is geometrical rather than a force between bodies, gravitational mass does not make any more sense in the context of geometrized Newtonian theory than in GR.

To see where gravitational mass comes from, we need to take the second step in the limiting process. This step makes use of a theorem known as the Trautman Recovery Theorem (Malament, 2010, Prop. 4.2.5), due to Andrzej Trautman. Trautman's theorem tells us that, given a classical spacetime of the sort found in geometrized Newtonian theory, satisfying certain conditions, one can find[18] (1) another spacetime that is flat,[19] and (2) a scalar field $\varphi_{\mathcal{G}}$

[16]Geometrized Newtonian theory is sometimes called Newton-Cartan theory, as it was first developed during a lecture course by Élie Cartan in the early 1920s, as an attempt to understand how Newtonian gravitation related to GR (Cartan, 1923, 1924). Strictly speaking geometrized Newtonian theory, rather than standard Newtonian physics, is the classical limit of GR.

[17]It may be helpful to emphasize, here, that geometrized Newtonian theory is *not* a model within GR that is somehow suggestive of classical physics. It is an independent theory with the empirical content of Newtonian physics. Indeed, a classical spacetime as described here is not and could not be a relativistic spacetime because it does not have an appropriate metric structure. We really have left GR when we arrive at the classical limit. As we shall see, there is a strong sense in which standard Newtonian physics can be recovered from geometrized Newtonian theory, and so the "Newton" in its name is well justified.

[18]See theorem A.4 for details of this claim, and for a precise statement of the conditions a (geometrized) classical spacetime needs to satisfy in order to recover standard Newtonian theory, which are not trivial.

[19]A classical spacetime as considered by geometrized Newtonian theory is generally curved, though *space* is always flat. In standard Newtonian physics, spacetime taken as a whole has to be flat.

that satisfies Poisson's equation (i.e. that has the dynamical relationship to the distribution of matter in the universe that Newton's theory predicts for the gravitational potential) and which is such that for any free (inertial) massive test point particle,

$$\mathbf{a} = -\nabla\varphi_{\mathcal{G}}.^{20} \tag{2.2}$$

In other words, under certain circumstances, we can recover a flat spacetime and a *gravitational potential* $\varphi_{\mathcal{G}}$ that has just the relations to both the distribution of matter in the universe and the dynamics of a particle that we would expect from Newtonian physics. We have now recovered full-blown, standard Newtonian physics as a limit of GR.

I want to draw attention to an important feature of Eq. (2.2). It is a derived relation between the acceleration of a particle and the gradient of the gravitational potential. Compare Eq. (2.2) with Eq. (2.1). They differ by a proportionality term, $m_{\mathcal{G}}/m_{\mathcal{I}}$. In other words, Eq. (2.2) is just what we get if we set $m_{\mathcal{I}} = m_{\mathcal{G}}$ in (2.1). In the process of recovering Newtonian theory from GR, we have shown that *gravitational and inertial mass must be equal.* Another way of getting to the same conclusion would be as follows. Eq. (2.2) just tells us the relationship between acceleration and the gravitational potential. The gravitational mass was defined as the *coupling* to the gravitational field: in other words, it is the constant of proportionality moderating the relationship between force and the gravitational field. If we use $\mathbf{F} = m_{\mathcal{I}}\mathbf{a}$,[21] we find that the force on a massive test point particle arising from the gravitational potential $\varphi_{\mathcal{G}}$ is

$$\mathbf{F} = m_{\mathcal{I}}\mathbf{a} = -m_{\mathcal{I}}\nabla\varphi_{\mathcal{G}}. \tag{2.3}$$

Eq. (2.3) tells us directly that the coupling to the gravitational field in Newtonian physics

[20] Here I am using the notation of standard Newtonian physics.

[21] It might be worth emphasizing that $\mathbf{F} = m_{\mathcal{I}}\mathbf{a}$ holds for massive point particles generally in GR, geometrized Newtonian theory, and standard Newtonian gravitation. Sometimes one hears that "GR does not have forces, whereas Newtonian physics does," but this is not correct. There are forces in GR; the difference with Newtonian physics is that *gravitation* is not a force in GR. One way of seeing why this should be is that the worldlines of massive particles under the influence of only gravitational effects are *non-accelerating*, which means that the force acting on them (in GR) vanishes.

is given by the *inertial* mass. The reason that gravitational and inertial mass are always equal is that gravitational mass simply is inertial mass. And thus we have an answer to the original question.

3 What have I just done?

Now that I have offered an answer to one of my questions, I can ask what kind of explanation I have given. It seems clear without further argument that this explanation is *not* a causal explanation, in any of the senses of causal explanation that have been articulated over the last few decades.[22] It likewise does not fit into any of the earlier statistical accounts of explanation, such as the statistical relevance account of explanation or the inducto-statistical account.[23] For one, in all of these cases (causal/mechanical, SR, IS), one explains an *event* or perhaps a class of events. The explanation under consideration in this paper regards a more general feature of the world, in particular as it relates to a specific theory.[24] For another, the present explanation is (at least roughly) an *argument*; usually, causal explanations take the form of a narrative of events that lead to the event to be explained. In the remainder of this section, I will focus on two other prominent accounts of explanation that, at least *prima facie*, have a better chance of capturing the kind of explanation I gave in the previous section: the deductive-nomological (DN) account and the unificationist account.[25]

[22]Some prominent examples are of course Salmon (1984), but also Cartwright (1983) and Woodward (2003). Strevens (2004, 2008) also offers a kind of causal account of explanation, though his "kairetic account" also includes some of the desirable features of the unificationist account.

[23]For more on either of these, see (Salmon, 1989).

[24]See, too, Kitcher (1989, Sec. 3.3) for more on the idea that the causal account has difficulty with explanation of theoretical regularities, as opposed to propositions concerning singular events.

[25]Batterman (2002) also discusses a form of explanation that is not well-treated by the causal and unificationist accounts. He dubs it "asymptotic explanation." I think that he has correctly identified a form of explanation that the received accounts miss; however, I want to emphasize that the present example is strikingly different from the examples Batterman offers. Asymptotic explanation involves explanations of "universality"—properties that families of systems have at a given distance scale, irrespective of their microscopic details. His most familiar example concerns phase changes: virtually all substances undergo phase changes (from gas to liquid, liquid to solid, etc.), even though they have dramatically different microscopic properties. The form of the explanation for such universal features involves the renormalization group, which is a method for moving between different levels of description of a physical system. The form of explanation described here relies on a very different kind of limiting procedure and concerns relations between theories, not between distance scales.

3.1 The Deductive-Nomological Account

The DN account of explanation, originally proposed by Hempel and Oppenheim (1948), was long the received view of explanation. On this account, an explanation is a (logical, first-order) argument by which the thing to be explained, the *explanandum*, is deduced from a set of true premises, the *explanans*. It is taken to be necessary that the *explanans* include at least one law of nature; generally, it will also include particulars such as initial conditions or boundary conditions. The thing to be explained is a proposition. The intuition is that an explanation is a demonstration of law-like expectability. To explain a proposition is to show that one could expect it to be true, given the laws of nature and some given set of circumstances.[26]

Does the explanation I give in section 2 fit the DN mold? It certainly has many of the central features of a DN explanation. The explanation consists of an argument by which the *explanandum* is derived. The *explanans* is perhaps a bit broad: all of the central principles of GR must be included in order to set up the limiting process necessary to recover geometrized Newtonian theory. But among these principles are several law-like propositions, and at least one law plays a central role in the explanation insofar as $\mathbf{F} = m_{\mathcal{I}}\mathbf{a}$ enters at the end of the story to show that inertial mass provides the coupling of a body to the Newtonian gravitational potential.[27] Finally, there is a strong sense in which the argument's explanatory power comes from the fact that the result is derived from central principles of GR, which means that the explanation is nomological in an important way.

But not all is well. This explanation poses some difficulties for the DN model, too. First, the present *explanandum* is most naturally expressed as a proposition in second-order logic. We are trying to explain why in all propositions concerning inertial and gravitational mass (or, in all instances of Newton's force law), the two values appear to be equal. Hence we

[26]See, as ever, Salmon (1989). I am glossing over many difficulties concerning what might count as a law of nature and what kind of deduction is necessary (for instance, are all explanations deductions in first-order logic?).

[27]Other laws, such as Einstein's equation, must also be included in the *explanans* because they are necessary for constructing the limiting process.

are required to quantify over sentences of the logic. In first-order logic, it is not possible to express the full force of the conclusion: one can show that in any particular instance, gravitational mass is equal to inertial mass, but not that inertial mass is *always* equal to gravitational mass, which is what was demanded. Is this an insurmountable problem for the DN model? As typically described (as, for instance, in Salmon (1989)), the DN account begins with a first-order logic. But I do not see that first-order deductions are a necessary feature of the account; one can easily imagine extending the account to include more general deductions. In any case, a second-order derivation is certainly in the spirit of the DN model.

A second worry, however, is more troubling. It is not clear that, construed on the DN model, the present explanation is possible as given. In particular, at the last step of the explanation I re-interpreted the term $m_{\mathcal{I}}$ as the gravitational mass with the argument that the gravitational mass is defined in Newtonian theory as the term that moderates the strength of a body's coupling to a gravitational potential. But this move is not open within a DN explanation since the terms of the language are fixed in advance. The re-interpretation of the term $m_{\mathcal{I}}$ on the right hand side of Eq. (2.3) as gravitational mass necessary to show that insofar as Newtonian gravitation emerges in an appropriate limit from GR, gravitational mass *is* inertial mass is not possible within the fixed logical framework.[28] At best, one can show the somewhat weaker result that a term corresponding to gravitational mass is always equal to a term corresponding to inertial mass by beginning with an augmented language that already includes gravitational mass (which would not otherwise be natural in GR). This result gives the nomic expectability characteristic of the DN account: one can show that one can always expect inertial mass and gravitational mass to be equal. But without the additional re-interpretation, one has not successfully explained *why* they are equal, at least not in the way that one can if the last step of the explanation is permitted. And so we have a situation in which substantial explanatory power is contained in a step of the explanation

[28]Does the logical structure need to be fixed in advance? If an explanation is really to be understood as an argument in the strictest sense, I think that it does. The interpretation of the terms enters in the semantics of the logic; an argument, meanwhile, is syntactic. There cannot be a logical operation that forces a re-interpretation of the terms that one began with, as would be necessary.

that would not be expressible if we had to begin with a fixed set of terms coming from GR.

So the present explanation bears some family resemblance to DN explanations, though it does not have just the form described by Hempel and Oppenheim (because it requires second-order logic). More troubling, the DN account does not permit the final (important) step of the explanation. All that said, if one were committed to the claim that *all* explanations, or at least all deductive explanations, need to fit the DN model, it may be possible to extend the model to include the present example. Alternatively, if disinclined to extend the DN model, one might bite the bullet and claim that the last step of the present explanation does not have any explanatory power. I am not inclined to make either of these moves, however. Insofar as I already accept a pluralistic view about explanation, I do not see any virtues in a procrustean reading of the present explanation as an example of DN explanation.

But suppose that one is not convinced that the two concerns I have raised pose problems for a DN reading of the present example. Even so, many writers have pointed out that Hempel and Oppenheim offer neither necessary nor sufficient conditions for an argument to be explanatory (Godfrey-Smith, 2003; Salmon, 1989; Kitcher, 1989). In other words, even if the present explanation *does* fit the DN model, it does not follow that it is explanatory *because* it fits that model. One still needs to give an account of what makes a given argument explanatory, even if it has a law as a necessary premise. A prominent attempt to describe what additionally may be required to make an argument explanatory is given by the unificationist account of explanation, which I will turn to presently.

3.2 The Unificationist Account

Kitcher (1989) provides the authoritative manifesto on the unificationist account of explanation, so I will focus on the version of the account given there.[29] The basic idea of the unificationist account of explanation is that science aims to explain phenomena by showing how a phenomenon fits into a unified systematization of our beliefs. Kitcher makes this idea

[29] For more on the history of the unificationist account, see Kitcher (1989) and Salmon (1989), as well as references therein.

precise in the following way. Suppose that K is the set of all statements endorsed by the scientific community. An acceptable explanation is a member of the *explanatory store* over K, denoted $E(K)$. How can we characterize the members of $E(K)$? First, we say that $E(K)$ is a set of arguments by which some members of K are derived from other members of K. But in general there will be many such sets of arguments. To pick one, Kitcher introduces *general argument patterns*, ordered triples consisting of (1) a *schematic argument*, which are sequences of sentences with key terms replaced by dummy variables (i.e. sequences of *schematic sentences*); (2) sets of sets of filling instructions corresponding to each schematic argument, where a given set of filling instructions tells you how to fill in the dummy variables of a schematic sentence; and (3) a classification of the schematic argument, which tells you which of the sentences in the argument are supposed to count as premises and which are conclusions. One general argument pattern is more *stringent* than another if its classification and the structure of its schematic sentences together make it more difficult to instantiate. Given this machinery, Kitcher says that $E(K)$ consists of the (not necessarily unique)[30] minimal set of maximally stringent general argument patterns from which a maximal number of conclusions can be drawn. More roughly, one wants to find the smallest subset of K from which the other members of K can be derived, using the fewest possible stringent argument patterns.

As on the DN account, an explanation on the unificationist account is an argument. But the unificationist account offers an additional set of constraints on the kinds of arguments that count as explanatory, above and beyond the conditions of the DN account. It is not sufficient (or necessary) that a given conclusion be derived from a set of premises including a law, as in the DN model; now, an argument is explanatory if it is a member of the explanatory store of a set of statements endorsed by the scientific community.[31] To understand whether

[30]See Kitcher (1989, pg. 435). In general, Kitcher seems to think that the explanatory store over K may not be determined by the constraints given. In historical cases, however, he claims that these constraints *are* mutually sufficient to determine the explanatory store of K.

[31]On Kitcher's account, one would add an "only if" to this last sentence. But if we adopt a pluralist view of explanation on which some, but not necessarily all, explanations are unificationist explanations then we want to understand the condition of membership in $E(K)$, as defined by Kitcher, to be only a sufficient

the explanation in section 2 is a unificationist explanation, then, we want to determine whether we should expect it to be in the explanatory store over K. There are several reasons to think that it should not be.

First, it is not clear that the sentence "gravitational mass and inertial mass are equal" is a member of K. Kitcher assumes that K is deductively closed and consistent. Thus, if the statements of GR are members of K, then "there is no such thing as gravitational mass" is also a member of K. Perhaps one can consider statements of the form "in such and such theory, X" as members of K. But this seems like a strange move for a unificationist to make. It seems more in the spirit of the unificationist approach to skip over explanations of such uniformities in the Newtonian context and go right to explaining the motion of individual bodies in a way in which one never introduces the Newtonian concept of gravitational mass at all. Indeed, on Kitcher's account of reduction, one would reduce Newtonian gravitation to GR by showing that the general argument patterns of Newtonian gravitation can be recovered from and extended as general argument patterns of GR. But then one would not expect to be able to explain features of the world that can only be expressed in the Newtonian framework, because the very goal of the reduction would be to move one's explanatory arguments *out* of the Newtonian framework and into GR. The ideal would be able to explain all of the phenomena that can be explained by Newtonian physics with GR alone, eschewing the Newtonian argument patterns altogether. The explanation given above would not be in $E(K)$ because, once GR is endorsed by the scientific community, the *explanandum* of question 3 is no longer a member of K and thus not the kind of sentence that calls for explanation.

Another worry comes from the opposite direction. Even if K were construed in such a way that the *explanandum* were in K, it is still not clear that the explanation I have given would be included in the explanatory store. For one, even if you neglect explanations of the sort I have given, you can still explain *all the same phenomena* in the world. That is, adding

condition. A given explanation may be explanatory by virtue of being in $E(K)$, but it is possible that there are other kinds of explanations as well.

an additional singular argument of the sort I have given makes the explanatory store larger, without adding any payoff in terms of explanations of particulars. Given the conditions on choosing $E(K)$, one could conclude that the explanation I have given should not be in $E(K)$ even if the *explanandum* were included (since if it were included, $E(K)$ would violate the minimality condition).

A third worry is more general. It is difficult to see what kind of argument pattern is being executed in the present example. The thing to be explained is a singular feature of the Newtonian theory. Neither the *explanandum* nor the explanation itself can be schematized without losing its essential character. One might be able to schematize the argument by reconstruing it as an argument concerning the inertial and gravitational masses for particular bodies, and then include a general argument pattern by which one shows that for any given body, the inertial and gravitational masses are identical. But this is not what was originally demanded and, as with the DN model, it is not clear that such explanations answer the original question. Really we want to know why inertial and gravitational mass are *always* the same, not why they happen to take the same value in any variety of situations.

I think that these points suggest that the explanation I have given is an awkward fit with the unificationist account of explanation, too. Once again, this does not mean that the unificationist account cannot be adapted to fit the explanation I have given. I think it probably can. Indeed, it is hard to prove that a given explanation, if successful, is not a member of the maximally unified set of arguments over a set of beliefs. My point is rather that to make the present explanation fit with the details of Kitcher's account, some adaptation of the unificationist position is likely necessary. More importantly, even if one can find a modified unificationist account that would fit more naturally with the present explanation, it is not clear that it would do justice to the explanation I have given. In other words, it does not seem that the *reason* that the explanation I have given is explanatory has anything to do with the fact that it is an instantiation of a general, stringent, and unified argument pattern. Indeed, I have suggested that it is not an instantiation of a general argument pattern at all—it is a singular explanation of a general feature of Newtonian theory.

If anything, attempting to adapt the unificationist picture to include explanations of the present sort threatens to gloss over the important features of an explanation that are not otherwise native to the unificationist account of explanation. I would rather say that very many explanations in science have the character that Kitcher describes: they consist of arguments that are explanatory by virtue of how they fit some particular or regularities in a systematized body of knowledge. I find Kitcher's analysis of his own examples convincing, and I think that a broad class of explanations in physics fit well with the unificationist picture of explanation. Modifying the unificationist account, which gets so many interesting cases right, seems counterproductive. What we have here is simply a different kind of explanation.

4 Let a thousand flowers bloom

So far, I have argued that the explanation given in section 2 is neither a causal explanation nor any kind of statistical explanation. It is, at least roughly, a deductive argument, though it cannot be made to fit the strict logical structure demanded by the DN model without sacrificing a crucial explanatory step. Even if it *could* be made to fit with the DN model (or if the DN model could be expanded to include such explanations), one would still owe an account of why this *particular* deductive-nomological argument is explanatory, given that the Hempel-Oppenheim criteria are not sufficient alone. The leading candidate to provide the additional sufficient conditions for deductive arguments to count as explanatory is the unificationist account; I have argued, however, that as presented by Kitcher (1989), the unificationist account does not allow for explanations of the sort I have given. Indeed, the trouble is not necessarily with the unificationist account; rather, the explanation I have given is simply not explanatory by virtue of its status as an instantiation of a maximally unified argument pattern. It should be clear from my remarks in section 1 that I do not take these accounts' inability to deal with the present explanation as an argument against the accounts. I have simply presented a scientific explanation of a different sort (and there are many possible sorts).

The remaining work is simply to identify some of the features of the present explanation and to try to characterize what makes it explanatory. I have identified the most important features along the way, but it seems worthwhile to tie them together now. First, as we have seen, the present explanation is a deductive argument. But it is a *singular* deductive argument: it explains a broad generality in Newtonian physics and there are no other generalities that can be explained by the same argument.[32] Moreover, the explanation requires second-order logic to capture the spirit in which the question is asked. Most importantly, there is a crucial step of translation involved in the theory that makes it difficult to fully capture in purely logical form. This translation is necessary because the question is asked in the language of one theory (indeed, it only makes sense in that theory), but it is of a form that *a fortiori* cannot be answered without appealing to physics that goes beyond the theory within which it is asked. Meanwhile, it demands an answer on its own terms: it is not enough to *dissolve* the question by appealing to a superseding theory. The explanatory demand is to show how, given some superseding theory, a general fact within an old theory that seemed unexplainable given the apparatus of that theory is really necessary or to be expected. The explanation consists in filling in the details of a general response to such questions along the following lines: such and such is the case in *this* theoretical framework because really we believe this *other*, superseding theory more.

I take it that the explanatory work here is done by spelling out how two theories relate to one another and showing that, taken as a limit of a superior theory, an old theory must display the regularity in question. It is this relation between an old theory and a preferred superseding theory what makes the present argument explanatory. For these reasons, one might call explanations of the sort I have given intertheoretic explanations, or perhaps aspirational explanations, since the questions that call for such explanations are often demands

[32] That is not to say that there are not other generalities of Newtonian theory that can be explained by the limiting process I have described. There are. Two examples are the fact that space is flat in Newtonian theory (Malament, 1986b) and that momentum is conserved. But the explanations for these differ in important details from the explanation I have given, even though they are also explained via the limit from GR.

for new theories. This last feature reiterates something I have already suggested, that such explanations are particularly important: these explanations guide inquiry by setting out the questions that a new theory is expected to answer (once appropriate translation work into the old theory has been done). That said, they are not the only kinds of explanation that could serve in this role. And as Geroch suggests in the above quoted passage, to say any more would be a mistake.

A Technical Details of Answer to Question 3

In the body of the paper I offer an answer to the question, "Why are inertial and gravitational mass equal in Newtonian physics?" This answer concerns the sense in which standard Newtonian gravitational theory can be understood as a limit of GR, via geometrized Newtonian theory. The explanation offered in the main text is reasonably precise, but non-technical; here I offer the technical details of that argument. This appendix is not intended as a complete or pedagogical introduction to geometrized or covariant formulations of Newtonian gravitation (or GR, the treatment of which will be especially brief). I will provide only a brief, formal review of geometrized and covariant standard Newtonian theory and GR to establish notation and terminology. For a systematic treatment of these subjects, including proofs of the theorems stated here and an explanation of the "abstract index notation" I will use throughout, see Malament (2010, Esp. Ch. 4). For a systematic treatment of the sense in which geometrized Newtonian theory arises as a limit from GR, see Malament (1986a); Künzle (1976); Ehlers (1981).

A.1 Preliminary definitions

We begin by defining the geometrical structures we will work with. First, to fix notation and sign conventions, we will describe GR.

Definition A.1 *A* relativistic spacetime *is an ordered pair* (M, g_{ab})*, where* M *is a smooth, connected, four-dimensional manifold and* g_{ab} *is a smooth, non-degenerate semi-Riemannian metric on* M *with Lorentz signature* $(+,-,-,-)$*.*

In a relativistic spacetime, the metric defines a lightcone structure at every point as follows. Given any point p and any vector ξ^a in the tangent space M_p, we say that ξ^a is *timelike* if $g_{ab}\xi^a\xi^b > 0$, *spacelike* if $g_{ab}\xi^a\xi^b < 0$, and *null* if $g_{ab}\xi^a\xi^b = 0$. The length of any vector ξ^a at a point is given by $||\xi^a|| = |g_{ab}\xi^a\xi^b|^{1/2}$. A (smooth)[33] curve is timelike (resp. spacelike or null) if its tangent vector is at every point of the curve. A spacetime is *temporally orientable* if there exists a continuous timelike vector field on all of M; such a vector field determines a *temporal orientation.* If a relativistic spacetime has a temporal orientation, then it is possible to consistently distinguish between future- and past-directed timelike vector fields. In what follows, we assume all relativistic spacetimes have a temporal orientation.

Since the metric is non-degenerate, there exists an inverse metric g^{ab} such that $g^{an}g_{nb} = \delta^a{}_b$. We can move easily between vector fields and covector fields on M by "raising" and "lowering" indices with g^{ab} and g_{ab}, respectively, so for instance if ξ^a is a vector field on M, then $\xi_b = g_{ab}\xi^a$ is a covector field on M, and likewise for more complicated tensor fields. The metric determines a unique derivative operator on M, ∇_a, satisfying the compatibility condition $\nabla_a g_{bc} = \mathbf{0}$. The derivative operator allows us to define the *curvature* of spacetime via the *Riemann curvature tensor*, which is the unique smooth tensor field $R^a{}_{bcd}$ such that for all smooth vector fields ξ^b, $R^a{}_{bcd}\xi^b = -2\nabla_{[c}\nabla_{d]}\xi^a$. We say that a spacetime is *flat* if $R^a{}_{bcd} = \mathbf{0}$. From the Riemann curvature tensor, we can define the *Ricci tensor* by $R_{ab} = R^n{}_{abn}$.

Massive point particles are represented by their worldlines, which are smooth future-directed timelike curves parametrized by arc-length. (Point particles have an attenuated status here—really, we are thinking of a field theory, and point particles are some appropriate idealization.) With every point particle, there is an associated *four-momentum*, P^a, defined at every point of the particle's worldline, whose length is the (inertial) *rest mass.* For a point particle with non-zero mass $m_{\mathcal{I}}$,[34] we can write $P^a = m_{\mathcal{I}}\xi^a$, where ξ^a is the tangent vector

[33]Here and in what follows, it should be assumed that we are limiting attention to smooth (i.e. infinitely differentiable) curves, fields, manifolds, etc, whether stated explicitly or not.

[34]Since keeping track of the distinction between inertial and gravitational mass is important for the ultimate moral of the present discussion, I will label masses as inertial even in the context of GR and geometrized Newtonian gravitation, were strictly speaking there can be no ambiguity. In keeping with the notation from the body of the paper, I will use capitalized calligraphic symbols for subscripts indicating labels to distinguish

field to the particle's worldline (called the particle's *four-velocity*). More generally, we can associate with any matter field a smooth symmetric field T^{ab}, called the energy-momentum tensor. T^{ab} can be thought to encode the four-momentum density of the matter field as determined by any future-directed timelike observer at a point: For all points $p \in M$ and all unit, future-directed timelike vectors at p, ξ^a, the four-momentum of a matter field at p as determined by ξ^a is $P^a = T^a{}_b\xi^b$. The curvature of spacetime is related to the energy-momentum tensor by *Einstein's equation*,

$$R_{ab} = 8\pi(T^{ab} - \frac{1}{2}Tg_{ab}), \tag{A.1}$$

where $T = T^a{}_a$.

We can now proceed to define a parallel structure for classical theories.

Definition A.2 *A* classical spacetime *is an ordered quadruple* $(M, t_{ab}, h^{ab}, \nabla)$*, where* M *is a smooth, connected, four-dimensional manifold;* t_{ab} *is a smooth symmetric field on* M *of signature* $(1, 0, 0, 0)$*;* h^{ab} *is a smooth symmetric field on* M *of signature* $(0, 1, 1, 1)$*; and* ∇ *is a derivative operator on* M *compatible with* t_{ab} *and* h^{ab}*, i.e. it satisfies* $\nabla_a t_{bc} = \nabla_a h^{bc} = \mathbf{0}$*. We additionally require that* t_{ab} *and* h^{ab} *are orthogonal, i.e.* $t_{ab}h^{bc} = \mathbf{0}$*.*

Note that "signature," here, has been extended to cover the degenerate case. We can see immediately from the signatures of t_{ab} and h^{ab} that neither is invertible. Hence in general neither t_{ab} nor h^{ab} can be used to raise and lower indices.

t_{ab} can be thought of as a temporal metric on M in the sense that given any vector ξ^a in the tangent space at a point, p, $||\xi^a|| = (t_{ab}\xi^a\xi^b)^{1/2}$ is the temporal length of ξ^a at that point. If the temporal length of ξ^a is positive, ξ^a is timelike; otherwise, it is spacelike. At any point, it is possible to find a covector t_a, unique up to a sign, such that $t_{ab} = t_a t_b$. If there is a continuous, globally defined vector field t_a such that at every point, $t_{ab} = t_a t_b$, then the spacetime is temporally orientable (we will encode the assumption that a spacetime is temporally oriented by replacing t_{ab} with t_a in our definitions of classical spacetimes). h^{ab},

them from subscripts indicating index (i.e. tensor) structure. So, $\mathcal{I}$ will be used to indicate "inertial" and $\mathcal{G}$ to indicate "gravitational."

meanwhile, can be thought of as a spatial metric. However, since there is no way to lower the indices of h^{ab}, we cannot calculate the spatial length of a vector directly. Instead, we rely on the fact that if ξ^a is a spacelike vector (as defined above), then there exists a covector σ_a such that $\xi^a = h^{ab}\sigma_b$. The length of ξ^a can then be defined as $(h^{ab}\sigma_a\sigma_b)^{1/2}$. (If ξ^a is not a spacelike vector, then there is no way to assign it a spatial length.) Note, too, that it is possible to define the Riemann curvature tensor $R^a{}_{bcd}$ and the Ricci tensor R_{ab} with respect to ∇ as in GR (or rather, as in differential geometry generally). Flatness ($R^a{}_{bcd} = \mathbf{0}$) carries over intact from GR; we say a classical spacetime is *spatially flat* if $R^{abcd} = R^a{}_{nmq}h^{bn}h^{cm}h^{dq} = \mathbf{0}$. It turns out that this latter condition is equivalent to $R^{ab} = h^{an}h^{bm}R_{nm} = \mathbf{0}$.[35]

We describe matter in close analogy with GR. Massive point particles are again represented by their worldlines, which are smooth future-directed timelike curves parameterized by elapsed time. For a point particle with (inertial) mass $m_{\mathcal{I}}$, we can always define a smooth unit vector field ξ^a tangent to its worldline, again called the four-velocity, such that we can define a four-momentum field $P^a = m_{\mathcal{I}}\xi^a$. The mass of the particle is now given by the temporal length of its four-momentum. In similar analogy to the relativistic case, we can associate with any matter field a smooth symmetric field T^{ab}, now called the mass-momentum tensor. T^{ab} once again encodes the four-momentum density of the matter field as determined by a future directed timelike observer at a point, but in this case all observers agree on the four-momentum density at p: $P^a = t_bT^{ab}$. Contracting once more with t_b yields the mass density, $\rho = t_at_bT^{ab}$.

In the present covariant, four-dimensional language, standard Newtonian theory can be expressed as follows. Let (M, t_a, h^{ab}, ∇) be a classical spacetime. We require that ∇ is flat (i.e. $R^a{}_{bcd} = \mathbf{0}$). We begin by considering the dynamics of a test point particle with inertial mass $m_{\mathcal{I}}$ and four-velocity ξ^a. We can define the force on such a particle by $F^a = m_{\mathcal{I}}\xi^a\nabla_a\xi^b$ (literally, $\mathbf{F} = m_{\mathcal{I}}\mathbf{a}$). In the absence of external forces, a massive test point particle undergoes geodesic motion. If the total mass-momentum content of spacetime is described by T^{ab}, we

[35]See Malament (2010, Prop. 4.15).

require that the conservation condition holds, i.e. at every point $\nabla_a T^{ab} = \mathbf{0}$. To add gravitation to the theory, we can represent the gravitational potential as a smooth scalar field φ on M. φ is required to satisfy Poisson's equation, $\nabla_a \nabla^a \varphi = 4\pi\rho$ (where ∇^a is shorthand for $h^{ab}\nabla_b$). Gravitation is considered a force; in general, the gravitational force on a point particle is moderated by its gravitational mass, according to $F^a_{\mathcal{G}} = -m_{\mathcal{G}} \nabla^a \varphi$. (Indeed, this relationship can be taken as a definition of the gravitational mass.)

In geometrized Newtonian theory we again begin with a classical spacetime (M, t_a, h^{ab}, ∇), but now we allow ∇ to be curved. The dynamics of a point particle with inertial mass $m_{\mathcal{I}}$ and four-velocity ξ^a are again given by $F^a = m_{\mathcal{I}} \xi^n \nabla_n \xi^a$; likewise, free massive test point particles undergo geodesic motion. However, the geodesics are now determined relative to ∇, which is not necessarily flat. The conservation condition is again expected to hold. Gravitation is no longer a force and so there is no longer a "gravitational mass" term in the theory. Instead, gravitational interactions are seen to be the result of the curvature of spacetime, which in turn is determined by a geometrized form of Poisson's equation,

$$R_{ab} = 4\pi\rho t_a t_b. \tag{A.2}$$

Since the Riemann curvature tensor (and by extension, the Ricci tensor) is determined by ∇, the geometrized Poisson's equation places a constraint on the derivative operator. In particular, ∇ must be such that, for all smooth vector fields ξ^a, $R_{ab}\xi^a = -2\nabla_{[b}\nabla_{n]}\xi^n = 4\pi\rho t_a t_b \xi^a$. Note, too, that the geometrized Poisson's equation forces spacetime to be spatially flat, because if Poisson's equation holds, then $R^{ab} = 4\pi\rho h^{an} h^{bm} t_n t_m = \mathbf{0}$ by the orthogonality condition on the metrics.

A.2 Relations between the theories

We are particularly interested in the relationship between these three theories. Several results are available. First, it is always possible to "geometrize" a gravitational field on a flat classical spacetime—that is, we can always move from the covariant formulation of standard Newtonian gravitation to geometrized Newtonian gravitation, via a result due to Andrzej Trautman (1965).

Proposition A.3 (Trautman Geometrization Lemma.) (Slightly modified from Malament, 2010, Prop. 4.2.1.) *Let* $(M, t_a, h^{ab}, \overset{f}{\nabla})$ *be a flat classical spacetime. Let* φ *and* ρ *be smooth scalar fields on* M *satisfying Poisson's equation,* $\overset{f}{\nabla}_a \overset{f}{\nabla}{}^a \varphi = 4\pi\rho$. *Finally, let* $\overset{g}{\nabla} = (\overset{f}{\nabla}, C^a{}_{bc})$,[36] *with* $C^a{}_{bc} = -t_b t_c \overset{f}{\nabla}{}^a \varphi$. *Then* $(M, t_a, h^{ab}, \overset{g}{\nabla})$ *is a classical spacetime;* $\overset{g}{\nabla}$ *is the unique derivative operator on* M *such that given any timelike curve with tangent vector field* ξ^a,

$$\xi^n \overset{g}{\nabla}_n \xi^a = \mathbf{0} \Leftrightarrow \xi^n \overset{f}{\nabla}_n \xi^a = -\overset{f}{\nabla}{}^a \varphi; \tag{G}$$

and the Riemann curvature tensor relative to $\overset{g}{\nabla}$, $\overset{g}{R}{}^a{}_{bcd}$, *satisfies*

$$\overset{g}{R}_{ab} = 4\pi\rho t_a t_b \tag{CC1}$$

$$\overset{g}{R}{}^{a}{}_{b}{}^{c}{}_{d} = \overset{g}{R}{}^{c}{}_{d}{}^{a}{}_{b} \tag{CC2}$$

$$\overset{g}{R}{}^{ab}{}_{cd} = \mathbf{0}. \tag{CC3}$$

Trautmann showed that it is also possible to go in the other direction. That is, given a curved classical spacetime, it is possible to recover a flat classical spacetime and a gravitational field, φ—so long as the curvature conditions (CC1)-(CC3) are met.

Proposition A.4 (Trautman Recovery Theorem.) (Slightly modified from Malament, 2010, Prop. 4.2.5.) *Let* $(M, t_a, h^{ab}, \overset{g}{\nabla})$ *be a classical spacetime that satisfies eqs.* (CC1)-(CC3) *for some smooth scalar field* ρ. *Then, at least locally on* M, *there exists a smooth scalar field* φ *and a flat derivative operator on* M, $\overset{f}{\nabla}$ *such that* $(M, t_a, h^{ab}, \overset{f}{\nabla})$ *is a classical spacetime;* (G) *holds; and* φ *and* $\overset{f}{\nabla}$ *together satisfy Poisson's equation,* $\overset{f}{\nabla}_a \overset{f}{\nabla}{}^a \varphi = 4\pi\rho$.

It is worth pointing out that the pair $(\overset{f}{\nabla}, \varphi)$ is not unique. It is also worth pointing out that whenever we begin with standard Newtonian theory and move to geometrized Newto-

[36] This notation is explained in Malament (2010, Prop. 1.7.3). Briefly, if ∇ is a derivative operator on M, then any other derivative operator on M is determined relative to ∇ by a smooth symmetric (in the lower indices) tensor field, $C^a{}_{bc}$, and so specifying the $C^a{}_{bc}$ field and ∇ is sufficient to uniquely determine a new derivative operator.

nian theory, it is always possible to move back to the standard theory, because Prop. A.3 guarantees that the curvature conditions (CC1)-(CC3) are satisfied.

We can now ask how either of these classical theories relate to GR. The answer is that geometrized Newtonian theory arises as a limiting case of GR, for a properly constructed limit. (For full details of this limiting procedure, see Malament (1986a, Sec. 5).) Intuitively, we will begin with a relativistic spacetime, and then allow the lightcone structure at every point to open in such a way that, in the limit, the lightcones at every point become degenerate. Since the lightcone structure in a sense determines the speed of light, allowing the lightcones to widen in this fashion captures a sense in which one might allow the speed of light to go to infinity.

To motivate what follows, it is useful to see how the limit works in detail in so-called Minkowski spacetime, which is a relativistic spacetime (M, g_{ab}) in which (a) M is the manifold $\mathbb{R}^4$, (b) g_{ab} is flat, and (c) the manifold and the derivative operator associated with g_{ab} together are geodesically complete. In this case, we can write the metric at any point as a matrix in terms of standard coordinates (t, x, y, z) and a constant c, the speed of light, as $g_{ab}(c) = \text{Diag}(1, -1/c^2, -1/c^2, -1/c^2)$. Because of the special properties of Minkowski space, it makes sense to speak of the lightcone widening uniformly around the fixed $t-$axis at all points. The metric has a well defined limit as $c \to \infty$, which can be expressed at any point p as, $\lim_{c\to\infty} g_{ab}(c) = \text{Diag}(1, 0, 0, 0) = t_{ab}$, where t_{ab} is a suggestively named degenerate metric on M with (generalized) signature $(1, 0, 0, 0)$. The inverse metric $g^{ab}(c)$ does not itself have a well-defined limit, but if we rescale it as $g^{ab}(c)/c^2$, it does. In this case, we find, $\lim_{c\to\infty} g^{ab}(c)/c^2 = \text{Diag}(0, -1, -1, -1) = -h^{ab}$, where now h^{ab} is a degenerate metric on M with signature $(0, 1, 1, 1)$. So in Minkowski space, we can recover the metrical structure of classical spacetime simply by allowing the speed of light to diverge.

In a general spacetime, however, we cannot assume that the metric will behave so nicely—for instance, if space is curved, we do not even know if there is a global coordinate system in which we can write the metric at an arbitrary point. So we proceed more carefully. Consider a manifold M admitting a one-parameter family of nondegenerate Lorentz metrics $g_{ab}(\lambda)$ (where λ ranges over some interval $(0, k) \subseteq \mathbb{R}$) that satisfy two conditions:

(Lim1) $\lim_{\lambda \to 0} g_{ab}(\lambda) = t_a t_b$ for some non-vanishing closed field t_a;[37]

(Lim2) $\lim_{\lambda \to 0} \lambda g^{ab}(\lambda) = -h^{ab}$ for some field h^{ab} of signature $(0, 1, 1, 1)$.

For any $\lambda \in (0, k)$, we can associate with $g_{ab}(\lambda)$ the unique covariant derivative operator compatible with $g_{ab}(\lambda)$, $\overset{\lambda}{\nabla}$, as well as the Ricci curvature tensor associated with $\overset{\lambda}{\nabla}$, $\overset{\lambda}{R}_{ab}$. Thus the one-parameter family of metrics generates a one-parameter family of compatible derivative operators and curvature tensors. Suppose further that, for any $\lambda \in (0, k)$, we can define a smooth symmetric field $T^{ab}(\lambda)$ that together with $g_{ab}(\lambda)$ and its associated Ricci tensor satisfy:

(Lim3) For all $\lambda \in (0, k)$, $\overset{\lambda}{R}_{ab} = 8\pi \left(T_{ab}(\lambda) - \frac{1}{2} g_{ab}(\lambda) T(\lambda) \right)$, where $T(\lambda) = T_{ab}(\lambda) g^{ab}(\lambda)$; and

(Lim4) $\lim_{\lambda \to 0} T^{ab}(\lambda) = T^{ab}$ for some smooth symmetric field T^{ab} on M.

When these conditions hold, it is possible to show that in the limit as $\lambda \to 0$, the family of relativistic spacetimes $(M, g_{ab}(\lambda))$ converges to a classical spacetime on which T^{ab} and R_{ab} satisfy the geometrized Poisson equation, Eq. (A.2). This result can be formulated as follows:

Proposition A.5 (Classical Limit of GR.) (Adapted from Malament, 1986a, Props. on Limits 1 & 2) *Fix a smooth, connected, four-dimensional manifold M and assume λ is a real-valued variable taking all values on an interval $(0, k)$. Suppose that for each λ on an interval $(0, k)$, there exist smooth symmetric fields $g_{ab}(\lambda)$ and $T_{ab}(\lambda)$ on M such that $(M, g_{ab}(\lambda))$ is a relativistic spacetime and for each λ, $g_{ab}(\lambda)$ and $T_{ab}(\lambda)$ collectively satisfy conditions (Lim1)-(Lim4). Then there exists a derivative operator ∇_a on M such that $\lim_{\lambda \to 0} \overset{\lambda}{\nabla}_a = \nabla_a$,*[38] *and for which $(M, t_a, h^{ab}, \nabla_a)$ is a classical spacetime satisfying $R^{a}{}_{b}{}^{c}{}_{d} = R^{c}{}_{d}{}^{a}{}_{b}$. Moreover, there exists a smooth field ρ on M such that $\lim_{\lambda \to 0} T_{ab}(\lambda) = \rho t_a t_b$, which satisfies $R_{ab} = 4\pi \rho t_a t_b$.*

[37] If t_a is a non-vanishing closed field, the product $t_{ab} = t_a t_b$ automatically has signature $(1, 0, 0, 0)$.

[38] What does it mean for a sequence of derivative operators to converge? Suppose that $\tilde{\nabla}_a$ is a fixed auxiliary derivative operator on M. Then for each $\overset{\lambda}{\nabla}_a$, there is a smooth symmetric field $C^{a}{}_{bc}(\lambda)$ such that $\overset{\lambda}{\nabla}_a = (\tilde{\nabla}_a, C^{a}{}_{bc}(\lambda))$. Now suppose that there is another derivative operator on M, $\nabla_a = (\tilde{\nabla}_a, C^{a}{}_{bc})$. We can say that $\lim_{\lambda \to 0} \overset{\lambda}{\nabla}_a = \nabla_a$ if $\lim_{\lambda \to 0} C^{a}{}_{bc}(\lambda) = C^{a}{}_{bc}$.

Prop. A.5 gives the precise sense in which geometrized Newtonian gravitation is a limiting case of GR.

A.3 Gravitational Mass in Newtonian Theory

We have now done sufficient groundwork to offer a technically precise formulation of the explanation given in the body of the paper. The argument was that by beginning with GR and then moving in the limit to standard Newtonian theory, one finds that a massive point particle's coupling to the gravitational field is given by its inertial mass. This limit proceeds in two steps. First, using Prop. A.5, one shows that geometrized Newtonian gravitation is a limiting case of GR. Prop. A.4, meanwhile, shows that when three curvature conditions, (CC1)-(CC3), are satisfied, we can recover (covariant) standard Newtonian theory from geometrized Newtonian theory. It is in the course of executing this two step process that one is forced to associate gravitational and inertial mass.

There is an important subtlety here. To connect Props. A.4 and A.5 and show that we can recover standard Newtonian theory as a limit from GR, we need to show that the classical spacetime we reach in the limit from GR in fact meets the three curvature conditions necessary to recover the standard theory. Prop. A.5 gives that two of the curvature conditions, (CC1) and (CC2), are satisfied automatically. But what about (CC3), $R^{ab}{}_{cd} = \mathbf{0}$? In general, $R^{ab}{}_{cd}$ need *not* vanish in a classical spacetime reached in the limit from GR. It turns out that there is a more general recovery theorem, due to Hans-Peter Künzle (1976) and Jürgen Ehlers (1981), that holds when $R^{ab}{}_{cd} \neq \mathbf{0}$. But the theory that you recover in this case is not standard Newtonian gravitation—it is a non-geometrized generalization of standard Newtonian gravitational theory in which the gravitational potential field is replaced by a vector field and there is an additional contribution to the force law for a particle arising from a kind of universal rotation. The third curvature condition is sufficient to guarantee that this rotational contribution vanishes and that the gravitational vector field can be written as the covariant derivative of a scalar potential.

There are several circumstances under which one can guarantee that the condition $R^{ab}{}_{cd}$ is satisfied.[39] But for present purposes, establishing when the condition holds is unnecessary. The important point is that condition (CC3) is both a necessary and sufficient condition for recovery of standard Newtonian physics via Prop. A.4. This is not a problem for the explanation, *per se*, since insofar as we can recover standard Newtonian theory at all, we can do so only from the class of families of relativistic spacetimes that converge to classical spacetimes satisfying (CC3). In other words, in order for the explanation I propose to succeed, one needs to assume a special curvature condition that is not otherwise guaranteed to hold in the limit from GR—but this curvature condition is just what one need to assume in order to recover standard Newtonian theory in the first place.

I can now state the precise claim, the proof of which amount to the formal explanation.

Proposition A.6 *Let* $(M, t_a, h^{ab}, \overset{f}{\nabla})$ *be a flat classical spacetime and let* φ *be a gravitational field defined on that spacetime as in standard Newtonian gravitation. Suppose further that* $(M, t_a, h^{ab}, \overset{f}{\nabla})$ *and* φ *arise via the two-step limiting process just described (which is only possible if the intermediate curved classical spacetime satisfies* (CC3)*). Consider a massive point particle with inertial mass* $m_{\mathcal{I}}$ *traversing a timelike curve in* M*,* γ*, with tangent vector field* ξ^a*, under the influence of only gravitational force. Then the gravitational force experienced by the massive point particle is*

$$F^a_{\mathcal{G}} = m_{\mathcal{I}} \xi^n \overset{f}{\nabla}_n \xi^a = -m_{\mathcal{I}} \overset{f}{\nabla}{}^a \varphi. \tag{A.3}$$

In other words, the particle's gravitational mass is equal to its inertial mass.

Proof. By assumption, $(M, t_a, h^{ab}, \overset{f}{\nabla})$ and φ arise via the two-step limiting process described above. Thus there exists a (curved) classical spacetime $(M, t_a, h^{ab}, \overset{g}{\nabla})$ satisfying (CC3) from which $(M, t_a, h^{ab}, \overset{f}{\nabla})$ can be recovered. Since the particle experiences no non-gravitational force, we know from the geodesic principle of geometrized Newtonian gravitation that γ must be a geodesic relative to $\overset{g}{\nabla}$. Meanwhile, by Prop. A.4, we know that if γ is a geodesic relative to $\overset{g}{\nabla}$, then $\xi^n \overset{f}{\nabla}_n \xi^a = -\overset{f}{\nabla}{}^a \varphi$. Thus we have the acceleration of the particle's worldline, which

[39] For instance, $R^{ab}{}_{cd}$ is automatically satisfied in a classical spacetime that is, in a certain precise sense, "asymptotically flat" (See Malament, 2010, Sec. 4.5 for details).

we can plug into $F^a = m_{\mathcal{I}} \xi^n \overset{f}{\nabla}_n \xi^a$ to find the gravitational force on the particle. We see that $F^a_{\mathcal{G}} = m_{\mathcal{I}} \xi^n \overset{f}{\nabla}_n \xi^a = -m_{\mathcal{I}} \overset{f}{\nabla}{}^a \varphi$, as required. It follows that the particle's coupling to the gravitational field is given by its inertial mass. □

References

Batterman, Robert (2002). *The Devil in the Details*. New York: Oxford University Press.

Cartan, Elie (1923). "Sur les variétés à connexion affine, et la théorie de la relativité généralisée (première partie)." *Annales scientifiques de l'École Normale Supérieure*, *40*, 325–412.

Cartan, Elie (1924). "Sur les variétés à connexion affine, et la théorie de la relativité généralisée (première partie) (suite)." *Annales scientifiques de l'École Normale Supérieure*, *41*, 1–25.

Cartwright, Nancy (1983). *How the Laws of Physics Lie*. New York: Oxford University Press.

Ehlers, Jürgen (1981). "Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie." *Grundlagen Probleme der Modernen Physik*, Ed. Jürgen Nitsch, Joachim Pfarr, and Ernst-Walther Stachow. Zurich: Bibliographisches Institut.

Einstein, Albert (1920). *Relativity: The Special and General Theory*. Ed. Robert W. Lawson. New York: Henry Holt and Co.

Fisher, Ronald Aylmer (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press.

Geroch, Robert (1978). *General Relativity from A to B*. Chicago: University of Chicago Press.

Godfrey-Smith, Peter (2003). *Theory and Reality*. Chicago: University of Chicago Press.

Hempel, Carl and Paul Oppenheim (1948). "Studies in the Logic of Explanation." *Philosophy of Science*, *15*, 135–175.

Kitcher, Philip (1989). "Explanatory Unification and the Causal Structure of the World." *Scientific Explanation*, Ed. Philip Kitcher and Wesley Salmon. Minneapolis: University of Minnesota Press.

Kuhn, Thomas S. (1977). "Concepts of Cause in the Development of Physics." *The Essential Tension*. Chicago: The University of Chicago Press. 21–30.

Künzle, H. P. (1976). "Covariant Newtonian Limit of Lorentz Space-Times." *General Relativity and Gravitation*, *7*(5), 445–457.

Malament, David (1986a). "Newtonian Gravity, Limits, and the Geometry of Space." *From Quarks to Quasars*, Ed. Robert Colodny. Pittsburgh: University of Pittsburgh Press.

Malament, David (1986b). "Gravity and Spatial Geometry." *Logic, Methodology and Philosophy of Science*, Ed. Ruth Barcan Marcus, George Dorn, and Paul Weingartner. Volume VII. New York: Elsevier Science Publishers. 405–411.

Malament, David B. (2010). "Relativistic and Newtonian Spacetime Structure." Most recent version available at http://www.lps.uci.edu/home/fac-staff/faculty/malament/FndsofGR/.

Nagel, Ernest (1998). "Issues in the Logic of Reductive Explanations." *Philosophy of Science*, Ed. Martin Curd and J. A. Cover. New York: W. W. Norton.

Nickles, Thomas (1998). "Two Concepts of Intertheoretic Reduction." *Philosophy of Science*, Ed. Martin Curd and J. A. Cover. New York: W. W. Norton.

Odom, B., D. Hanneke, B. D'Urso, and G. Gabrielse (2006). "New Measurement of the Electron Magnetic Moment Using a One-Electron Quantum Cyclotron." *Physical Review Letters*, *97*(3), 030801.

Sachs, Mendel (1976). "On the Logical Status of Equivalence Principles in General Relativity Theory." *British Journal for the Philosophy of Science*, *27*(3), 225–229.

Salmon, Wesley (1984). *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.

Salmon, Wesley (1989). "Four Decades of Scientific Explanation." *Scientific Explanation*, Ed. Philip Kitcher and Wesley Salmon. Minneapolis: University of Minnesota Press.

Skyrms, Brian (1996). *Evolution of the Social Contract.* New York: Cambridge University Press.

Strevens, Michael (2004). "The Causal and Unification Approaches to Explanation Unified–Causally." *Nous*, *38*(1), 154–176.

Strevens, Michael (2008). *Depth: An Account of Scientific Explanation.* Cambridge, MA: Harvard University Press.

Trautman, Andrzej (1965). "Foundations and Current Problem of General Relativity." *Lectures on General Relativity*, Ed. Stanley Deser and Kenneth W. Ford. Englewood Cliffs, NJ: Prentice-Hall.

Weatherall, James Owen (2005). *Effective Field Theory and the Pragmatics of Explanation.* Undergraduate thesis. Cambridge, MA: Harvard College.

Weinberg, Steven (1972). *Gravitation and Cosmology.* New York: Wiley.

Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation.* Oxford: Oxford University Press.