Studies in History and Philosophy of Modern Physics 46 (2014) 142–153

Contents lists available at ScienceDirect

Studies in History and Philosophy of Modern Physics

journal homepage: www.elsevier.com/locate/shpsb

Foundations of quantum gravity: The role of principles grounded in empirical reality

Marc Holman

Institute for History and Foundations of Science, Utrecht University, P.O. Box 80.010, 3508 TA Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 2 February 2013
Received in revised form
22 August 2013
Accepted 27 October 2013
Available online 4 December 2013

Keywords:
Quantum gravity
Thermodynamic arrow of time
CMB uniformity
Quantum nonlocality
Irreducible randomness

ABSTRACT

When attempting to assess the strengths and weaknesses of various principles in their potential role of guiding the formulation of a theory of quantum gravity, it is crucial to distinguish between principles which are strongly supported by *empirical data* – either directly or indirectly – and principles which instead (merely) rely heavily on theoretical arguments for their justification. Principles in the latter category are not necessarily invalid, but their a priori foundational significance should be regarded with due caution. These remarks are illustrated in terms of the current standard models of cosmology and particle physics, as well as their respective underlying theories, i.e., essentially general relativity and quantum (field) theory. For instance, it is clear that both standard models are severely constrained by symmetry principles: an effective homogeneity and isotropy of the known universe on the largest scales in the case of cosmology and an underlying exact gauge symmetry of nuclear and electromagnetic interactions in the case of particle physics. However, in sharp contrast to the cosmological situation, where the relevant symmetry structure is more or less established directly on observational grounds, all known, nontrivial arguments for the "gauge principle" are purely theoretical (and far less conclusive than usually advocated). Similar remarks apply to the larger theoretical structures represented by general relativity and quantum (field) theory, where – actual or potential – empirical principles, such as the (Einstein) equivalence principle or EPR-type nonlocality, should be clearly differentiated from theoretical ones, such as general covariance or renormalizability. It is argued that if history is to be of any guidance, the best chance to obtain the key structural features of a putative quantum gravity theory is by deducing them, in some form, from the appropriate empirical principles (analogous to the manner in which, say, the idea that gravitation is a curved spacetime phenomenon is arguably implied by the equivalence principle). Theoretical principles may still be useful however in formulating a concrete theory (analogous to the manner in which, say, a suitable form of general covariance can still act as a sieve for separating theories of gravity from one another). It is subsequently argued that the appropriate empirical principles for deducing the key structural features of quantum gravity should at least include (i) quantum nonlocality, (ii) irreducible indeterminacy (or, essentially equivalently, given (i), relativistic causality), (iii) the thermodynamic arrow of time, (iv) homogeneity and isotropy of the observable universe on the largest scales. In each case, it is explained – when appropriate – how the principle in question could be implemented mathematically in a theory of quantum gravity, why it is considered to be of fundamental significance and also why contemporary accounts of it are insufficient. For instance, the high degree of uniformity observed in the Cosmic Microwave Background is usually regarded as theoretically problematic because of the existence of particle horizons, whereas the currently popular attempts to resolve this situation in terms of inflationary models are, for a number of reasons, less than satisfactory. However, rather than trying to account for the required empirical features *dynamically*, an arguably much more fruitful approach consists in attempting to account for these features directly, in the form of a lawlike initial condition within a theory of quantum gravity.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Modern Physics*

> It may be that a real synthesis of quantum and relativity theories requires not just technical developments but radical conceptual renewal.
>
> J. S. Bell (1987)

E-mail address: m.holman@uu.nl

1355-2198/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.shpsb.2013.10.005

1. Introduction

The problem of how to unify quantum theory and general relativity into a consistent theoretical framework has baffled physicists for more than eight decades, but has become particularly pressing in recent years. Each of the two theories in question is a fundamental theory of physics in its own right – each theory is, in Einstein's terminology, a *theory of principle*; see below – and is corroborated by a very impressive body of experimental evidence. But physical reality constitutes an undivided whole and it obviously makes no sense to have *two* distinct fundamental theories for a single reality. That physicists have been able to get away with this state of affairs for so long is simply due to the fact that the intersection of the domains of applicability of the two theories is in some sense small and at any rate is usually thought to lie far beyond anything technologically feasible, at present or in the nearby future. Indeed, according to the received view, because of gravity's intrinsic weakness as compared to the nuclear and electromagnetic interactions, general relativistic considerations play no role in quantum theory until particle scattering energies reach Planckian dimensions and, conversely, because quantum theory is at first glance effectively restricted to the (sub-)atomic regime, it can seemingly play no role in any nontrivial general relativistic effects, as these typically involve distance scales many orders of magnitude beyond the atomic scale. But, apart from the questionable nature of these arguments, the lack of a unifying theoretical framework for the entire realm of natural phenomena is arguably a serious problem from the perspective of the foundations of physical theory.[1]

As said, both quantum theory and general relativity are theories of principle. This means that they are characterized by defining principles which are supposed to be valid universally, and which should apply, without restriction, to *all* physical phenomena[2] (such theories are to be distinguished from *constructive theories*, which are theories which attempt to explain a limited group of natural phenomena by means of some model or set of equations, but always so within the context of a specific theory of principle[3]). But it is not difficult to see that the two sets of principles associated with these theories are not fully inter-compatible. For instance, according to general relativity, spacetime is a dynamical entity, whereas the standard unitary dynamics of states and observables in quantum theory is conditional upon a decomposition of this entity into "space" and "time", which are moreover taken as non-dynamical. Similarly, according to quantum theory it makes sense to talk of a point mass being in a quantum superposition centred about two different spatial locations, whereas according to general relativity this makes no sense, since such a superposition would (in principle) entail a corresponding superposition of two spacetime geometries and there is no natural way to mathematically implement such a superposition (i.e., to identify the points of two non-identical spacetimes with each other).[4] Logic thus dictates that at least one set of principles cannot be the complete story and it is a matter of simple historical fact that the "mainstream", often implicit, viewpoint in theoretical physics with respect to this issue has always been that the principles of general relativity are somehow to be subjugated to those of quantum theory (in a sense, this view is of course also encapsulated in the very phrase "quantization of gravity").[5] As will be seen in the next two sections however, there are ample grounds to question the validity of this viewpoint (at the very least, it is undeniable that it has so far failed to lead to a consistent resolution of the unification problem). If the mainstream perspective is abandoned, the incompatibility of the two sets of principles raises some profound conceptual and methodological dilemmas. For instance, on the one hand, it might seem that the most important lesson that can be drawn from the historical development of general relativity is that the strong insistence on a principle of locality, in spite of a number of serious earlier obstacles, ultimately payed off.[6] On such an account, it would thus not seem unreasonable to expect that quantum theory will eventually also be seen to be founded upon a more local, "classical" basis. On the other hand, it might be asked on what precise grounds a physical theory is required to be local. One seemingly obvious answer is that *if* physical reality is structured that way, human intuition would presumably manage to latch onto such a structure somehow, and so the *fact* that locality is a very intuitive notion would then (according to this line of reasoning) imply the validity of some form of locality as a fundamental principle of nature. Yet, such a position appears dangerously close to the Kantian notion of the synthetic a priori, whereas the intuitive appeal of a locality principle seems equally well explainable on the basis that the structure of "macroscopic" physical reality happens to be local (to a large extent) and that human intuition is able to latch onto that structure through sense experience. But there would then be no guarantee that such intuition could be validly extended into the "microscopic" domain, not directly accessible to the senses. In fact, as recalled in more detail later, it is well established that an underlying local description in the case of quantum theory is *impossible* for empirical reasons (at least if such a description is interpreted in terms of real spacetime phenomena).

On this basis, it is then perhaps tempting to conclude that the entire doctrine of locality – although extremely useful in arriving at the successful theories of electromagnetism and relativity – is ultimately untenable and that, because of this, the principles of quantum theory carry more weight than those of general relativity. However, even if the locality of general relativity and classical electrodynamics ultimately turns out to be "merely" an emergent quality – as the present evidence strongly suggests – there is an important reason for attaching at least as much weight to the line of research that culminated in Einstein's general relativity and, consequently, also to the principles of general relativity themselves. This is the circumstance that the development of the theories of relativity – as well as that of classical electromagnetism – were strongly driven by the requirement that physical theories be *intelligible*, whereas, by contrast, the development of quantum theory was strongly dominated by an attitude which was essentially instrumentalist.

These remarks are further illustrated in Sections 3.1 and 3.2 respectively. As to the unification problem, the upshot of the

[1] That is, the lack of such a framework *is* a serious problem assuming an appropriate form of scientific realism (although that is of course the household doctrine of theoretical physicists working on these matters – at least when explaining their activities to the lay public or when writing up research grants!). It is not necessary however to assume realism to point out the inadequacies of existing theories (see Section 2 for further discussion).

[2] See Stachel (2002).

[3] For instance, the $SU(3) \times SU(2) \times U(1)$ standard model of particle physics and quantum field theory models more generally are constructive theories within a quantum theoretical context (see Section 3.2), while the so-called ΛCDM "concordance model" is a constructive theory of cosmology within a general relativistic context (see Section 3.1). It is sometimes argued that general relativity is (at least partly) a constructive theory in view of the fact that it can be regarded as a classical field theory, but that position will not be adopted here. General relativity describes the properties of space and time in relation to those of ponderable matter and fields and as such it forms the general setting for particular constructive theories of ponderable matter and fields.

[4] Cf. Penrose (1996, 2009). See also Károlyházy, Frenkel, & Lukács (1986) and Diósi (1989).

[5] It is somewhat of an irony that as early as 1916, Einstein himself already argued for a modification of "the new theory of gravitation" *because* of quantum theory (of which the general principles had yet to be formulated of course). See Einstein (1916). For an explicit quantum-universality claim in this regard, see e.g. Kiefer (2010).

[6] The notion of locality will be further explicated and differentiated in Section 4.1 but the intuitive content of this notion should suffice for present purposes.