arXiv:0909.1408v1 [quant-ph] 8 Sep 2009

THE BACKGROUND AS A QUANTUM OBSERVABLE: EINSTEIN'S HOLE ARGUMENT IN A QUASICLASSICAL CONTEXT

I. SCHMELZER

Abstract. We consider a thought experiment measuring the decoherence for quasiclassical superpositions of gravitational field. The hole argument allows to prove that a covariant (background-free) theory is completely unable to define the outcome of this experiment and is therefore not viable. Instead, the results of experiments of this type allow to reconstruct a common background shared by all superposed gravitational fields.

1. Introduction

Einstein's hole argument uses the covariance of the Einstein equations to construct different solutions $g_{\mu\nu}(x)$, $g'_{\mu\nu}(x)$ which have the same initial values and boundary conditions. At a first look, this seems to imply that general relativity is not a well-defined theory, and that diffeomorphism invariance is not viable as a principle for physical theories. But it appears that in the classical case these different solutions cannot be distinguished by observation, so all physical predictions are well-defined: The equations of GR do not define a unique solution $g_{\mu\nu}(x)$ in terms of some spacetime background variable x, but only an equivalence class of solutions.

The aim of this paper is to show that the situation is different in quasiclassical gravity, where we can consider superpositions of classical gravitational fields. Replacing one of the fields $g_{\mu\nu}(x)$ in such a superpositional state by a $g'_{\mu\nu}(x)$ constructed in the hole argument changes observable predictions for a simple thought experiment.

The experiment measures decoherence caused by gravitational interaction. We prepare a superpositional state of a (sufficiently heavy) source particle. Some gravitational interaction with a lightweight particle possibly causes decoherence. We measure the remaining interference pattern for the source particle, ignoring the state of the lightweight particle.

The remaining degree of interference, together with a possible phase shift, defines a complex number. In the Newtonian quantum gravity (multi-particle Schrödinger theory with Newtonian interaction potential) we obtain a simple approximation for this observable. If the source particle is located in a δ-like way near $x_{l/r}$ and is heavy enough, a good approximation may be obtained using the wave functions $\psi_{l/r}(\mathbf{x}, t)$ of the lightweight particle in the two external potentials $V_{l/r}(x) = |x - x_{l/r}|^{-1}$. Then the complex number we observe in our experiment can be approximated by

Key words and phrases. quantum gravity, superposition.

Berlin, Germany.

the scalar product

$$\langle\psi_l|\psi_r\rangle = \int \bar{\psi}_l(\mathbf{x})\psi_r(\mathbf{x})d^3\mathbf{x}. \tag{1.1}$$

It is natural to assume that this formula remains approximately valid beyond the Newtonian limit, in the quasiclassical approximation where the wave functions $\psi_{l/r}(\mathbf{x},t)$ of the lightweight particle have to be computed as one-particle semiclassical approximations on the background of gravitational fields $g^{l/r}_{\mu\nu}(\mathbf{x},t)$ defined by a classical metric theory of gravity.

But with this assumption about the physical meaning of $\langle\psi_l|\psi_r\rangle$, the underlying classical theory of gravity needs a background. If it would be background-free, we could use the hole argument to replace one of the gravitational fields, say $g^l_{\mu\nu}(\mathbf{x},t)$, by another field ${g'}^l_{\mu\nu}(\mathbf{x},t)$, but leave the other, $g^r_{\mu\nu}(\mathbf{x},t)$, unchanged. If we replace $g^l_{\mu\nu}(\mathbf{x},t)$ by ${g'}^l_{\mu\nu}(\mathbf{x},t)$, we have to replace in a corresponding way $\psi_l(\mathbf{x},t)$ by some different $\psi'_l(\mathbf{x},t)$, leaving $\psi_r(\mathbf{x},t)$ (together with $g^r_{\mu\nu}(\mathbf{x},t)$) unchanged. But in this case their scalar product $\langle\psi_l|\psi_r\rangle$ changes in an unpredictable way, thus, it is not defined by the theory. This is incompatible with the status of $\langle\psi_l|\psi_r\rangle$ as the approximation of an observable. Thus, a background-free theory cannot predict the result of our experiment and is therefore not viable.

We formalize this as an impossibility theorem for background-free theories. The assumptions we have to make are rather weak – we do not need assumptions about strong fields, the quasiclassical approximation is already sufficient. We also do not need the full power of covariance – all we need is covariance for spatial diffeomorphisms which preserve some given foliation of spacetime. Thus, not only fully background-free theories are impossible: Even the idea to solve the "problem of time" with a preferred foliation, but to preserve background-freedom for spatial diffeomorphisms, fails.

But the message is not only negative: Quantum theory allows to measure something qualitatively new, unobservable in classical theory, and we can identify this new quantum observable with the background. Given all the information which is accessible by experiments of this type one can in principle recover the background. This invalidates relational and positivistic argumentation against the background: It is at least in principle observable, and defines the relation between different metrics in a superpositional steate.

The background can be also supported by beautiful mathematics: To introduce it into GR, we have a candidate of unique beauty – harmonic coordinates. Then the nice formula (1.1) for the result of our experiment can be applied for quasiclassical superpositions of relativistic gravitational fields too.

2. A thought experiment with a superposition of gravitational fields

To create a superpositional state of the gravitational field is, at least in principle, quite easy – all one has to do is to create a usual superpositional state of some sufficiently heavy source particle. Thus, at least in principle, we do not need more than a usual double slit arrangement, with a left and a right slit at $x_{l/r}$, which creates a superpositional state for a sufficiently heavy source particle.

The resulting superpositional state interacts with it's environment, and this interaction leads, if it is sufficiently strong, to decoherence. What we are interested in is the decoherence caused by the gravitational interaction. So we assume that there is some lightweight particle which interacts gravitationally with the superpositional state, and we want to compute the resulting decoherence effect. In practice, this effect is far too small to be important – the other forces are so much stronger that they cause complete decoherence before the decoherence effect of the gravitational interaction becomes observable. But this is a practical problem, and nothing suggests that the theoretical principles which work for the much stronger electromagnetic interaction become invalid for gravity.

While the result itself is quite trivial, we split it into two theorems: The advantage of this splitting is that the first theorem needs almost nothing about the lightweight environment particle, but considers only the source particle. It gives us some complex number as an observable. The second theorem then allows to compute this observable in terms of a scalar product of states of the lightweight particle.

Let's denote the states where the source particle is at the left resp. right slit with $|g_{l/r}\rangle$. Then, our initial state is the superposition $\frac{1}{\sqrt{2}}(|g_l\rangle + |g_r\rangle)$.

Theorem 1 (the result of partial position measurement)**.** *Assume a pure superpositional state $\frac{1}{\sqrt{2}}(|g_l\rangle+|g_r\rangle)$ of some system interacts with some part of its environment. Assume that the interaction is (approximately) diagonal in the basic states $|g_l\rangle$, $|g_r\rangle$ and leaves them (approximately) unchanged. Then the resulting state $\hat{\rho}_\theta$ of the system has (approximately) the form*

$$\hat{\rho}_\theta = (1-|\theta|)\hat{1} + |\theta||g_\theta\rangle\langle g_\theta|; \qquad |g_\theta\rangle = \frac{1}{\sqrt{2}}(|g_l\rangle + e^{i\varphi}|g_r\rangle) \tag{2.1}$$

which is uniquely characterized uniquely by the complex number $\theta = |\theta|e^{i\varphi}$. This number θ is observable.

Proof. Once the basic states $|g_l\rangle$, $|g_r\rangle$ remain unchanged, the resulting state for their superposition has to be a state generated from these basic states, thus, a 2×2 density matrix. Then, for the superposition $\frac{1}{\sqrt{2}}(|g_l\rangle + |g_r\rangle)$ the probability of measuring $|g_l\rangle$ and $|g_r\rangle$ in this partial measurement is equal. This restricts the general form of the density matrix to a two-parameter set of the form of $\hat{\rho}_\theta$. Observability is obvious as well – a standard interference pattern will show some observable degree of remaining interference defining $|\theta|$ and for $|\theta| > 0$ an additional phase shift caused by the interaction defining $e^{i\varphi}$. □

Thus, a completely general interaction with some yet completely unspecified part of it's environment, which will be ignored, gives an observable effect characterized by some complex number θ.

Theorem 2 (the result as a scalar product)**.** *If, for the initial states $|g_{l/r}\rangle$ of theorem 1 the resulting full state (the system together with the part of the environment it has interacted with) is $|g_{l/r}\rangle|\psi_{l/r}\rangle$, where the wave functions $|\psi_{l/r}\rangle$ describe the state of the environment, then $\theta = \langle\psi_l|\psi_r\rangle$.*

Proof. Assuming that the initial state of the part of the environment is $|\psi_0\rangle$, the initial state of the whole system is $\frac{1}{\sqrt{2}}(|g_l\rangle + |g_r\rangle)\otimes|\psi_0\rangle$. It evolves into some final state $\frac{1}{\sqrt{2}}(|g_l\rangle\otimes|\psi_l\rangle + |g_r\rangle\otimes|\psi_r\rangle)$. Once we ignore the state of the environment, we have to describe the final state of the system by a density matrix. If the environment is described by a single particle, we can measure it's position $\mathbf{x}$ and then trace over $\mathbf{x}$. This leads to an expression which depends only on the scalar product $\langle\psi_l|\psi_r\rangle = \int \bar{\psi}_l(\mathbf{x})\psi_r(\mathbf{x})d^3\mathbf{x}$:

$$\begin{aligned}\hat{\rho}_\theta &= \frac{1}{2}\int d^3\mathbf{x}(|g_l\rangle \otimes \psi_l(\mathbf{x}) + |g_r\rangle) \otimes \psi_r(\mathbf{x}))(|g_l\rangle \otimes \bar{\psi}_l(\mathbf{x}) + |g_r\rangle) \otimes \bar{\psi}_r(\mathbf{x})) \\ &= (1 - |\langle\psi_l|\psi_r\rangle|)\frac{1}{2}(|g_l\rangle\langle g_l| + |g_r\rangle\langle g_r|) + |\langle\psi_l|\psi_r\rangle||g_\theta\rangle\langle g_\theta|.\end{aligned} \tag{2.2}$$

Comparison with (2.1) gives immediately $\theta = \langle\psi_l|\psi_r\rangle$. The same result will be obtained for other measurements λ of the environment: We obtain $\theta = \langle\psi_l|\psi_r\rangle$, only with $\langle\psi_l|\psi_r\rangle = \int \bar{\psi}_l(\lambda)\psi_r(\lambda)d\lambda$. □

3. The case of quasiclassical gravity

Let's specify now the details for the case of quasiclassical gravity. We use this notion to describe the approximation of full quantum gravity which allows to consider finite superpositions of semiclassical states, where a semiclassical state is defined by a classical gravitational field – a Lorentz metric $g_{\mu\nu}(x)$ – and a quantum field moving on this background.

For our purpose there is no necessity to consider questions related with different foliations of spacetime. So we consider one foliation of spacetime $x = (\mathbf{x}, t)$ to be fixed. It does not mean that this somehow restricts the applicability of our considerations to theories with a preferred foliation: It only extends the applicability to theories with such a preferred foliation too. For theories which do not have a preferred foliation, our consideration may be applied to any foliation.

Thus, we assume now that the basic states $|g_l\rangle$ and $|g_r\rangle$ are defined by solutions $g^l_{\mu\nu}(\mathbf{x}, t)$, $g^r_{\mu\nu}(\mathbf{x}, t)$ of some classical metric theory of gravity (not necessarily GR). We restrict ourself to sufficiently weak gravitational fields and a sufficiently plane foliation. In this case, we can ignore effects of particle creation and destruction by the gravitational field and consider a one-particle approximation. The details of the evolution are not relevant – all we need is that, given the solutions $g^{l/r}_{\mu\nu}(\mathbf{x}, t)$ and some initial wave function $|\psi_0\rangle$, we can compute wave functions $|\psi_{l/r}\rangle$ for all times. Note that the evolution should not even be unitary, because there will be some probability of particle creation and destruction which will be ignored in our approximation. All we would like to have is summarized in the following postulate:

Postulate 1. *For sufficiently weak, almost classical gravitational fields there exists some one-particle quasiclassical approximation, where in some fixed (but not necessarily preferred) foliation of space-time $x = (\mathbf{x}, t)$ the following properties hold:*

(1) *The quantum state $|g\rangle|\psi\rangle$ at time t_0 is (modulo the minimal uncertainties required by the uncertainty principle) completely defined by a classical gravitational field $g_{\mu\nu}(\mathbf{x}, t_0)$ and it's derivatives and a one-particle wave function $\psi(\mathbf{x}, t_0) \in \mathbb{C}$ in some Hilbert space $H_{|g\rangle}$ of wave functions.*

(2) *The Hilbert space structure on $H_{|g\rangle}$ is defined by the scalar product*

$$\langle\psi_1|\psi_2\rangle = \int \overline{\psi}_1(\mathbf{x})\psi_2(\mathbf{x})d^3\mathbf{x}.$$

(3) *For superpositional states $\frac{1}{\sqrt{2}}(|g_l\rangle + |g_r\rangle)$ of such one-particle semiclassical states, the experiment of theorems 1 and 2 defines an observable non-degenerated sesquilinear form $\theta = \langle\psi_l|\psi_r\rangle$ on $H_{|g_l\rangle} \times H_{|g_r\rangle}$*

(4) *This form is defined by*

$$\langle\psi_l|\psi_r\rangle = \int \overline{\psi}_l(\mathbf{x})\psi_r(\mathbf{x})d^3\mathbf{x}.$$

4. The classical hole argument

Let's remember Einstein's hole argument [5] and it's resolution in classical general relativity. Let $g_{\mu\nu}(x)$ be a solution of the Einstein equations of general relativity (GR). Let's consider some bounded spacetime region Σ – the "hole". It is located in the future of the initial time $x^0 = t_0$. [1] Then we consider a nontrivial smooth diffeomorphism $x' = x'(x)$, which is trivial ($x'(x) = x$) outside Σ. As a consequence of diffeomorphism invariance, the transformed metric $g'_{\mu\nu}(x')$, after formally replacing x' with x, gives a different solution $g'_{\mu\nu}(x)$ of the Einstein equations. It coinsides with $g_{\mu\nu}(x)$ outside Σ, thus, defines a different solution with the same boundary conditions and the same set of initial data at $x^0 = t_0$. Thus, at a first look, it seems that in GR the gravitational field cannot be uniquely defined by whatever set of initial values and boundary conditions, for matter fields as well as the gravitational field.

The GR solution of this problem is that the two solutions $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$, while different as functions of x, cannot be distinguished by observation. What naively looks like an observable – the value $g_{\mu\nu}(x)$ in a point x – appears to be unobservable: Observables are connected with events, which have to be identified by their relations to other events. For example, the event x_1 may be identified by the set of events at t_0 which intersect it's past light cone. But this same set at t_0 defines, on the $g'_{\mu\nu}(x)$, another event $x'_1 = x'(x_1)$.

Thus, despite the classical hole argument, in GR all real, physical observables are well-defined by the initial values and boundary conditions. But for the viability of the covariant Einstein equations it is essential that there is no observable O which allows to distinguish the two solutions $g_{\mu\nu}(x)$ and $g'_{\mu\nu}(x)$ connected via the coordinate transformation $x' = x'(x)$ in the hole Σ.

5. The impossibility theorem

In our considerations it is sufficient to consider the hole argument for purely spatial diffeomorphisms $(\mathbf{x}', t') = (\mathbf{x}'(\mathbf{x}, t), t)$ which preserve our fixed foliation of spacetime: Covariance for this subgroup of diffeomorphisms is already sufficient to prove our impossibility theorem. Of course, a fully covariant theory will be covered anyway – the theory is also covariant for all spatial diffeomorphisms. But in this way we cover also theories which are only spatially background-free, but use a preferred foliation. [2]

The definition of a spatially background-free theory of gravity is based on a straightforward notion of equivalence of semiclassical configurations:

---

[1] In Einstein's version it was located in a region without any material processes. The conclusion was, correspondingly, that the gravitational field cannot be uniquely defined by the distribution of matter. This contradicts Mach's principle, but this is something we have accepted. The gravitational field has it's own degrees of freedom, gravitational waves. Thus, this version of the hole argument has only historical interest.

[2] Such theories play an important role in the canonical quantization concept, even if they are combined with the hope that, once such a theory has been found, one can get rid of the preferred foliation in some final step.

Definition 1 (equivalence)**.** *Assume the two configurations of one-particle semi-classical gravity $g_{\mu\nu}(\mathbf{x}), \psi(\mathbf{x})$ and $g'_{\mu\nu}(\mathbf{x}), \psi'(\mathbf{x})$ can be obtained from each other using a spatial coordinate transformations $x' = x'(x) = (\mathbf{x}'(\mathbf{x}, t), t)$ via the usual hole construction. That means, the application of the coordinate transformation $x'(x)$ to the fields $g_{\mu\nu}(\mathbf{x}), \psi(\mathbf{x})$, where $g_{\mu\nu}(\mathbf{x})$ is transformed like a metric, and $\psi(\mathbf{x})$ as the square root of a density, gives $g'_{\mu\nu}(\mathbf{x}')$, $\psi'(\mathbf{x}')$, with $\mathbf{x}'$ instead of $\mathbf{x}$. Then the two configurations are said to be equivalent.*

Based on this notion of equivalence we can define now the meaning of background independence:

Definition 2 (background independence)**.** *A quantum theory of gravity is spatially background-free if equivalent semiclassical one-particle configurations $g_{\mu\nu}(\mathbf{x}), \psi(\mathbf{x})$ and $g'_{\mu\nu}(\mathbf{x}), \psi'(\mathbf{x})$ define (modulo the minimal uncertainties required by the uncertainty principle) the same quantum state $|g\rangle|\psi\rangle$.*

In particular, a background-free theory in the usual sense will be also spatially background-free.

Now we can formulate our main impossibility result:

Theorem 3. *A quantum theory of gravity which fulfills postulate 1 cannot be spatially background-free.*

Proof. Let's consider two configurations $g^l_{\mu\nu}(\mathbf{x}, t), \psi_l(\mathbf{x}, t)$ and $g^r_{\mu\nu}(\mathbf{x}, t), \psi_r(\mathbf{x}, t)$. Assume that $|\langle\psi_l|\psi_r\rangle| \approx 1$ and that the wave functions $\psi_{l/r}$ have finite support U. These assumptions are uncritical, because $|\langle\psi_l|\psi_r\rangle| = 1$ is what happens if the gravitational interaction is too weak, which happens in all our current experiment, and instead of finite support we could as well use some sufficiently certain localization.

We consider now a spatial coordinate transformation $x'(\mathbf{x}, t) = (\mathbf{x}'(\mathbf{x}, t), t)$ which is trivial for $t \le t_0$ so that at times $t > t_1$ the image of U is contained in a set U' which has no intersection with U, thus, $U' \cap U = \emptyset$. We apply now the hole construction with $x'(\mathbf{x}, t)$ to $g^l_{\mu\nu}(\mathbf{x}, t), \psi_l(\mathbf{x}, t)$. The wave function transforms like a square root of a density, thus, $\psi_l(\mathbf{x}) = n'(\mathbf{x})\psi'_l(\mathbf{x}'(\mathbf{x}, t))$ for some (irrelevant) weight function $n'(\mathbf{x})$. As a consequence, the wave function $\psi'_l(\mathbf{x})$ has their support in U'. Thus, the supports of $\psi'_l(\mathbf{x}, t)$ and $\psi_r(\mathbf{x}, t)$ do not intersect for $t > t_1$. As a consequence

$$\langle\psi'_l|\psi_r\rangle = \int \overline{\psi}'_l(\mathbf{x}, t)\psi_r(\mathbf{x}, t)d^3\mathbf{x} = 0. \tag{5.1}$$

Now, the configurations $g^l_{\mu\nu}(\mathbf{x}), \psi_l(\mathbf{x})$ and ${g'}^l_{\mu\nu}(\mathbf{x}), \psi'_l(\mathbf{x})$ are by construction equivalent. Therefore, if the theory is background-independent, they define the same quantum state $|g_l\rangle|\psi_l\rangle$. Therefore, the superpositional state $\frac{1}{\sqrt{2}}(|g_l\rangle|\psi_l\rangle + |g_r\rangle|\psi_r\rangle)$ is also the same, and, therefore, should give the same predictions for all observable effects. But the prediction for our thought experiment is not the same: For the initial configurations we have assumed $|\langle\psi_l|\psi_r\rangle| \approx 1$, thus, a full interference picture, while for the alternative configurations we have found $\langle\psi'_l|\psi_r\rangle = 0$, thus, no interference effect at all. Thus, even if $\langle\psi_l|\psi_r\rangle$ has to define the observable θ only approximately, we obtain a contradiction: $1 \not\approx 0$. □

It is worth to note that we have applied the hole construction only to one of the two metrics in the superpositional state. One could ask if it wouldn't be more appropriate to apply the same diffeomorphism to above metrics. This would correspond to the notion of c-covariance considered by Anandan [1], to be distinguished from the stronger notion of q-covariance, where different diffeomorphisms can be applied to different metrics. But, in agreement with Anandan, it is the stronger notion of q-covariance which corresponds to background freedom in the sense of GR. In fact, the notion of "the same diffeomorphism" is simply not defined for different solutions of GR. To define it in a meaningful way, one needs – a background.

6. Recovering the common background

An impossibility theorem seems to be something negative: It is impossible for quantum gravity to extend a very nice and beautiful property of general relativity – background independence – into the quantum domain. Quantum gravity would have to be less symmetric and, therefore, less beautiful than expected by those who argue for a background-free theory of quantum gravity.

But we have to disagree here. In our opinion, the result is a positive one, if we look at it from another side – from the point of view of the possibilities of observation. Theoreticians have focussed their interest mainly on the loss of possibilities caused by quantum theory – the uncertainty relations. But there is also another side: Quantum theory gives also new possibilities for measurements. Almost all of the impressive progress of the accuracy of our measurement devices is based on the use of various quantum effects. But the improvement of accuracy using interference effects is not all what quantum measurements have to offer. It gives also qualitatively new observables – additional information which cannot be gained even in principle using only classical measurements.

A similar insight has been gained already by the Bohm-Aharonov effect: In a region (outside a toroidal solenoid) where the EM field strength $F_{\mu\nu}$ was identically zero it was nonetheless possible to measure a nontrivial EM effect – a phase shift in an interference picture. This phase shift may be computed by an integral over a closed path of the gauge potential A_μ. This is a strong hint that the true degrees of freedom of a gauge field are the gauge potentials A_μ and not the field strength $F_{\mu\nu}$. Unfortunately, this argument is not decisive – one cannot recover the gauge potential A_μ even if all the integrals $\oint A_\mu(x)dx^\mu$ over closed paths are known.

This leads to the question about the relation between the new quantum observables θ and the background. We have found that we need the background to compute the observable θ. But is it possible to recover the background given all possible observables of type θ? Or is there a similar difference as between the classical gauge potential A_μ and the observable integrals over closed paths $\oint A_\mu(x)dx^\mu$? This is answered by the following

Theorem 4. *Assume in a quantum theory of gravity postulate 1 points 1-3 holds and there is no superselection rule which prevents superpositions of different semi-classical states. Then we can define a common background shared by all gravitational fields in this theory.*

Proof. For an arbitrary given field $g_{\mu\nu}(\mathbf{x},t)$ we consider it's superposition with some fixed vacuum state $\eta_{\mu\nu}$. The observable θ defines, according to postulate 1 item 3, a non-degenerated sesquilinear form on $H_{|g\rangle} \times H_{|\eta\rangle}$. This form can be used to define an isomorphism $H_{|g\rangle} \cong H_{|\eta\rangle}$: Every element of $H_{|g\rangle}$ defines a linear functional on $H_{|\eta\rangle}$, thus, an element on this space. Given this isomorphism, the position measurement on the vacuum state, defined by some projector-valued measure on $H_{|\eta\rangle}$, defines uniquely a projector-valued measure on $H_{|g\rangle}$. This measurement on $H_{|g\rangle}$ defines the background. □

Here, we have not used the explicit form of the scalar product as postulated in item 4 of postulate 1. The abstract form defined by item 3 was sufficient. But the explicit product form guarantees better properties:

Theorem 5. *Assume in a quantum theory of gravity postulate 1 holds. Then the measurement of the background position commutes with the measurement of position on the metric* $g_{\mu\nu}(\mathbf{x}, t)$.

Proof. Indeed, the coordinate $\mathbf{x}$ used in item 4 of postulate 1 defines position on above metrics involved in the superpositional state. Thus, it defines a position measurement on the metric $g_{\mu\nu}(\mathbf{x}, t)$ as well as on the background $\eta_{\mu\nu}$. □

The construction of the background in the proof of theorem 4 contains a certain freedom of choice – the choice of the reference state $\eta_{\mu\nu}$ we have named "vacuum". But a different choice of the vacuum does not change much – it simply corresponds to the freedom of choice of the coordinates of the common background. The identification of "the same point" in different gravitational fields will remain unchanged.

Of course, a single experiment gives always only one value θ, while for the recovery of the background in our construction we need all the values $\langle\psi_l|\psi_r\rangle$ for arbwe assume thatitrary wave functions $|\psi_l\rangle$, $|\psi_r\rangle$. But this situation is typical in physics: To specify whatever field configuration completely by measurement, we would need an infinity of measurements, moreover of infinite accuracy, which is impossible. The positivistic idea that all we use in physics has to be based on results of measurements is simply nonsensical. The purpose of the theorem is a different one: It shows that all the information which is, in each peace of it, available in principle by measurement is sufficient to recover the background. This makes our quantum gravity observables different from the observables of the Bohm-Aharonov effect, which do not allow to reconstruct the gauge potentials themself, but only gauge-invariant integrals of the gauge potential over a closed path.

Note that to prove that the background is unobservable classically, we need the covariance of classical general relativity. In this sense, the insight of classical GR is far away from being lost – it is what allows us to prove that quantum measurements are really more powerful than classical measurements.

7. Discussion

The background-freedom of GR is a really beautiful property, and the idea to extend this property into the quantum domain has its beauty and should not be given up too early. But those trying to find such a theory already know that this job is almost impossible, and that in particular the problem of defining the observables is very critical.[3] Despite this, they seem quite optimistic. This suggests that they will hardly give up reading about our impossibility theorem.

[3] In particular Smolin [11] notes that "...one cannot define the physical observables of the theory without solving the dynamics", and Thiemann [14] writes "...one must find a complete set of Dirac observables (operators that leave the space of solutions invariant) which is an impossible task to achieve even in classical general relativity."

Are there any objections against our impossibility theorem which may justify some hope? There is of course the general hope that impossibility theorems in physics often do not hold their promises. There are lot's of examples of impossibility theorems which have later appeared to be based on unreasonable assumptions.[4]

This is not much, only a quite abstract hope. But it is sufficient to motivate us to consider possible objections in a very careful way.

7.1. **Isn't it problematic that only approximations are used?** Our postulate 1 uses only a quite weak approximation: Finite superpositions of semiclassical states, moreover, only the one-particle approximation of semiclassical theory.

One can argue that using such a weak approximation we cannot expect any nontrivial insight into full quantum gravity.

First, to obtain such insights there are certainly better approximations. In particular we can do multi-particle theory in the semiclassical domain, or, even better, field theory, taking into account not only particle creation and destruction by the gravitational field, but also the problems with the definition of the vacuum state.

But even this clearly would not be sufficient: Superpositions of semiclassical states are something beyond semiclassical theory, thus, simply cannot be handled appropriately today. To obtain some nontrivial insight into full quantum gravity, we need better theories.

But this argumentation misses the logic of impossibility theorems: An impossibility theorem which bases it's argumentation on an approximation is even stronger than one which makes some precise assumptions: To circumvent such an impossibility theorem, the theory would have to differ from the ones covered by the theorem already in the approximation. And, given this logic, the weaker the approximation, the better: If, in particular, already the Newtonian limit is sufficient to prove the impossibility, a viable theory would have to differ from those covered by the theorem already in the Newtonian limit. If the one-particle approximation is sufficient, the theory would have to differ already in the one-particle approximation. In all these cases, the theory which circumvents the impossibility theorem has to differ in a more rigorous way from our expectations, and the impossibility theorem becomes therefore stronger.

7.2. **What about strong gravitational fields?** One may ask what our considerations can tell us about strong gravitational fields.

The answer is simple: Nothing. Strong gravitational fields are irrelevant for our impossibility argument. A viable theory of gravity should be able to describe all gravitational fields, weak ones as well as strong ones. We show that a background-free theory is unable to handle weak gravitational fields appropriately. This is sufficient to prove that the theory is not viable as a general theory.

7.3. **What about different foliations of spacetime?** We consider everything only in a given foliation. But a background-free theory clearly should not have a preferred foliation. So, one may ask, what about different foliations? In particular, what about the problems with unitarity related with different foliations of spacetime already in a Minkowski context, as described in [15]?

[4]The classical example are the various impossibility theorem for hidden variable theories, starting with von Neumann's, where one "impossible" theory is explicitly known today as de Broglie-Bohm pilot wave theory.

Again, our answer is simple: We don't have to care about this. If the final quantum theory of gravity is able to handle different foliations of spacetime, or if it solves the "problem of time" in a quite trivial way, using some preferred foliation, does not matter at all. Whatever, the theory should be able to handle appropriately at least one foliation of spacetime. Our theorem shows that even spatially background-free theories are not viable, thus, a background-free theory would not be viable even if restricted to a single foliation.

7.4. **Maybe the formula for the scalar product is false?** The formula for the scalar product (1.1) is of central importance for our considerations. But it is a formula supposed to be valid in a domain – superpositions of relativistic gravitational fields – where we have no established theory. So one may argue that this formula may be invalid in this domain.

But there is good evidence for this formula. First, it holds in the non-relativistic limit. This is simply many-particle Schrödinger theory, a well-known and well-established theory, which is in parts even supported by experiment [8]. In this theory, the source particle and the lightweight particle are two particles which interact via the Newtonian potential. Once the source particle is assumed to be much heavier, and to be in states $|g_r\rangle$ and $|g_l\rangle$ strongly localized near the slits $x_{r/l}$, one can use even the one-particle theory with the Newtonian potential created by the source particle in $\mathbf{x}_{r/l}$ as an external field to compute $\psi_{l/r}(\mathbf{x}, t)$ for the initial value $\psi_{l/r}(\mathbf{x}, t_0) = \psi_0(\mathbf{x})$:

$$i\partial_t \psi_{l/r}(\mathbf{x}, t) = (-\frac{1}{2m}\Delta - \frac{mM}{|\mathbf{x} - \mathbf{x}_{l/r}|})\psi_{l/r}(\mathbf{x}, t) \tag{7.1}$$

The resulting functions $\psi_{l/r}(\mathbf{x}, t)$ define two approximate solutions of the two-particle problem

$$\Psi_{l/r}(\mathbf{x}_1, \mathbf{x}_2, t) = \delta(\mathbf{x}_1 - x_{l/r})\psi_{l/r}(\mathbf{x}_2, t). \tag{7.2}$$

The derivation given in the proof of theorem 2 is then valid in two-particle theory, with the scalar product for the lightweight particle is defined by (1.1). So, in this limit, (1.1) holds.

But there is also another limit where (1.1) holds: Semiclassical theory. This theory is applicable if the gravitational fields $g^r_{\mu\nu}(\mathbf{x})$ and $g^l_{\mu\nu}(\mathbf{x})$ are equal. In this case, the gravitational interaction is irrelevant. But the formula nonetheless holds. If the source particle and the lightweight particle interact in other ways, for example electromagnetically, the formula gives nontrivial predictions for electromagnetic interference effects in external gravitational fields. Experiments in this domain are standard, and, given that gravitational redshift is a relativistic effect, relativistic gravity is required.

Thus, the formula holds in two different limits: The non-relativistic as well as the semiclassical one.

Third, we have the proof of theorem 2. This gives, using quite general quantum principles, that the result has to be of the same general type, namely a scalar product.

Moreover, we do not need the formula nor for strong gravitational field, nor as an exact formula even for weak fields. All we need is an approximate formula for weak relativistic fields. An approximate formula is sufficient because for the different states constructed in our hole argument the value of $|\theta|$ varies from the maximal possible value 1 to the minimal 0, which is clearly too much for an approximation error.

But even in the worst case, even if we do not have the explicit formula (1.1), theorem 4 gives us an observable background. This background may not have the nice properties following from the theorem 5, but a background is a background is a background, and our goal was to prove that no background-free theory is viable.

7.5. **Maybe Newton-Cartan is more appropriate than Newton?** In the previous point, we have used Newtonian theory as the non-relativistic limit to justify the scalar product formula. But is this justified? It is known that Newton-Cartan theory the correct $c \to \infty$ limit of GR, and this theory is already quantized [4].

But the Newtonian theory as a limit is not wrong – we simply need additional restrictions to obtain it. We have preferred it because of it's simplicity, in particular in application to our two-particle problem. Moreover, as we have already mentioned, we do not have to use the most advanced approximation if we want to prove an impossibility theorem: The weaker the approximation, the larger the class of theories covered.

7.6. **Maybe in quantum gravity we have no certain subdivision into systems?** The proof of theorem 2 requires that we have a subdivision into subsystems – the heavy source particle and the lightweight environment particle. The Hilbert space of the composition has to be the tensor product of the Hilbert spaces of the subsystems. But maybe this rule no longer holds in quantum gravity?

We can repeat here all what has been said in the previous point about the validity in two different limiting cases and that we need the scalar product only as an approximation. In particular, the subdivision into systems, with a separate scalar product on each factor space, may be an approximate structure as well.

Nonetheless, we want to add yet another point: Even in this case, theorem 1 gives us some complex value as the result of a measurement. A complex value which has to be predicted by the theory, and we have no idea how to do this in a background-free theory.

7.7. **Maybe in full quantum gravity it is meaningless to talk about interacting particles?** In our thought experiment, we use particles as if they were simply quantum billiard balls. But already in semiclassical field theory we have to handle particle creation and destruction by the gravitational field, and it seems much more appropriate to consider particles only as phenomenological, derived objects: The fields seem to be more fundamental. In full quantum gravity the situation may be even worse. It may become completely meaningless to talk about particles – we don't know yet. So, one may argue that our thought experiment becomes meaningless in full quantum gravity.

But even in full quantum gravity there will be a domain where particles are a good approximation. This is certainly so for the physical situations which we observe today. And this will remain to be true in the domain of non-relativistic, Newtonian quantum gravity. In this domain, our thought experiment clearly makes sense.

Moreover, the theorems 1 and 2 do not require at all that we consider interactions with a single particle, not even with some particles – in theorem 1 we don't make any assumptions about the part of the environment our source particle interact with, and all we need in theorem 2 is that it interacts with some other system, which has a separate Hilbert space with a scalar product in it.

7.8. **Maybe the experiment makes sense only for such weak fields that the Newtonian limit is sufficient to handle it?** If one accepts that the experiment makes sense for some weak fields, where Newtonian gravity is able to handle it approximately, one could try to argue the other way around: Maybe the domain where the experiment makes sense is sufficiently small, essentially containing only the Newtonian limit? In this case, one could argue that in this limit we can use the fixed background of Newtonian theory to predict the result of the experiment.

There is much to answer. First, it would be doubtful that a theory which uses the background in the Newtonian limit is background-free. It is known, in particular, that the true $c \to \infty$ limit of GR is not Newtonian theory, but Newton-Cartan theory, which leads to Newtonian theory only if one imposes an additional condition [4]. What will be the true $c \to \infty$ limit of a background-free full quantum gravity? It seems reasonable to guess that it will be a background-free quantization of Newton-Cartan theory, where some principle of background-independence similar to definition 2 holds. This theory would face the same impossibility problem as a background-free GR quantization.

Of course, in Newton-Cartan theory it is natural to use the background structure which exists in this theory for quantization. But one should not expect that this quantum theory appears as the limit of a background-free theory. It is much more reasonable to expect this theory as the limit of a GR quantization which uses some background too, for example, of a quantization of GR in harmonic gauge.

Second, the background exists only in the limit. It does not exist for weakly relativistic fields. However weak the relativistic effects may be in a certain situation, there is always some level of accuracy such that the Newtonian limit is unable to make a sufficiently accurate prediction. For this level of accuracy, the Newtonian limit itself already does not provide any fixed background. Of course, one would need only an approximate background, but this approximate background should be already a structure of relativistic gravity, different from the background in the limit.

Third, it seems unreasonable to expect that the experiment makes sense only in an extremely small environment of the Newtonian limit. Especially if we look at theorem 1 and it's proof, we find that the thing our source particle interacts with does not have to be specified at all. We will obtain some observable complex value θ even if the lightweight particle appears highly relativistic, if there will be a lot of relativistic particle creation and destruction and so on. In theorem 2 we don't need much more: Only the quite general principle of quantum theory that the system consisting of the source particle and it's environment has to be described by a tensor product of the two Hilbert spaces, with a scalar product defined in each of them. Even if we doubt that this principle survives in full quantum gravity, it is hard to expect that it starts to fail already for weak to moderate relativistic fields.

7.9. **Maybe one has to consider the full experiment?** There is some standard way to describe the double slit experiment, with the particle starting from one starting point and ending at some other point at the screen. One could reasonably hope that a consideration which includes this full experiment does not require the consideration of any superpositions of gravitational fields: We do not have to consider the intermediate states, we only have to compute the amplitudes between initial and final states.

But this simple picture of the double slit experiment is not a necessity. Instead, it is an idealization. A more general quantum experiment starts with states which already are superpositional states, and never have been localized in one point. In particular, the superpositional state may be the ground state of some appropriate double slit potential. Now, ground states may be prepared starting with an arbitrary state and waiting long enough for the energy being radiated away. In the same way, one may measure if the particle has remained in the ground state after the interaction: We look if the state radiates the energy which has been obtained during the interaction.

Then, we do not need a description of a particular experiment, but a general theory which allows to describe them all. But at every finite moment of time there will be not only semiclassical states, but also their superpositions. Thus, one would need a theory which is quite different from quantum theory: Possibly without any information about finite times, like an S-matrix theory, possibly without superpositions as legitimate states.

7.10. **But maybe the true theory of quantum gravity will be something like an S-matrix theory?** In quantum field theory, we obtain physical predictions by computing the S-matrix. How to compute predictions beyond the S-matrix, for finite distances, in a gauge-invariant way is not clear. But this is widely considered as irrelevant – all what we can compare with observations is the S-matrix.

In this situation, one may expect that in quantum gravity all we can compute is some analogon of the S-matrix as well. Taken together with some positivistic philosophy, it could be argued that some S-matrix-like theory of quantum gravity would be sufficient. For such a theory, it is not clear if our thought experiment makes sense – we have used in our argument bounded states at finite times. Thus, one may hope that restriction to some S-matrix-like theory allows to avoid this problem.

Such a line of argumentation should be clearly rejected: The S-matrix is sufficient only FAPP (for all practical purposes). But FAPP we simply don't need quantum gravity. We need quantum gravity for theoretical reasons: The true, final theory of the universe should be able to handle effects of quantum gravity, even if we, as human beings, are unable to test these predictions. A theory which does not cover quantum gravity is, therefore, insufficient theoretically, even if it covers, like semiclassical QFT on a GR background, all we can observe.

But this same line of argumentation which makes full quantum gravity interesting for us also shows that an S-matrix-like theory, which would be unable to make predictions for finite distances, is insufficient too. Our human experience is bounded to finite distances, thus, the infinite limits gives by the S-matrix are not what we observe, but only an approximation. If this approximation is sufficient FAPP or not is completely irrelevant for it's theoretical status as an approximation. The true theory of the universe should be able to predict our observations at least in principle with an accuracy higher than that given by the S-matrix.

7.11. **The relationalist argument.** One reason for the attractiveness of the background-independent approach is it's philosophical background. It goes back to the position of Leibniz, who has proposed arguments for a relational view, against an absolute notion of space and time proposed by Newton. A nice introduction, from point of view of the modern background-independent approach, can be found in [11]:

> "Leibniz's argument for relationalism was based on two principles, which have been the focus of many books and papers by philosophers to the present day. The *principle of sufficient reason* states that it must be possible to give a rational justification for every choice made in the description of nature. . . . A theory that begins with the choice of a background geometry, among many equally consistent choices, violates this principle.. . .
>
> One way to formulate the argument against background spacetime is through a second principle of Leibniz, *the identity of the indiscernible*. This states that any two entities which share the same properties are to be identified. Leibniz argues that were this not the case, the first principle would be violated, as there would be a distinction between two entities in nature without a rational basis. If there is no experiment that could tell the difference between the state in which the universe is here, and the state in which it is translated 10 feet to the left, they cannot be distinguished. The principle says that they must then be identified. In modern terms, this is something like saying that a cosmological theory should not have global symmetries, for they generate motions and charges that could only be measured by an observer at infinity, who is hence not part of the universe."

From point of view of the Popperian scientific method [9, 10] relationalism has to be rejected as a variant of positivism, based on the priority of observation: The observation that some entities are indiscernible requires their identification in the theory. Instead, Popperian science is based on the priority of theory: Theories are free guesses about Nature, and they are, in particular, free to postulate differences between indiscernibles if this gives some advantages. But this rejection of relationalism in principle does not diminish the heuristic value of it's principles. While we are free to postulate differences between indiscernibles, we have to take into account Ockham's razor: Without good reasons to introduce differences between indiscernibles we should nonetheless identify them.

Therefore it is important to recognize that our thought experiment invalidates the relationalistic argumentation against the background. First, the new observable defines sufficient reason for the introduction of a common background: The background solves the problem of computing a prediction for the new observable, and there is no obvious alternative way to solve this problem. It is also no longer possible to apply the principle of "identity of the indiscernible" against the background. Indeed, superpositional states with different values for $\langle\psi_l|\psi_r\rangle$ are no longer indiscernible. Last but not least, the background becomes itself a relational object: As constructed here, the background defines a *relation* between physical objects – two gravitational fields which are part of one superpositional state.

8. Consequences for classical gravity

We conclude that a background-free theory of quantum gravity is not viable. This has also consequences for classical gravity, because one thing we know for sure: That classical gravity has to be the $\hbar \to 0$ limit of quantum gravity. Once quantum gravity has a background, then there will be such a background in classical gravity too.

Thus, even if GR seems adequate for classical gravity, we have to incorporate a background into classical gravity. In particular, we need an equation for the background. Fortunately, the choice is simple – essentially, there is only one sufficiently beautiful candidate: The harmonic coordinate condition

$$\partial_\mu(g^{\mu\nu}\sqrt{-g}) = \Box X^\nu = 0 \tag{8.1}$$

is not only very beautiful condition in itself, but also simplifies the Einstein equations essentially. Harmonic coordinates have been used to prove local existence and uniqueness results for general relativity [2, 3], so that there is no non-uniqueness and no hole problem in GR in harmonic coordinates.

The identification of the background with harmonic coordinates already leads to some interesting modifications of GR: The topology of the solutions has to be trivial, we have a new notion of completeness (a solution is complete if it is defined for all values $-\infty < X^\mu < \infty$ of the harmonic coordinates X^μ, while the metric $g_{\mu\nu}(x)$ no longer has to be geodesically complete), and we have a symmetry preference for the flat FRW universe (it is the only homogeneous universe).

Further modifications appear if we don't want to add the harmonic condition as an additional external equation, but want to obtain it as an Euler-Lagrange equation. The natural way to do this is to add a covariance-breaking term

$$L_{GR} \to L_{GR} + \gamma_{\alpha\beta} g^{\mu\nu} X^\alpha_{,\mu} X^\beta_{,\nu}\sqrt{-g} \tag{8.2}$$

for some constants $\gamma_{\alpha\beta}$ so that the harmonic condition becomes an Euler-Lagrange equation

$$\frac{\delta S}{\delta X^\alpha} = \gamma_{\alpha\beta}\Box X^\beta. \tag{8.3}$$

This leads to additional terms in the Einstein equations. While we obtain the Einstein equations in the natural limit $\gamma_{\alpha\beta} \to 0$, this leads to even more qualitative modifications of GR even for arbitrary small values $\gamma_{\alpha\beta}$: Gravitons obtain a mass, we obtain stable "frozen stars" instead of black holes, and a "big bounce" instead of a big bang, above without the GR big bang and black hole singularities [7, 6, 12].[5]

9. Conclusion

We have proven that a background-free quantum theory of gravity is not viable, is unable to predict the result of simple quantum experiments.

This is not simply yet another item on the long list of problems of "we don't know what to do" type on the way to such a theory. We know what to do, namely to use a theory with background. The main objection against the background –

[5] Once a classically unobservable space-time background is introduced anyway, this background can be used to solve the "problem of time" of canonical quantum theory too – in the most trivial way of introducing a preferred time. The corresponding ADM decomposition allows a condensed matter interpretation of gravity in terms of density, velocity and stress tensor so that the harmonic condition transforms into continuity and Euler equations [12, 13]. Because we know how to quantize condensed matter theories, such an interpretation may be useful for quantization. A preferred time may be useful for other problems too, in particular it may be used in realistic hidden variable interpretations of quantum theory like de Broglie-Bohm pilot wave theory, or for the condensed matter interpretation of the SM fermions and gauge fields proposed in [13].

it's unobservability – is no longer valid, because the new quantum observables we have found allow to recover the background.

This result is not only relevant for quantum gravity – we have to introduce the background into classical gravity too, which leads to important modifications.

Let's conclude with a remarkable observation: As in this case, as in the case of the Bohm-Aharonov effect the new quantum observables – the background as well as the gauge potential – have been known long before quantum theory. This is a lecture about the remarkable power of mathematical simplicity and beauty, and the counterproductiveness of the positivistic rejection of unobservables.

References

[1] J. Anandan, Gravitational phase operator and cosmic string, Phys. Rev. D53, nr. 2, p.779-786, 1996

[2] Y. Choquet-Bruhat, Theoreme d'existence pour certain systemes d'equations aux derivees partielles non lineaires, Acta Math., 88, 141 (1952)

[3] Y. Choquet-Bruhat, Cauchy problem, in L. Witten (ed.) Gravitation: an Introduction to Current Research, Wiley, New York (1962)

[4] J. Christian, Exactly soluble sector of quantum gravity, gr-qc/9701013 (2008)

[5] A. Einstein, Die formale Grundlage der allgemeinen Relativitaetstheorie, Preussische Akademie der Wissenschaften, Sitzungsberichte (1914), 1030-1085.

[6] A.A. Logunov et.al., Theor.Math.Phys. 69 Nr 3, 1986

[7] A.A. Logunov, Theor.Math.Phys. 70 Nr 1, 1987

[8] V. Nesvizhevsky et al, Nature, vol. 415, p. 297, 2001

[9] K.R. Popper, Conjectures and Refutations, Routledge & Kegan Paul, London, 1963

[10] K.R. Popper, Logic of scientific discovery, Routledge, London, 1959

[11] L. Smolin, The case for background independence, arXiv:hep-th/0507235 (2005)

[12] Schmelzer, I.: A generalization of the Lorentz ether to gravity with general-relativistic limit. arXiv:gr-qc/0205035 (2002)

[13] Schmelzer, I.: A Condensed Matter Interpretation of SM Fermions and Gauge Fields, Foundations of Physics, vol. 39, 1, p. 73 (2009)

[14] T. Thiemann, Introduction to modern canonical quantum general relativity, arXiv:gr-qc/0110034 (2001)

[15] Torre, C. G., Varadarajan, M., Functional evolution of free quantum fields, arXiv:hep-th/9811222, Class.Quantum Grav. 16, 2651 (1999)

E-mail address: `ilja.schmelzer@gmail.com`
URL: `ilja-schmelzer.de`