An Observer in de Sitter Space

Edward Witten

March 15, 2023

In ordinary quantum mechanics, we do not usually incorporate the observer as part of the system.

That is fine for most purposes, but if one is studying gravity, one has to take into account the fact that the observer gravitates. Still, in an open universe, it is likely not essential to explicitly incorporate an observer in the description, because in a general open universe, the gravity of the observer can be arbitrarily weak. It is in a closed universe that one might expect that it is essential to include the observer in the description. (This might be the situation also in some open universes in which the region of space visible to the observer is compact).

I will give a concrete example – the static patch of de Sitter space – where it is important to explicitly include the observer in the description. (Based on arXiv:2206.10790 by Chandrasekharan, Longo, Penington, and EW.) However first I want to discuss what an observer can observe.

We are going to need two classic but not that well known results about quantum field theory:

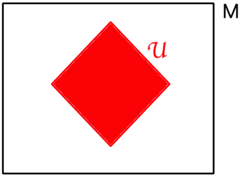
H. J. Borchers, "Field Operators as $\mathbb{C}^{\infty}$ Functions In Spacelike Directions," Il Nuovo Cimento **33** (1964) 1

and

H. J. Borchers, "Uber die Vollstandigkeit lorentzinvarianter Felder in einer zeitartigen Röhre," Il Nuovo Cimento **19** (1961) 787.

H. Araki, "A Generalization Of Borchers' Theorem," Helv. Phys. Acta 36 (1963) 132-9.

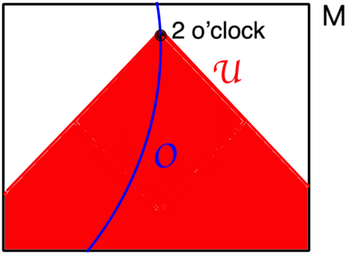
In ordinary quantum field theory – without gravity – in a spacetime M, one can arbitrarily specify any open set $\mathcal{U} \subset M$ and define an algebra $\mathcal{A}_{\mathcal{U}}$ of operators in $\mathcal{U}$:



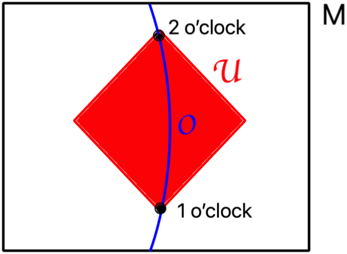
In the presence of gravity, since spacetime fluctuates, it doesn't make sense talk about the region $\mathcal{U}$ unless we have an invariant way to identify it.

For example, in an asymptotically flat spacetime, in the presence of a black hole, we could talk about the region outside the black hole horizon. That is invariantly defined and presumably makes sense (at least to all finite orders in G) even when spacetime fluctuates. We could introduce an observer who is more or less at rest near infinity and describe the region outside the horizon as the region visible to this observer. However, as I've remarked, in an open universe we do not expect that it is essential to incorporate the observer in the description.

If we assume the existence of an observer, we can invariantly identify various regions in spacetime, For example, if the observer carries a clock, we can discuss the region that is visible to the observer prior to 2 o'clock



or, alternatively, the region that is causally accessible to the observer in a stated time interval (meaning that the observer can both see and influence this region)



But what can an observer actually measure? I will assume a very simple model in which the observer is described by a timelike worldline, and what the observer can measure are simply the quantum fields along this worldline. (The observer is assumed to also carry a clock, and measuring equipment, and has access to operators that act on the measuring equipment, but I won't try to make all that explicit.) This seems like a rather minimal model of what an observer is and what the observer can measure, but it raises two immediate questions, which we can answer nicely in the absence of gravity:

(1) Can well-defined operators be defined by smearing a quantum field along a timelike worldline?

(2) Given a "yes" answer to this question, what sort of thing is the algebra generated by these operators?

Let me elaborate a bit on the first question. We are accustomed in QFT to talking about "local operators" $\phi(x)$, but a local operator isn't really a Hilbert space operator, since acting on a Hilbert space state it takes us out of Hilbert space. In the case of the vacuum state Ω in Minkowski space, this is clear from the fact that $|\phi(x)|\Omega\rangle|^2 = \infty$ or equivalently

$$\langle\Omega|\phi^\dagger(x)\phi(x)|\Omega\rangle = \infty,$$

due to a short distance singularity. Since the leading short distance singularity is universal, it is also true that $|\phi(x)|\Psi\rangle|^2 = \infty$ for any state Ψ in any spacetime M.

Thus matrix elements of $\phi(x)$ (between suitable states, such as Fock space states in a free field theory) make sense, but eigenvectors and eigenvalues of $\phi(x)$ do not make sense. If we could measure $\phi(x)$, the answer would be one of its eigenvalues, so $\phi(x)$ (for a point x along the worldline) isn't an example of what our observer can measure.

What are actually measureable are suitable smeared versions of $\phi(x)$. Let us discuss which ones. Suppose we are going to smear $\phi(x)$ over a set S to get a smeared "operator"

$$\phi_f = \int_S \mathrm{d}\mu f(x)\phi(x).$$

If this is going to be a real operator, the smearing has to be such that the $\phi^\dagger(x')\phi(x)$ OPE singularity is integrable, when smeared in this fashion.

For example, spatial smearing will only succeed if ϕ has rather low dimension. Spatial smearing (at, say, $t = 0$) produces an operator $\phi_f = \int \mathrm{d}\vec{x}\, f(\vec{x})\phi(\vec{x}, 0)$, leading to an OPE singularity

$$\int \mathrm{d}\vec{x}\, \mathrm{d}\vec{x}'\, f(\vec{x})\overline{f}(\vec{x}')\, \phi^\dagger(\vec{x}', 0)\phi(\vec{x}, 0).$$

In d space dimensions, this is integrable if and only if ϕ has dimension less than $d/2$. For example, in QCD, $d = 3$, and the lowest dimension operator is a quark bilinear $\overline{q}q$, of dimension 3, which is more than $d/2 = 3/2$. So in QCD, no operators can be defined by smearing in space.

Smearing in Euclidean space is only slightly better. If we try to define a smeared operator $\phi_f = \int \mathrm{d}x\, f(x)\phi(x)$ in $D = d + 1$ dimensional Euclidean space, we will run into the OPE singularity

$$\int \mathrm{d}x\, \mathrm{d}x' f(x)\overline{f}(x')\, \phi^\dagger(x')\phi(x).$$

This is integrable if and only if ϕ has dimension less than $D/2 = (d+1)/2$, so we have slightly extended the range in which we can define a true operator, but not enough to be able to define any operators in QCD, for example.

How then do we get true operators by smearing of "local operators"? The secret is the Feynman $\mathrm{i}\varepsilon$. No amount of smearing in space will produce a true operator (unless we start with a "local operator" of small dimension) but smearing in time turns a local operator of any dimension into a true operator. This is an old result (Borchers 1964) but I will pause to explain it a little. In doing so I will take the timelike curve to be a straight line in Minkowski space, and I will consider only the leading OPE singularity. Neither of these restrictions is essential. Subleading OPE singularities and general timelike curves in any spacetime can be treated similarly.

Suppose that at $\vec{x}=0$, we smear a "local operator" ϕ by a smooth function $f(t)$ of compact support. If ϕ has dimension Δ, then the leading OPE singularity that we run into will be

$$\int \mathrm{d}t'\,\mathrm{d}t\,\overline{f}(t')f(t)\frac{1}{(t'-t+\mathrm{i}\varepsilon)^{2\Delta}}.$$

(There are also subleading OPE singularities, but the $\mathrm{i}\varepsilon$ disposes of them in a similar way.) The integral is obviously well-defined for $\varepsilon>0$ and we want to show that no divergence appears in the limit $\varepsilon\to 0$. For this we just write

$$\frac{1}{(t'-t+\mathrm{i}\varepsilon)^{2\Delta}} = C_n\frac{\partial^n}{\partial t^n}(t-t'+\mathrm{i}\varepsilon)^{n-2\Delta},$$

for any integer $n>0$, with a constant C_n. Inserting this in the OPE integral and integrating by parts n times, we replace the original integral with

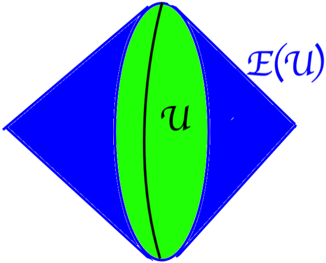
$$(-1)^n C_n\int \mathrm{d}t'\,\mathrm{d}t\,\overline{f}(t')f^{[n]}(t)\,(t-t'+\mathrm{i}\varepsilon)^{n-2\Delta}.$$

For large enough n, this is manifestly convergent for $\varepsilon\to 0$.

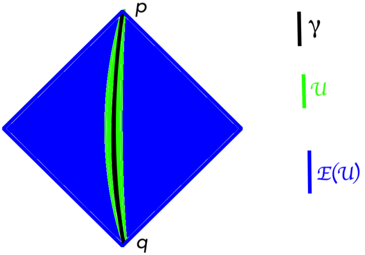
So quantum fields, smeared along the observer worldline, give well-defined operators, and it makes sense to talk about the algebra generated by such operators. We can limit the support of the smearing functions to any interval, say the interval in which the observer's proper time is in the range $\tau_1 \leq \tau \leq \tau_2$. So we can define an algebra associated to any interval on the worldline.

But what are the algebras that we make this way? In the context of quantum field theory without gravity, this question is answered by the "timelike tube theorem." This theorem was originally formulated for quantum fields in Minkowski space (Borchers 1961, Araki 1963). It was generalized to free field theories in curved spacetime by Strohmaier (2000), and in forthcoming work, Strohmaier and I hope to prove a version of the theorem for non-free theories in curved spacetime.

If $\mathcal{U}$ is an open set in spacetime, its "timelike envelope" $\mathcal{E}(\mathcal{U})$ is the smallest set that contains all points that can be reached by deforming timelike curves in $\mathcal{U}$ through a family of timelike curves:



The timelike tube theorem asserts that the algebra of operators in $\mathcal{U}$ is the same as the algebra of observables in the possibly much larger region $\mathcal{E}(\mathcal{U})$. I find this result quite striking. I will try to give at least a hint of why it is true, but first I want to explain an implication (noted by Strohmaier 2000). Suppose we are actually interested in a timelike curve γ, say one of finite extent that connects points q, p:



We can thicken γ slightly to an open set $\mathcal{U}$, such that the timelike envelope $\mathcal{E}$ does not depend on $\mathcal{U}$ – it is just the region causally accessible to γ.

The timelike tube theorem says that the algebra $\mathcal{A}(\mathcal{U})$ of operators in $\mathcal{U}$ doesn't really depend on $\mathcal{U}$ but only on γ – it is the timelike envelope $\mathcal{E}(\gamma)$. So we can define an algebra for every (possibly bounded) timelike curve γ. That should not surprise us *per se*, since we deduced the same result (in a much more elementary way) by looking at the OPE singularities. That is a pleasing consistency check. More important for our purposes, we've learned that the algebra $\mathcal{A}(\gamma)$ of operators supported on a curve γ is a good stand-in for the algebra $\mathcal{A}(\mathcal{E})$ of the spacetime region $\mathcal{E}$ with which γ is in causal contact.

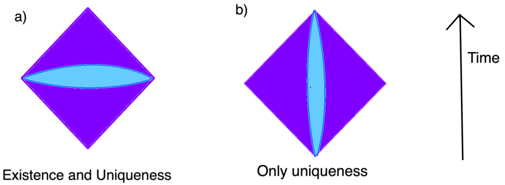
For our purposes, I think what is important about this is that $\mathcal{A}(\gamma)$ – the algebra generated by the fields along γ – is:

(1) more operationally meaningful than $\mathcal{A}(\mathcal{E})$, since it is more directly what an observer can measure

(2) better defined in the presence of gravity than $\mathcal{A}(\mathcal{E})$ would appear to be.

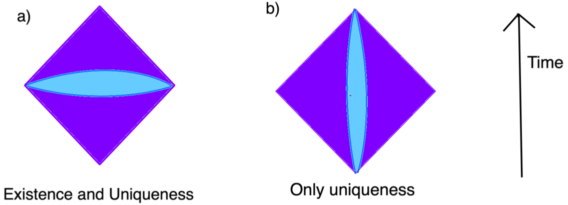
So $\mathcal{A}(\gamma)$ seems like a good substitute for algebras associated to open sets, which one would consider in the absence of gravity.

To try to give at least a hint of why the timelike tube theorem is true, I will explain the classical limit. Suppose we have a reasonable relativistic wave equation like the Klein-Gordon equation $(\Box + m^2)\phi = 0$. We are given a solution in one region $\mathcal{U}$ (light blue) and we want to predict the solution in a larger region $\mathcal{E}(\mathcal{U})$ (purple). Here are two cases:

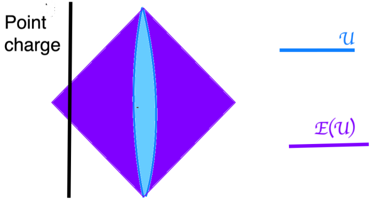


In a), one is given a solution in a "spacelike tube" and one wants to extend it over the "spacelike envelope" (which is called the domain of dependence). In b), one is given a solution in a "timelike tube" and one wants to extend the solution over the "timelike envelope."

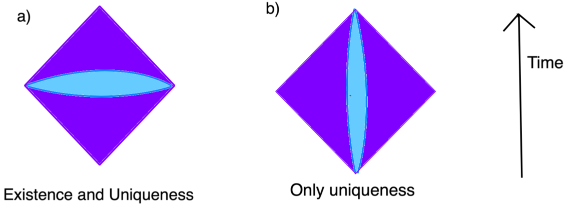
The "Holmgren uniqueness theorem" of PDE's asserts that the extension is unique, if it exists, in both cases a) and b), but existence is a more special result and only holds in case a):



To see that an existence result cannot possibly hold in case b), here is a counterexample:



The existence and uniqueness result in the case of a) is the basis for much of physics. It says that the solution can be predicted from initial data – physics is causal. But by contrast the uniqueness result without a guarantee of existence in case b)



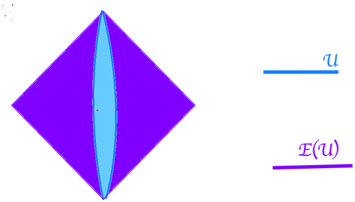
is not usually useful at the classical level because in fact the extension over $\mathcal{E}(\mathcal{U})$ of a solution on $\mathcal{U}$ usually does not exist and it is very hard to predict when it does.

Suppose, however, that we are doing quantum field theory and for simplicity consider a free field ϕ with the action

$$I = \frac{1}{2}\int_M \mathrm{d}^D x \sqrt{g}\left(-D_\mu\phi D^\mu\phi - m^2\phi^2\right).$$

In this case, we can view ϕ as an operator-valued solution of the Klein-Gordon equation $(\Box + m^2)\phi = 0$. If we are studying this quantum field theory on M, then the field $\phi(x)$ does exist throughout M and therefore existence of the extension from $\mathcal{U}$ to $\mathcal{E}(\mathcal{U})$ is not an issue.

But what does uniqueness mean?



In some sense, uniqueness means "the field $\phi(x)$ for $x \in \mathcal{E}(\mathcal{U})$ is uniquely determined by $\phi(y)$ for $y \in \mathcal{U}$." As explained by Borchers and Araki in the early 1960's, the quantum meaning of this statement is really "$\phi(x)$ for $x \in \mathcal{E}(\mathcal{U})$ is contained in the algebra generated by $\phi(y)$ for $y \in \mathcal{U}$" or equivalently the operator algebras of the two regions are the same:

$$\mathcal{A}(\mathcal{E}(\mathcal{U})) = \mathcal{A}(\mathcal{U}).$$

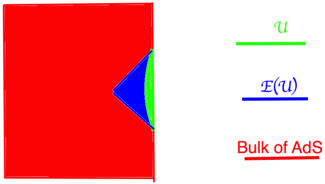
This is the timelike tube theorem.

(As I already noted, the theorem was originally proved in Minkowski space and those proofs were not limited to free field theory, while in curved spacetime the presently available proof – Strohmaier 2000 – is for free field theories.) One last detail about this: the statement "$\phi(x)$ for $x \in \mathcal{E}(\mathcal{U})$ is uniquely determined by $\phi(y)$ for $y \in \mathcal{U}$" cannot be expressed, even in free field theory, in the existence of a formula

$$\phi(x) \stackrel{?}{=} \int_{\mathcal{U}} \mathrm{d}y \, G(x, y)\phi(y)$$

for some Green function $G(x, y)$. A suitable Green function does not exist; if it did, this would imply existence of the extension over $\mathcal{E}(\mathcal{U})$ for every solution $\phi(y)$ on $\mathcal{U}$, and we have seen that in general the extension does not exist.

In the AdS/CFT correspondence, this problem has been noted in the context of causal wedge reconstruction – also called HKLL reconstruction. In that context, $\mathcal{U}$ is a small neighborhood of the conformal boundary and one wants to determine the field in the causal wedge $\mathcal{E}(\mathcal{U})$, which (usually) is the same as the timelike envelope, from the field in $\mathcal{U}$:



I believe that the timelike tube theorem is actually the correct formulation of HKLL reconstruction.

This completes what I will say about what an observer can observe.

In the rest of this talk (following a paper with Chandrasekharan, Penington, and Longo that was mentioned previously) I am going to analyze a concrete example of an observer in a closed universe with cosmological horizons. This will be de Sitter space dS_D, which is the maximally symmetric solution of Einstein's equations in $D = d + 1$ dimensions with a positive cosmological constant. It can be described by the metric

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + R^2 \cosh^2(t/R)\mathrm{d}\Omega^2$$

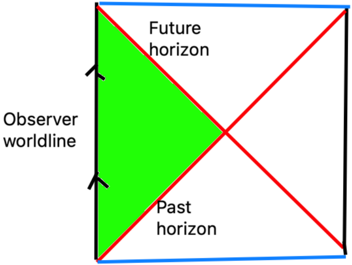
where R is the radius of curvature and $\mathrm{d}\Omega^2$ is the metric of a round sphere of unit radius. This is compact, so dS_D is an example of a closed universe. At time t, the sphere has radius

$$R(t) = R \cosh(t/R)$$

so it exponentially grows for $t \to +\infty$ (or $t \to -\infty$). The exponential growth for $t >> R$ is believed to be a good approximation to what is currently beginning to happen in the real world.

In the 1970's, Gibbons and Hawking studied de Sitter space as a simple example of a spacetime with a cosmological horizon – in which an observer cannot see the whole universe. They attached an entropy to the de Sitter horizon.

A Penrose diagram of de Sitter space is as follows



Given an observer traveling on a geodesic in dS_D, coordinates can be chosen so that the worldline of the observer is the left boundary of the diagram. If this is done, the observer has past and future horizons which are the diagonals in the picture. These diagonals bound the green region, which is the region causally accessible to the observer, and is called a “static patch.”

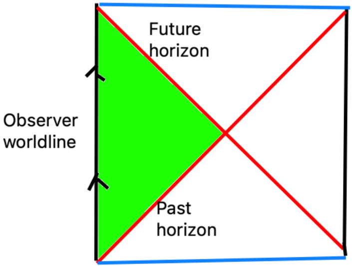
In Euclidean signature, dS_D becomes simply a D-sphere, with metric

$$\mathrm{d}\tau^2 + R^2 \cos^2(\tau/R)\mathrm{d}\Omega^2.$$

In ordinary quantum field theory in de Sitter space (and also in the presence of semiclassical gravity) there is a natural de Sitter state Ψ_{dS} such that correlation functions in this state can be obtained by analytic continuation from Euclidean signature. Let H be the generator of a one-parameter rotation subgroup of the sphere. In Euclidean signature, it obeys $\exp(-2\pi RH) = 1$. When continued to Lorentz signature, this leads to the fact that correlation functions in the state Ψ_{dS} have a thermal interpretation at the de Sitter temperature $T_{\mathrm{dS}} = 1/\beta_{\mathrm{dS}}$, where $\beta_{\mathrm{dS}} = 2\pi R$ (Gibbons and Hawking; Figari, Hoegh-Krohn, and Nappi). A slightly abstract way to describe this thermal interpretation is to say (Sewell, 1982) that the "modular Hamiltonian" of the state Ψ_{dS} is

$$H_{\mathrm{mod}} = \beta_{\mathrm{dS}} H.$$

The one parameter symmetry generator H can be chosen so that in Lorentz signature it is a symmetry of this picture



moving the observer worldline and the green region forwards in time (and moving the region spacelike separated from the observer backwards in time). In thinking about the experience of the observer, it is common to think of H as a time-translation generator and to refer to the H-invariant green region as a "static patch." What motivates this name is that coordinates can be chosen so that $H \sim \frac{\partial}{\partial t}$ and the metric of the static patch is time-independent; thus the static patch looks time-independent to the observer.

In ordinary quantum field theory, we would associate to the static patch an algebra of observables, which actually is a possibly unfamiliar Type III von Neumann algebra. This is an algebra with an infinite amount of quantum entanglement built in, giving an abstract explanation of the fact that entanglement entropy is ultraviolet divergent in quantum field theory. Including weakly coupled gravitational fluctuations does not qualitatively change the picture, but what does really change the picture is that in a closed universe, such as de Sitter space, the isometries have to be treated as constraints. This means that we should replace $\mathcal{A}_0$ by $\mathcal{A}_0^H$, its invariant subalgebra. But that does not work: the invariant subalgebra is trivial. Basically, anything that commutes with H can be averaged over all the thermal fluctuations and replaced by its thermal average, a c-number.

To get a reasonable algebra of observables, we include an observer in the analysis. Of course, in principle an observer should really be described by the theory, not injected from outside. What it really means to include an observer is that we consider a "code subspace" of states in which an observer is present in the static patch, and then we consider operators that can be defined in the low energy effective field theory in this code subspace, though they are not well-defined on the whole Hilbert space.

Should we be surprised that we need to include the observer in the analysis to get a sensible answer? As we discussed at the outset, in ordinary quantum mechanics without gravity, one can consider the observer who studies a quantum system to be external to the system. With gravity included, the observer inevitably gravitates and cannot truly be considered external to the system. However, in an open universe – for example one that is asymptotically flat – the gravity of the observer can be negligible. It is in a closed universe that it may be impossible to ignore the gravity of the observer. That is exactly the situation that we are in here because de Sitter space is a simple model of a closed universe, that is, a universe with compact spatial sections. And indeed we find that to get a sensible result we need to take into account the gravity of the observer.

As a minimal model of the observer, we consider a clock with Hamiltonian

$$H_{\mathrm{obs}} = q.$$

It is physically reasonable to assume that the observer's energy is bounded below by 0, so we assume $q \geq 0$. Thus the effect of including the observer is to modify the Hilbert space by

$$\mathcal{H}_0 \to \mathcal{H}_0 \otimes L^2(\mathbb{R}_+).$$

(Positive half-line since $q \geq 0$.) The algebra is likewise extended from $\mathcal{A}_0$ to

$$\mathcal{A}_1 = \mathcal{A}_0 \otimes B(L^2(\mathbb{R}_+)).$$

The last factor is the (Type I) algebra of all bounded operators on $L^2(\mathbb{R}_+)$; it is generated by q and by $p = -\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}q}$.

Finally the constraint becomes the total Hamiltonian of the quantum fields plus the observer:

$$H \to \widehat{H} = H + H_{\mathrm{obs}}.$$

The "correct" algebra of observables taking account of the presence of the observer is therefore

$$\mathcal{A} = \mathcal{A}_1^{\widehat{H}},$$

that is, the $\widehat{H}$-invariant part of $\mathcal{A}_1$.

This answer makes sense, unlike the previous one. The reason is that once an observer is present, we can "gravitationally dress" any operator to the observer's world-line. For any $\mathsf{a} \in \mathcal{A}_0$, the operator

$$\widehat{\mathsf{a}} = e^{\mathrm{i}pH} \mathsf{a} e^{-\mathrm{i}pH}$$

commutes with the constraint $\widehat{H} = H + q$. One more operator that commutes with the constraint is q itself (or equivalently $-H$). It follows from classic results of Connes and Takesaki from the 1970's that (1) there are no more operators that commute with the constraint, and (2) the algebra $\mathcal{A}$ that is generated by $\widehat{\mathsf{a}},\ \ \mathsf{a} \in \mathcal{A}_0$ along with q is actually a von Neumann algebra of Type II.

Since von Neumann algebras of Type II and Type III may be quite unfamiliar, for now let me just say that a Type II von Neumann algebra is much simpler than one of Type III, because it has a trace. As a result, familiar ideas like density matrices and entropies make sense. To a global state Ψ of de Sitter space plus the observer, reduced to the static patch, one can associate a density matrix $\rho \in \mathcal{A}$, characterized by

$$\langle\Psi|\mathsf{a}|\Psi\rangle = \mathrm{Tr}\,\mathsf{a}\rho, \quad \forall \mathsf{a} \in \mathcal{A}.$$

Therefore, we can define a von Neumann entropy

$$S(\rho) = -\mathrm{Tr}\,\rho \log \rho.$$

There is no such definition in the absence of gravity. The fact that gravity turns the Type III algebra into a Type II algebra gives an abstract explanation for why entropy is better defined in the presence of gravity than in ordinary quantum field theory. I should point out, however, that from a physical point of view, Type II entropy is a renormalized entropy from which an infinite constant has been subtracted.

There are actually two relevant varieties of Type II algebra and we can get either one of them in this construction:

(1) If we put no lower bound on the observer energy, we get an algebra of Type II_∞. In such an algebra, depending on ρ there is no upper bound on the von Neumann entropy. This is appropriate for describing a black hole (whose entropy can grow indefinitely as the mass increases) but not for describing de Sitter space.

(2) If we do put a lower bound on the observer energy, we get an algebra of Type II_1. The main difference for our purposes today is that in a Type II_1 algebra, there is a state of maximum possible entropy, which is the "maximally mixed" state with density matrix $\rho = 1$. (This is consistent with $\mathrm{Tr}\,\rho = 1$ because the trace in a Type II_1 algebra is usually normalized so that $\mathrm{Tr}\,1 = 1$.)

De Sitter space is believed to have a state of maximum entropy – namely "empty de Sitter space," with all the entropy in the cosmological horizon (Bousso 2000). So to get a reasonable model of de Sitter space, we should assume that the observer's energy is bounded below, which is a more reasonable assumption anyway. The maximum entropy state is then the one with density matrix $\rho = 1$. Density matrix $\rho = 1$ is the Type II_1 analog of a maximally mixed state in ordinary quantum mechanics, in which the density matrix is a multiple of the identity.

We can now compare with some claims made in the past by others (such as Banks; Susskind; Dong, Silverstein, and Torroba). First of all, since the maximum entropy state has $\rho = 1$, it has a "flat entanglement spectrum" (all eigenvalues of the density matrix are equal) and accordingly the Rényi entropies are constant:

$$S_\alpha(\rho) = \frac{1}{1-\alpha} \log \mathrm{Tr}\, \rho^\alpha = 0.$$

Given the assertion that de Sitter space has a state of maximum entropy, this is what one should expect: In ordinary quantum mechanics, the maximum entropy state of a system is "maximally mixed," with a "flat entanglement spectrum" (the density matrix is a multiple of the identity and all its eigenvalues are equal) and its Rényi entropies are independent of α.

Now, suppose that the observer makes a measurement with two outcomes that correspond to the projection operators Π and $1 - \Pi$. The probabilities of the two outcomes are $\mathrm{Tr}\,\Pi$ and $\mathrm{Tr}\,(1 - \Pi) = 1 - \mathrm{Tr}\,\Pi$. All values $0 \leq \mathrm{Tr}\,\Pi \leq 1$ are possible. If the outcome corresponding to Π is observed, then after this measurement, the density matrix is

$$\sigma = \frac{1}{\mathrm{Tr}\,\Pi}\Pi.$$

Since the two eigenvalues of σ are 0 and $1/\mathrm{Tr}\,\Pi$, one has $\sigma \log \sigma = \sigma \log(1/\mathrm{Tr}\,\Pi)$ so the entropy after the observation is

$$S(\sigma) = -\mathrm{Tr}\,\sigma \log \sigma = -\log(1/\mathrm{Tr}\,\Pi).$$

The entropy reduction from knowing the outcome is therefore $\Delta S = \log(1/\mathrm{Tr}\,\Pi)$, and this is related to the probability $p = \mathrm{Tr}\,\Pi$ of the given outcome by

$$p = e^{-\Delta S}.$$

However, the probability of a (low entropy) energy E fluctuation of the static patch is

$$p = e^{-\beta_{\mathrm{dS}} E},$$

according to the thermal interpretation of de Sitter space. Since also $p = e^{-\Delta S}$, we must have for consistency of the two descriptions

$$e^{-\beta_{\mathrm{dS}} E} = e^{-\Delta S}.$$

In other words, "'thermal" suppression of a fluctuation can be understood as purely entropic suppression. This is surprising, but it has been argued before on other grounds, notably by considering the case that the "fluctuation" is a small black hole at the center of the static patch.

Which part of this is surprising? The formula $p = e^{-\Delta S}$ for the probability of an outcome is an inevitable consequence of having a maximum entropy state in which all states are equally probable. In other words, if all states are equally likely, then the probability of a given outcome is just proportional to the number of microstates that are compatible with that outcome. Here I am using language appropriate for an ordinary quantum system with a finite-dimensional Hilbert space. A moment ago, I explained how to reach the same conclusion in the context of a Type II_1 algebra.

The surprise is not that $p = e^{-\Delta S}$, which one should expect for a maximum entropy state, but that the maximum entropy state also has a thermal interpretation. Let us discuss how to see this in the context of the Type II_1 algebra. First of all, ignoring the constraint for the moment, the time dependence of an operator $\mathsf{a} \in \mathcal{A}_0$ is defined in the usual way by

$$\mathsf{a}(t) = e^{\mathrm{i}Ht}\mathsf{a}e^{-\mathrm{i}Ht}.$$

Then time-dependent correlations such as

$$\langle \Psi_{\mathrm{dS}} | \mathsf{a}(t_1)\mathsf{a}'(t_2) | \Psi_{\mathrm{dS}} \rangle$$

have thermal properties that reflect the fact that these correlation functions can be computed by analytic continuation from Euclidean signature.

After imposing the constraint, we replace a with the dressed version $\widehat{\mathsf{a}} = e^{\mathrm{i}pH}\mathsf{a}e^{-\mathrm{i}pH}$, and again we define its time dependence by

$$\widehat{\mathsf{a}}(t) = e^{\mathrm{i}Ht}\widehat{\mathsf{a}}e^{-\mathrm{i}Ht}.$$

Then, because

$$H\Psi_{\mathrm{dS}} = 0,$$

we rather trivially find

$$\langle\Psi_{\max}|\widehat{\mathsf{a}}(t_1)\widehat{\mathsf{a}}'(t_2)|\Psi_{\max}\rangle = \langle\Psi_{\mathrm{dS}}|\mathsf{a}(t_1)\mathsf{a}'(t_2)|\Psi_{\mathrm{dS}}\rangle.$$

So correlators of gravitationally dressed operators after imposing the constraints have the same thermal properties that correlators of "bare" operators had before imposing the constraints.

Thus weakly coupled gravity does not disturb the thermal interpretation of de Sitter space, but it leads to a new interpretation, which we would not have without gravity:

The natural de Sitter state is a maximally mixed state of maximum possible entropy.

In sum, I have explained a concrete example of including an observer in order to get a sensible answer in a cosmological model with a closed universe. And, at least in the example of de Sitter space, I have explained that gravity makes the notion of entropy better defined that it is in ordinary quantum field theory. Something important that is discussed in the paper, but which I have not had time to explain, is that entropy defined this way agrees – up to an overall additive constant – with generalized gravitational entropy, as defined by formulas whose earliest version goes back to Bekenstein (1972). There is actually a similar story for a black hole, but that will have to be for another time.