On Probabilities in the Many Worlds Interpretation of Quantum Mechanics[†]

UNIVERSITY OF COLOGNE
INSTITUTE FOR THEORETICAL PHYSICS
BACHELOR THESIS

Author:
Florian BOGE
Matr.-Nr. 5733715

Supervisor:
PD Dr. Rochus KLESSE
Second Supervisor:
Prof. Dr. Claus KIEFER

[†]The present document is a slightly corrected version of the thesis originally handed in at April 4, 2016 and defended and graded at May 9, 2016.

Abstract. Quantum Mechanics notoriously faces a *measurement problem*, the problem that the unitary time evolution, encoded in its dynamical equations, together with the kinematical structure of the theory generally implies the non-existence of definite measurement outcomes. There have been multiple suggestions to solve this problem, among them the so called *many worlds interpretation* that originated with the work of Hugh Everett III [24, 28]. According to it, the quantum state and time evolution fully and accurately describe nature as it is, implying that under certain conditions multiple measurement outcomes that are seemingly mutually exclusive can be realized at the same time – but as different 'worlds' contained in a global, quantum mechanical structure, sometimes referred to as 'the multiverse' [81]. The many worlds interpretation has, however, been confronted with serious difficulties over the course of its development, some of which were solved by the advent of *decoherence theory* [46, 66]. The present thesis critically investigates the state of play on a key remaining problem of the many worlds interpretation, the problem of the meaning and quantification of *probabilities* in a quantum multiverse. Recent attempts of deriving the pivotal statistical ingredient of quantum mechanics, *Born's rule*, from either principles of decision theory [21,78–81] or from quantum mechanics alone, supplemented with a few general premises about probability [86, 87] are analyzed and their premises are scrutinized. It will be argued that, though both approaches yield promising results, they both ultimately fail to clearly establish the validity of Born's rule in the context of the many worlds interpretation. It is hence suggested that further research on this problem is indicated.

Contents

1 Introduction

Quantum Mechanics (QM) is one of science's greatest success-stories. It has produced numerous successfully confirmed predictions to unforeseen accuracy, and spawned off numerous technical implementations of different kinds, thereby giving rise to what many have thought of as nothing short of two separate *scientific revolutions*, rooted in electronics and information technology respectively [4, 48]. Yet while the formalism of QM has found completion in the form of standard, non-relativistic QM as well as (special) relativistic quantum field theory (QFT), up to date there is considerable debate about how to *interpret* the quantum formalism. In the following, we review the basic structure of the theory and subsequently introduce what has come to be known as the *measurement problem* (MP). We will here restrict our attention to non-relativistic QM, since we take it that the issues arising from the MP transfer seamlessly to QFT, due to the appearance of superpositions of field configurations therein [8, 73].

1.1 The Quantum Postulates and Basic Properties of QM

The fundamental structure of non-relativistic QM can be summarized in the form of at least the following postulates:

(I) A quantum system $\mathcal{S}$ is associated with a Hilbert space $\mathcal{H}$ and its state at time t is represented by a vector $|\psi(t)\rangle \in \mathcal{H}$.

(II) If $|a\rangle, |b\rangle \in \mathcal{H}$ represent states, then so does any linear combination $|\psi\rangle = \lambda|a\rangle + \mu|b\rangle$, $\lambda, \mu \in \mathbb{C}$, not prohibited by a superselection rule.

(III) A physical observable A is represented by a Hermitian operator $\hat{A}$ on $\mathcal{H}$ and the values of A for $\mathcal{S}$ are represented by the numbers in the spectrum of $\hat{A}$.

(IV) The temporal evolution of the vector $|\psi(t)\rangle$ associated with $\mathcal{S}$ is governed by the (time dependent) *Schrödinger equation* (SE): $\imath\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle$. $\hat{H}$ is the Hamilton operator, representing the total energy of the system $\mathcal{S}$.[1]

(V) An observable A has value a on $\mathcal{S}$ *iff* the state of $\mathcal{S}$ is given by $|a\rangle$, with $\hat{A}|a\rangle = a|a\rangle$.

(VI) If $\mathcal{S}$ is in a state represented by the (normalized) state vector $|\psi(t)\rangle$, A is an observable for $\mathcal{S}$ and $|a\rangle$ is a state such that $\hat{A}|a\rangle = a|a\rangle$, then $\mathrm{Pr}_A^{|\psi(t)\rangle}(a) = |\langle a|\psi(t)\rangle|^2$ gives the probability of finding value a for A on $\mathcal{S}$ in some measurement procedure for A.

(II) is usually called the *superposition principle*, (V) the *eigenvalue-eigenstate link*, (VI) *Born's rule*. (I) can be generalized to situations where the quantum state, resulting from some state preparation method, is uncertain, and where a whole range of states $|\psi_j\rangle$ are known to result with respective statistical weights p_j, representing (idealizations of) relative frequencies. In this case, the state is given by $\mathcal{S}$'s density operator (loosely speaking: density matrix) $\hat{\rho}_S = \sum_j p_j |\psi_j\rangle\langle\psi_j|$, with $\sum_j p_j = 1$. States of this sort are called *mixed*, those represented by (ket-)vectors $|\psi\rangle$ *pure*.[2] Born's Rule generalizes to $\mathrm{Pr}_A^{\hat{\rho}}(a) = \mathrm{Tr}(\hat{\rho}\hat{P}_a)$ in these cases, where $\hat{P}_a$ projects onto

[1] Throughout we will use $\imath$ to refer to the imaginary unit, to avoid confusions with indexing.

[2] A pure state density matrix is just a projector $|\psi\rangle\langle\psi|$ onto the subspace $\mathcal{H}_\psi \subset \mathcal{H}$ spanned by $|\psi\rangle$.

the subspace of $\mathcal{H}$ spanned by the eigenvectors of eigenvalue a for $\hat{A}$, $\mathrm{Tr}(\hat{\mathrm{O}}) = \sum_{\mathrm{j}} \langle j|\hat{O}|j\rangle$ is the trace operation, $\{|j\rangle\}_{j\in I}$ an orthonormal basis (ONB) of $\mathcal{H}$, and I some indexing set which we deliberately leave rather unspecified. The SE can then be generalized to the *von Neumann equation*

$$\frac{\partial}{\partial t}\hat{\rho}(t) = -\frac{\imath}{\hbar}[\hat{H},\hat{\rho}(t)]\,. \tag{1}$$

So far so good. But upon closer inspection, these postulates appear somewhat odd. In particular, the superposition principle (together with (V)) implies that systems can be in physical states where for a whole bunch of obervables, they do not possess definite values. Indeed, this is fundamentally rooted in the structure of the theory, since any state $|\psi\rangle$ can always be expanded as a a superposition $\sum_j \beta_j |\varphi_j\rangle$ in some suitable basis of $\mathcal{H}$, and the $|\varphi_j\rangle$ may be eigenstates of some suitable observable. The superposition principle is crucial the theory's extraordinary predictive power though, and to date this is not suitably reproduced by any more 'classical' theory [12, 44].

In virtue of (IV), the quantum mechanical time evolution is *linear*, whence it preserves superpositions over time. More precisely, it is provided in terms of *unitary* operators, i.e., operators which satisfy $\hat{U}\hat{U}^\dagger = \hat{U}^\dagger\hat{U} = \mathbb{1}$. To see this, note that if $\hat{H}$ is time-independent, one straightforwardly has

$$\imath\hbar\frac{\partial}{\partial t}\exp\Big(-\frac{\imath}{\hbar}\hat{H}t\Big)|\psi(0)\rangle = \hat{H}\exp\Big(-\frac{\imath}{\hbar}\hat{H}t\Big)|\psi(0)\rangle\,, \tag{2}$$

whence it must hold, in accord with (IV), that

$$\exp\Big(-\frac{\imath}{\hbar}\hat{H}t\Big)|\psi(0)\rangle = |\psi(t)\rangle\,. \tag{3}$$

This theme is invariant under a shift $0 \mapsto t_0$ to some arbitrary, fixed (temporal) starting point, whence we can identify

$$\exp\Big(-\frac{\imath}{\hbar}\hat{H}\cdot(t-t_0)\Big) =: \hat{U}(t;t_0) \tag{4}$$

as a general (unitary) time shift operator.[3] More precisely, $\hat{H}$ is the infinitesimal generator of time translations, because if we assume that there is *some* (linear) operator $\hat{D}$ such that

$$|\psi(t+\epsilon)\rangle = (1 + \hat{D}\epsilon + \mathcal{O}(\epsilon^2))\,|\psi(t)\rangle\,, \tag{5}$$

i.e., whose exponential translates the state by some small amount $\epsilon = t_f - t$ in time, we have

$$\frac{|\psi(t+\epsilon)\rangle - |\psi(t)\rangle}{\epsilon} = (\hat{D} + \mathcal{O}(\epsilon))\,|\psi(t)\rangle\,, \tag{6}$$

and letting $\epsilon \longrightarrow 0$, we obtain the Schrödinger equation for the choice $\hat{D} = -\frac{\imath}{\hbar}\hat{H}$.[4]

Unitarity also implies that the time evolution is completely deterministic and reversible, since

[3] For a time-dependent Hamiltonian $\hat{H}(t) = \frac{\hat{p}^2}{2m} + V(t)$, the theme generalizes to $\hat{U}(t;t_0) = \hat{\tau}\exp\Big(-\frac{\imath}{\hbar}\int_{t_0}^t \hat{H}(t)\,\mathrm{d}t\Big)$, where $\hat{\tau}$ is a time ordering operator whose effect can be written as $\hat{\tau}\hat{H}(t_1)\hat{H}(t_2) = \Theta(t_1-t_2)\hat{H}(t_1)\hat{H}(t_2) + \Theta(t_2-t_1)\hat{H}(t_2)\hat{H}(t_1)$ [71].

[4] The von Neumann equation can be derived in an analogous fashion from the Heisenberg picture, where operators such as the density matrix evolve under the unitary time evolution as $\hat{\rho}(t+\tau) = \hat{U}^\dagger(\tau)\hat{\rho}(t)\hat{U}(\tau)$.

given any initial state $|\psi(t_0)\rangle$, its future states are fully determined by the set of Hamiltonians that apply to it over time, and since the evolution can always be 'undone' by an operator $\hat{U}^\dagger$. Unitary operators are also *norm preserving*, meaning that if $|\tilde{\psi}\rangle = \hat{U}\,|\psi\rangle$ then $\langle\tilde{\psi}|\tilde{\psi}\rangle = \langle\psi|\hat{U}^\dagger\hat{U}|\psi\rangle = \langle\psi|\psi\rangle$. This implies that probabilities, as defined by the Born rule, are conserved over time as well.

That this linear and unitary time evolution, in concert with (II) and (V), does not generally lead to definite outcomes—i.e., definite values of observable physical quantities, including positions, speeds, and the like—as we should expect from experimental practice and everyday life experience, can be seen by appeal to a simple toy example (cf. [9]). Consider a spin-$\frac{1}{2}$ system in a magnetic field $\boldsymbol{B} = B\cdot\boldsymbol{n}_z$ (here treated classically). Then the spin magnetic moment $\boldsymbol{\mu}$ of the system will couple to the field, and the interaction part of the Hamiltonian is given by

$$\hat{H}_{\text{int}} = -\boldsymbol{\mu}\boldsymbol{B} = \underbrace{-\frac{\gamma\hbar B}{2}}_{=:\varepsilon}\hat{\sigma}_z, \tag{7}$$

where $\gamma = g\frac{e}{2m}$ is the gyromagnetic ratio, $g \approx 2$, and $\hat{\sigma}_z \doteq \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$ a Pauli matrix.[5] This Hamiltonian has eigenstates $|{\uparrow_z}\rangle$ and $|{\downarrow_z}\rangle$ of eigenvalue $\pm\varepsilon$. Assume, however, that the system was prepared – e.g. by blocking one beam in a suitable Stern-Gerlach-arrangement – with its spin pointing up along the x-axis, $|{\uparrow_x}\rangle = \frac{1}{\sqrt{2}}(|{\uparrow_z}\rangle + |{\downarrow_z}\rangle) \doteq \frac{1}{\sqrt{2}}\binom{1}{0}$, which by (II) adequately defines a state of the system, and that we subject it to the time evolution given by the interaction Hamiltonian. In matrix representation, we then have

$$\begin{aligned}\hat{U}(t) = \exp\!\Big(-\frac{\imath}{\hbar}\hat{H}_{\text{int}}t\Big) &\doteq \sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{\imath\varepsilon t}{\hbar}\right)^n\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}^n = \\ &= \sum_{n=0}^{\infty}\frac{1}{2n!}(-1)^n\left(\frac{\varepsilon t}{\hbar}\right)^{2n}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \imath\sum_{n=0}^{\infty}\frac{1}{(2n+1)!}(-1)^n\left(\frac{\varepsilon t}{\hbar}\right)^{2n+1}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} = \\ &= \begin{pmatrix}e^{-\imath\frac{\varepsilon}{\hbar}t} & 0\\ 0 & e^{\imath\frac{\varepsilon}{\hbar}t}\end{pmatrix}.\end{aligned} \tag{8}$$

Applying this to $|{\uparrow_x}\rangle$, we arrive at

$$|\psi(t)\rangle \doteq \begin{pmatrix}e^{-\imath\frac{\varepsilon}{\hbar}t} & 0\\ 0 & e^{\imath\frac{\varepsilon}{\hbar}t}\end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}e^{-\imath\frac{\varepsilon}{\hbar}t}\\ e^{\imath\frac{\varepsilon}{\hbar}t}\end{pmatrix} \doteq \frac{1}{\sqrt{2}}\left(e^{-\imath\frac{\varepsilon}{\hbar}t}\,|{\uparrow_z}\rangle + e^{\imath\frac{\varepsilon}{\hbar}t}\,|{\downarrow_z}\rangle\right) = \frac{e^{-\imath\frac{\varepsilon}{\hbar}t}}{\sqrt{2}}\left(|{\uparrow_z}\rangle + e^{\imath\frac{\Delta\varepsilon}{\hbar}t}\,|{\downarrow_z}\rangle\right), \tag{9}$$

with $\Delta\varepsilon = \varepsilon - (-\varepsilon) = 2\varepsilon$. The system now appears to be oscillating between the two states $|{\uparrow_x}\rangle, |{\downarrow_x}\rangle$ up to an overall phase of $e^{-\imath\frac{\varepsilon}{\hbar}t}$, so that in virtue of (V), at times $t \neq n\frac{2\pi\hbar}{\varepsilon}, n \in \mathbb{Z}$, it assumes no value for observable μ_x at all, where $\hat{\mu}_x := \frac{\varepsilon}{B}\hat{\sigma}_x$. According to Born's rule, (VI), one will respectively measure $\pm\varepsilon/B$ for the magnetic moment (and the spin) being up or down in x-direction with probabilities

$$\Pr_{\mu_x}^{|\psi(t)\rangle}(+\varepsilon/B) = |\langle{\uparrow_x}|\psi(t)\rangle|^2 = \cos^2\left(\frac{\varepsilon}{\hbar}t\right), \tag{10}$$

[5] We use $\doteq$ to denote a choice of representation.

$$\mathrm{Pr}_{\mu_x}^{|\psi(t)\rangle}(-\varepsilon/B) = |\langle\downarrow_x|\psi(t)\rangle|^2 = \sin^2\left(\frac{\varepsilon}{\hbar}t\right), \tag{11}$$

where $|\downarrow_x\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle - |\downarrow_z\rangle)$. The probability of finding one of the respective states thus equally 'oscillates' in time, with frequency $\omega = \frac{\varepsilon}{\hbar}$. It appears that for most of the time, the system is not in any definite state w.r.t. observable μ_x at all, but only assumes, for any given time t, a definite value upon *measurement*—in a predictable but stochastic manner.

Unitary operators pervade QM as a formal representation of state transformations throughout. An operator that represents shifts in position can be given by $\hat{U}(\boldsymbol{x}) = e^{-\frac{i}{\hbar}\boldsymbol{x}\hat{\boldsymbol{p}}}$; one that rotates a wavefunction (the position representation of a state vector) by an angle θ by $\hat{U}(\theta) = e^{\frac{i}{\hbar}\boldsymbol{n}_\theta\hat{\boldsymbol{L}}}$ ($\hat{\boldsymbol{L}}$ the angular momentum operator). But these operators do not produce states which contain definite values for all observables, as is equally present in the SE. The fact that we find a given system to yield, however, a definite value upon a suitable measurement, and that the objects we encounter in everyday life appear to have rather definite speeds and positions, say, leads to the so called *measurement problem*. We will outline its formal subtleties in more detail below, but we should first make the notion of measurement more precise for the present context.

1.2 Measurements in QM and the Measurement Problem

Generally speaking, there are multiple intricacies in spelling out what a measurement is. Intuitively, the word includes *someone*, a conscious agent or 'observer', who executes a certain procedure on a physical system to determine (a selection of) its physical properties, i.e. the values for certain observable magnitudes (observables) as they pertain to the system, by physically interacting with it with the aid of suitable equipment. Stated thusly, one can, in principle, determine three stages of this process:

1. interaction of the system under investigation with the equipment,
2. interaction of the observer with the equipment,
3. registration of a value (measurement result) by the observer.

Formally modeling these three stages respectively, however, we arrive at rather different views of the measurement process. From the point of view of the observer, in the third stage the system appears to have evolved as

$$|\psi\rangle \longrightarrow \frac{\hat{P}_a|\psi\rangle}{\|\hat{P}_a|\psi\rangle\|}, \tag{12}$$ [6]

if observable A is measured to have value a and its state before the measurement was $|\psi\rangle$ (on account of the eigenvalue-eigenstate link). That this is the actual evolution a system undergoes in suitable circumstances was explicitly postulated, in addition to (I)-(VI), by Dirac [25] and von Neumann [77] in their respective seminal works. This postulate usually goes by the name *projection postulate*. We will refer to (I)-(VI) plus the projection postulate as 'the standard formulation', as it is basically the way that QM is *used* in practice.

[6] The norm $\|\cdot\|$ of a ket-vector is defined as $\||\phi\rangle\| := \sqrt{\langle\phi|\phi\rangle}$.

In the above case, the measurement is represented by a family $\{\hat{P}_{o_j}\}_{j\in I}$ of orthogonal projectors ($\hat{P}_{o_j}\hat{P}_{o_k} = \delta_{jk}\hat{P}_{o_j}$) that satisfy $\sum_j \hat{P}_{o_j} = \mathbb{1}$, with $\hat{P}_{o_j}$ the projector onto the (not necessarily one dimensional) subspace of j-th eigenvalue for observable O. Such sets are called *projector valued measures* (PVM) and the associated (projective) measurements are repeatable in the sense that $\hat{P}^2_{o_j} = \mathbb{1}$. Up to degeneracy, they uniquely single out a definite pure quantum state. A generalization of the projection postulate to situations of non-repeatable measurements (e.g. measuring a photon's position by a silvered screen [58]) is *Lüders' Rule* [51] in its most general form [42]:

$$\hat{\rho} \longrightarrow \frac{\hat{M}_m\hat{\rho}\hat{M}^\dagger_m}{\mathrm{Tr}(\hat{M}_m\hat{\rho}\hat{M}^\dagger_m)} \tag{13}$$

Here, the operators $\hat{M}_m$ are called *measurement operators*, and they satisfy $\hat{M}^\dagger_m\hat{M}_m = \hat{E}_m$,[7] with $\{\hat{E}_m\}_{m\in J}$ a *positive operator-valued measure* (POVM), and where the ms represent outcomes not uniquely indicative of a particular pure quantum state. The POVM is defined by the properties that $\sum_m \hat{E}_m = \mathbb{1}$ and $\langle\psi|\hat{E}_m|\psi\rangle \geq 0, \forall\,|\psi\rangle \in \mathcal{H}, \forall m \in J$ and can generally be understood as a coarse-graining over projectors that need not preserve the orthogonality constraint, e.g. in virtue of the known statistics of an ancillary system, coupled to the system under investigation [58,60].

The first stage of the process as we have identified it above, however, is entirely different in nature. It describes an interaction between two physical systems (sometimes also referred to as a *premeasurement* interaction [15, 66]) and should hence be describable by a suitable interaction Hamiltonian. That this is indeed possible was first discovered by von Neumann [77], and the argument has been restated in various forms (e.g. [46, 56]). One version [56] proceeds by defining a (symbolic) 'pointer momentum' $\hat{P}$ for the measuring device $\mathcal{M}$, that satisfies a canonical commutation relation $[\hat{X}, \hat{P}] = \imath\hbar$ with the (symbolic) 'pointer position' $\hat{X}$. Suppose now that the 'pointer' (which could equally be a counter or an arrangement of light bulbs or what have you) is in a *ready state* $|X_0\rangle$, and is shifted by the interaction with $\mathcal{S}$ to a value indicative of $\mathcal{S}$'s state $|o_j\rangle$ (an eigenstate of observable O). Assuming, as a kind of idealization, that the effect of the interaction on the system is negligible, i.e., that its state stays almost the same, the interaction should effect an evolution $|o_j\rangle|X_0\rangle \longrightarrow |o_j\rangle|X_j\rangle$, with the pointer's state $|X_j\rangle$ indicative of the system's state $|o_j\rangle$. In fact, this – or a slightly loosened version thereof, where merely a probabilistic coupling is required [15] – is *constitutive* of what counts as a (good) measurement of a quantity; that the state of the apparatus is suitably calibrated with the state of the system measured.

With $\hat{H}_{\mathrm{int}} := -\frac{\lambda\hbar}{\Delta t}(\hat{O}\otimes\hat{P})$, where λ indicates strength and duration of the interaction, one now obtains $\hat{U}(\Delta t) = e^{\imath\lambda(\hat{O}\otimes\hat{P})}$ as the unitary operator effecting the state transition, and formally finds

$$\hat{U}(\Delta t)\,|o_j\rangle|X_0\rangle = |o_j\rangle\, e^{\imath\lambda(o_j\hat{P})}\,|X_0\rangle = |o_j\rangle|X_0 - \lambda o_j\rangle\,, \tag{14}$$

using the commutation relation between $\hat{X}$ and $\hat{P}$, and the formal analogy between the position-shift operator (cf. the previous section) and the operator $e^{\imath\lambda(o_j\hat{P})}$. The output of $\mathcal{M}$ will generally be a function of the physical value pertaining to $\mathcal{M}$ in virtue of the interaction, and the choice $f(X_j) := \frac{X_0 - X_j}{\lambda} = o_j$ establishes the desired joint evolution outlined above [56].

[7] We here assume the Hilbert space to be separable, whence we can always assume the existence of such operators $\hat{M}_m$, in virtue of the square-root lemma [36].

However, if the system is in a superposition $|\phi\rangle = \sum_j \alpha_j |o_j\rangle$ w.r.t. the observable measured, then the evolution proceeds as

$$\hat{U}(\Delta t)\,|\phi\rangle|X_0\rangle = \sum_j \alpha_j\,|o_j\rangle\,e^{\imath\lambda(o_j\hat{P})}\,|X_0\rangle = \sum_j \alpha_j\,|o_j\rangle|X_j\rangle\,, \tag{15}$$

so that $\mathcal{M}$ has evolved into a superposition of states as well, entangled with the state of $\mathcal{S}$.[8]

Above, we have assumed the ideal case where the state of $\mathcal{S}$ does not change as a function of the interaction, but in the non-ideal case, the evolution may be assumed to proceed in a similar fashion as

$$|\phi\rangle|X_0\rangle = \sum_j \alpha_j\,|o_j\rangle|X_0\rangle \xrightarrow{\hat{U}} \sum_{j,k} \alpha_{jk}\,|o_j\rangle|X_k\rangle = \sum_k \tilde{\alpha}_j\,|\tilde{o}_j\rangle|X_j\rangle\,, \tag{16}$$

with $|\tilde{\alpha}_k|^2 \approx |\alpha_k|^2$ and $|\tilde{o}_k\rangle \approx |o_k\rangle$ so that all that differs is a (small) change in the state of the system [14].[9]

This formally establishes the measurement problem, since now there is no definite outcome to be read out by the observer. This is, of course, equally the underlying difficulty of Schrödinger's world-famous cat-example [69], in which the evolution would proceed (symbolically) as $|\text{☢}\rangle|\text{🐈}\rangle \to \alpha\,|\text{☢}\rangle|\text{🐈}\rangle + \beta\,|\text{✷}\rangle|\;\rangle$. Suggested solutions to this problem come with vastly different implications about physical reality. Upon accepting (15) or (16) as well as (12), one is especially lead to wonder about stage 2 of the measurement process. This lead E. P. Wigner, in particular, to speculate about the impact of consciousness on physical reality [82].

Historically, the dominant solution to the MP was the denial of the possibility to treat measuring devices quantum mechanically, i.e., the denial of the above analysis of stage 1. This is an integral part of the so called *Copenhagen interpretation*, which subsumes various ideas of the founding fathers of QM, in particular of Bohr and Heisenberg. As a consequence, the interaction between observer and apparatus could (and would have to) be described entirely in the terms of everyday language and classical physics [38], whence stage 2 would not pose a special problem (for analysis, at least). But it is neither univocally clear what exactly should count as 'the' Copenhagen interpretation [40], nor what any version of it ultimately says about physical reality at the atomic scale or at the level of interaction between the classical and quantum 'realms'. In fact, David Mermin once coined a famous dictum to summarize his understanding of the Copenhagen interpretation: "Shut up and calculate!" [54]

Such an attitude may be healthy for engineers and possibly to some extent also for certain stages of the discovery of new physical theories; and it is certainly warranted to the extent that QM predicts statistical expectations perfectly well, in terms of PVMs and state vectors or

[8] Recall that any vector of the form $\sum_{j,k} \alpha_{j,k}\,|\phi_j^{(1)}\rangle|\psi_k^{(2)}\rangle$ from a tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is called *entangled* if there are for any basis at least two coefficients $\alpha_{j,k}, \alpha_{m,\ell}, (j,k) \neq (m,\ell)$ which are non-zero, i.e. if it *cannot* be written as a product state $|\phi_j^{(1)}\rangle|\psi_k^{(2)}\rangle$. That any such state can be rewritten in a "bi-orthogonal" form $\sum_{j=1}^d \alpha_j\,|\phi_j^{(1)}\rangle|\psi_j^{(2)}\rangle$ with $\{\,|\phi_j^{(1)}\rangle\}_{j=1}^{d_1}, \{\,|\psi_j^{(2)}\rangle\}_{j=1}^{d_2}$ ONBs of $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively and $d = \min\{d_1, d_2\}$, is a consequence of the *Schmidt-decomposition theorem* [36]. In infinite dimensions this generalizes to a Fock space representation, such as in the two-mode squeezed state $|\phi_\lambda\rangle = \sqrt{1-\lambda^2}\,\sum_{n=0}^{\infty} \lambda^n\,|n\rangle|n\rangle$ [39].

[9] In the worst case, the system gets destroyed though, and for these cases one merely requires the apparatus states to carry, after the interaction, information about the state of $\mathcal{S}$ *before* its destruction; i.e.: $|\phi\rangle\,|X_0\rangle \to \sum_k \alpha_k(o_j)\,|X_k(o_j)\rangle$, where in general $|X_k(o_j)\rangle \neq |o_j\rangle\,|X_k\rangle$ [79].

POVMs and density matrices. But at the same time this appears highly unscientific, at least if one subscribes to a view of science according to which it primarily strives for content-rich and *true* statements, laws, and theories [70]. It is thus understandable why many physicists have felt uncomfortable with the Copenhagen-attitude. In his 1957 thesis [24, 28], Hugh Everett III proposed an alternative view that – besides initial suspicion – enjoys much attention today, and has (apparently) even become one of the most popular interpretations of QM [68, 72]. However, this interpretation still faces foundational problems, in particular, how to adequately accommodate the probabilities predicted by Born's Rule. Below, we will first give a brief outline of the state of play of the MWI; the remainder of the thesis will confront and analyze the prospect of incorporating Born's rule into it.

2 The Many Worlds Interpretation

The central assumption of Everett's approach is that the unitary dynamics fully and adequately describes the evolution of the quantum state. There is no actual sudden and abrupt change of the system as described by (12). This, however, means that there can be no definite outcomes in the sense of (V), as required in stage 3 of the measurement process. But then, in turn, at least the *illusion* of such must arise in virtue of something else.

In fact, Everett [24] had the observer $\mathcal{O}$ end up entangled with system $\mathcal{S}$ and apparatus $\mathcal{M}$ as well, with a sequence $[\ldots]$ of respective memories indexed to $\mathcal{O}$'s state vector $|\xi_j\rangle$:[10]

$$|\psi_{\mathcal{S}*\mathcal{M}*\mathcal{O}}\rangle = \sum_j \alpha_j\, |o_j\rangle |X_j\rangle |\xi_{0,[\ldots]}\rangle \longrightarrow \sum_j \alpha_j\, |o_j\rangle |X_j\rangle |\xi_{j,[\ldots,o_j]}\rangle\,. \tag{17}$$

Thus in contrast to Wigner's or the Copenhagen proposal, stage 2 of the measurement process is here construed as a physical evolution exactly like that of stage 1. The world thus described is one unitarily evolving quantum mechanical state vector, where everything is (or becomes) entangled with everything else. The registration of one particular, definite outcome by $\mathcal{O}$ however, which, to recall, is familiar not only from experimental practice but also from primitive everyday life 'measurements', requires that the terms in this global superposition state constitute dynamically more or less independent 'branches' or 'worlds', containing different observers with different memory records. This later earned Everett's approach the name *many worlds interpretation* (MWI).

In essence, the MWI constitutes a fully deterministic version of QM which aims at solving the MP (section 1.2) by *removing* the need for *one* particular outcome where the theory seems to suggest the non-existence of such. It has been objected though, that neither do these 'worlds' appear naturally in the formalism of QM, nor is the assumption of *classical-looking* worlds always warranted. Central to these objections is the so called *preferred basis problem* (cf. [2, 66]), which results from the fact that a simple change of basis leads to a physically different situation. In a joint superposition state of a simple spin-system $\mathcal{S}$ and apparatus $\mathcal{M}$, say, where $\mathcal{M}$ has been suitably entangled with $\mathcal{S}$ to register the respective spin-values in both 'worlds',

$$|\psi_{\mathcal{S}*\mathcal{M}}\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow_z\rangle + |\downarrow_z\rangle\right)|X_0\rangle \longrightarrow \frac{1}{\sqrt{2}}\Big(|\uparrow_z\rangle|X_{\uparrow_z}\rangle + |\downarrow_z\rangle|X_{\downarrow_z}\rangle\Big), \tag{18}$$

it is no difficult task to rewrite the final state as

$$\frac{1}{\sqrt{2}}\left(|\uparrow_x\rangle|X_{\uparrow_x}\rangle + |\downarrow_x\rangle|X_{\downarrow_x}\rangle\right), \tag{19}$$

whence in any of the two worlds in the final state of (18), the spinful system could still be construed to be in a superposition w.r.t. z-spin, and $\mathcal{M}$ would be in a superposition w.r.t. results for z-spin now as well. Most importantly, any reasonable notion of 'classicality' requires the existence of states of definite (or at least: sufficiently narrow) *position*, not arbitrary superpositions thereof. As is well known though, Heisenberg discovered free narrow wave packet states to be

[10] We here introduce the otherwise unspecified symbol $*$ to refer to the combination of systems into composites. A first approach as to how to fill such a symbol with more meaning can be found in [33].

dynamically unstable and to spread enormously within seconds [37]. This and similar problems have been countered to the satisfaction of most MWI-theorists by the advent of decoherence theory. We will hence review its basic elements in the following section.

One may ask, at this point, what the reasons are to even accept an interpretation like this. It clearly introduces unobservable surplus structure into the theory, unobservable additional 'worlds' that evolve independently of ours (whatever 'ours' means in the present context). But we have seen that at least the 'spreading of superposition by entanglement', as it were, arises naturally as a consequence of applying quantum kinematics and dynamics at the level of apparatus and observers. This, in fact, is what many Everettians (adherents of the MWI) such as David Wallace think is one of the central virtues of the MWI:

> The 'Everett interpretation of quantum mechanics' is just quantum mechanics itself, 'interpreted' the same way we have always interpreted scientific theories in the past: as modelling the world. [81]

Interpreting QM "the same way we have always interpreted scientific theories" includes, for Wallace, reserving "no *fundamental* role for the notions of 'measurement' and 'observation'." (The emphasis is his.)

Indeed, it has been taken that one major concern in Everett's original thesis was that the theory be capable of yielding its own interpretation [23]. Another major concern of Everett, that is typically also mentioned as good a reasons to accept the MWI, was the application of QM to relativistic/gravitational – i.e., cosmological – problems [28]. Zurek, for instance, describes the MWI as "a natural choice for quantum cosmology, which describes the whole Universe by means of a state vector." [85] And similarly DeWitt: "Everett's view of the world is a very natural one to adopt in the quantum theory of gravity, where one is accustomed to speak without embarassment of the 'wave function of the universe.'" [22]

But for the MWI to really count as a suitable interpretation, it has to yield, as stated before, (a) at least the *impression* of definite values for dynamical quantities, in proper measurement procedures or otherwise, and (b) the empirical consequences of the standard formulation which grossly (if not exclusively) rest on Born's rule, (VI). That the empirical predictions so strongly depend on the Born rule can be easily seen: it is used in the prediction of interference patterns, averages of quantities used to confirm QM, and correlations peculiar to the QM formalism [10, 27] and not reproducible, in a satisfactory way, by any more classical theory [12, 44]. Our discussion of decoherence in the next section deals with (a); (b) will occupy the rest of the thesis.

2.1 Elements of Decoherence Theory

The foundations of Decoherence theory were discovered and outlined by H. D. Zeh in 1970 [83], and later developed further by Zurek [84], and with contributions by E. Joos, C. Kiefer, and others (cf. [46]). Decoherence essentially describes the vanishing of interference terms, the selection of a suitable preferred basis, and hence the 'emergence of classicality' due to the interaction of a system with its environment.

The general setup is as before, but this time the system is assumed to be surrounded by a non-negligible environment $\mathcal{E}$ of other systems. This is, of course, the suitable setup for almost all realistic examples, and sufficient isolation from $\mathcal{E}$, which usually first has to be prepared in a laboratory, is a crucial prerequisite for demonstrating, say, quantum interference in larger molecules [3] or similar subtle quantum phenomena. Including environment states $|\Xi_j\rangle$, the von Neumann evolution yields:

$$|\psi_{\mathcal{S}*\mathcal{M}*\mathcal{E}}\rangle = \sum_j \alpha_j\,|o_j\rangle|X_j\rangle|\Xi_0\rangle \longrightarrow |\psi'_{\mathcal{S}*\mathcal{M}*\mathcal{E}}\rangle = \sum_j \alpha_j\,|o_j\rangle|X_j\rangle|\Xi_j\rangle\,. \tag{20}$$

However, upon requiring that the environmental states are sufficiently distinguishing between different ‘pointer positions’ and system states, $\langle\Xi_j|\Xi_k\rangle \approx 0$, one now finds that the *reduced density matrix* for $\mathcal{S}*\mathcal{M}$ is of the form

$$\begin{aligned}\hat{\rho}_{\mathcal{S}*\mathcal{M}} = \mathrm{Tr}_{\mathcal{E}}(\hat{\rho}_{\mathcal{S}*\mathcal{M}*\mathcal{E}}) &= \sum_{i,j} \alpha_i\alpha_j^*\,|o_i\rangle\langle o_j| \otimes |X_i\rangle\langle X_j|\,\mathrm{Tr}\Big(|\Xi_i\rangle\langle\Xi_j|\Big) = \\ &= \sum_{i,j} \alpha_i\alpha_j^*\,|o_i\rangle\langle o_j| \otimes |X_i\rangle\langle X_j| \sum_k \langle\phi_k|\Big(\sum_{\ell,\ell'} \beta_\ell^{(i)}\beta_{\ell'}^{(j)*}\,|\phi_\ell\rangle\langle\phi_{\ell'}|\Big)|\phi_k\rangle = \\ &= \sum_{i,j} \alpha_i\alpha_j^*\,|o_i\rangle\langle o_j| \otimes |X_i\rangle\langle X_j| \sum_k \beta_k^{(i)}\beta_k^{(j)*} = \\ &= \sum_{i,j} \alpha_i\alpha_j^*\,|o_i\rangle\langle o_j| \otimes |X_i\rangle\langle X_j|\,\langle\Xi_j|\Xi_i\rangle \approx \sum_j |\alpha_j|^2\,|o_j\rangle\langle o_j| \otimes |X_j\rangle\langle X_j|\end{aligned} \tag{21}$$

where $\hat{\rho}_{\mathcal{S}*\mathcal{M}*\mathcal{E}} = |\psi'_{\mathcal{S}*\mathcal{M}*\mathcal{E}}\rangle\langle\psi'_{\mathcal{S}*\mathcal{M}*\mathcal{E}}|$, and we have used the expansion $|\Xi_j\rangle = \sum_\ell \beta_\ell^{(j)}\,|\phi_\ell\rangle$ in an orthonormal basis $\{\,|\phi_\ell\rangle\}_{\ell\in I}$ of $\mathcal{H}_{\mathcal{E}}$. Thus, upon assuming that the environmental states are sufficiently distinguishing between (joint) apparatus and system states, the quantum mechanically entangled system approximately transforms into something that looks like a statistical ensemble of systems prepared in eigenstates of the observable measured by $\mathcal{M}$, more precisely, into what is called an *improper mixture* [20].

Crucially, the preferred basis of states $|o_j\rangle$ thus established is not introduced *ad hoc*, but determined by the interaction between the systems $\mathcal{S}*\mathcal{M}$ and $\mathcal{E}$, namely as that basis of states which is robust under the interaction with $\mathcal{E}$, because all superposition (off-diagonal) terms w.r.t. the basis of pointer states $|X_j\rangle$ will be damped almost to zero in virtue of $\langle\Xi_j|\Xi_k\rangle \approx 0$. Stated differently, the preferred basis is the one in which the environmental interaction Hamiltonian takes on the desired diagonal form [46, 66], i.e., a basis of eigenstates of observables that commute with the interaction Hamiltonian [85].

Decoherence theory also implies that Heisenberg’s spreading of the wave packet vanishes, and wave packet states generally maintain a narrow localization under all realistic conditions [46, 66]. This is shown by considering a system $\mathcal{S}$ with a suitably narrow wave packet state $|\varphi\rangle = \int \mathrm{d}^3x\,\varphi(\boldsymbol{x})\,|\boldsymbol{x}\rangle$ and particles in initial quantum states $|\chi_i\rangle$ scattered off of it, constituting the environment $\mathcal{E}$ for $\mathcal{S}$. For illustrative purposes, think of a dust grain in outer space, scattering cosmic photons. This process will be described by a (unitary) scattering operator $\hat{S}$, and on each of the kets $|\boldsymbol{x}\rangle$ in the continuous superposition state $|\varphi\rangle$, this will have the effect

$$|\boldsymbol{x}\rangle|\chi_i\rangle \longrightarrow \hat{S}\,|\boldsymbol{x}\rangle|\chi_i\rangle\,. \tag{22}$$

Now using that $|\boldsymbol{x}\rangle = e^{-\frac{i}{\hbar}\boldsymbol{x}\hat{\boldsymbol{p}}}\,|\boldsymbol{x}=\boldsymbol{0}\rangle$ and that $\mathbb{1}\,|\boldsymbol{0}\rangle = e^{-\frac{i}{\hbar}\boldsymbol{x}\hat{\boldsymbol{q}}}e^{\frac{i}{\hbar}\boldsymbol{x}\hat{\boldsymbol{q}}}\,|\boldsymbol{0}\rangle$, where $\hat{\boldsymbol{q}}$ is the momentum operator for $|\chi_i\rangle$, we can rewrite this as

$$e^{-\frac{\imath}{\hbar}\boldsymbol{x}(\hat{\boldsymbol{p}}+\hat{\boldsymbol{q}})}\hat{S}\,|\boldsymbol{0}\rangle\, e^{\frac{\imath}{\hbar}\boldsymbol{x}\hat{\boldsymbol{q}}}\,|\chi_i\rangle\,, \tag{23}$$

assuming that $\left[\hat{S}, (\hat{\boldsymbol{p}} + \hat{\boldsymbol{q}})\right] = 0$, i.e., that the interaction is invariant under translations of the total system $\mathcal{S} * \mathcal{E}$. Assuming further that, due to its comparatively large mass, the dust grain's position is hardly affected by the interaction, we can rewrite this as

$$e^{-\frac{\imath}{\hbar}\boldsymbol{x}(\hat{\boldsymbol{p}}+\hat{\boldsymbol{q}})}\,|\boldsymbol{0}\rangle\,\hat{S}\,e^{\frac{\imath}{\hbar}\boldsymbol{x}\hat{\boldsymbol{q}}}\,|\chi_i\rangle = |\boldsymbol{x}\rangle\,\underbrace{e^{-\frac{\imath}{\hbar}\boldsymbol{x}\hat{\boldsymbol{q}}}\hat{S}\,e^{\frac{\imath}{\hbar}\boldsymbol{x}\hat{\boldsymbol{q}}}}_{=:\hat{S}_{\boldsymbol{x}}}\,|\chi_i\rangle \equiv |\boldsymbol{x}\rangle|\chi_{\boldsymbol{x}}\rangle\,, \tag{24}$$

meaning that the photons will carry away information about the position of the dust grain. The total density matrix right after the interaction can thus be given as

$$\hat{\rho} = \int \mathrm{d}^3\boldsymbol{x} \int \mathrm{d}^3\boldsymbol{x}'\,\underbrace{\varphi(\boldsymbol{x})\varphi^*(\boldsymbol{x}')}_{=:\rho(\boldsymbol{x},\boldsymbol{x}')}\,|\boldsymbol{x}\rangle\langle\boldsymbol{x}'| \otimes |\chi_{\boldsymbol{x}}\rangle\langle\chi_{\boldsymbol{x}'}|\,, \tag{25}$$

and by tracing out $\mathcal{E}$ again, one obtains the dependence

$$\rho(\boldsymbol{x}, \boldsymbol{x}') \longrightarrow \rho(\boldsymbol{x}, \boldsymbol{x}')\,\langle\chi_{\boldsymbol{x}'}|\chi_{\boldsymbol{x}}\rangle \tag{26}$$

which describes the influence of $\mathcal{E}$ on $\mathcal{S}$ after some scattering time τ.

The scattering operator is typically introduced [11] by considering the states $|\varphi\rangle$ to be *asymptotically free*, i.e.,

$$\lim_{t\to\pm\infty} \hat{U}(t)\,|\varphi\rangle = \lim_{t\to\pm\infty} \hat{U}_{\text{free}}(t)\,|\lambda^\pm\rangle\,, \tag{27}$$

with

$$\hat{U}(t) = \exp\!\left(-\frac{\imath}{\hbar}\hat{H}t\right), \qquad \hat{H} = \frac{\hat{\boldsymbol{p}}^2}{2m} + \hat{H}_{\text{int}}, \qquad \hat{U}_{\text{free}}(t) = \exp\!\left(-\frac{\imath}{\hbar}\frac{\hat{\boldsymbol{p}}^2}{2m}t\right) \tag{28}$$

and where $|\lambda^\pm\rangle$ are eigenstates of the free Hamiltonian $\frac{\hat{\boldsymbol{p}}^2}{2m}$. The state $|\varphi\rangle$ at $t = 0$ is related to these by

$$|\varphi\rangle = \lim_{t\to\pm\infty} \hat{U}^\dagger(t)\hat{U}_{\text{free}}(t)\,|\lambda^\pm\rangle =: \hat{\Omega}_\pm\,|\lambda^\pm\rangle\,.^{11} \tag{29}$$

With these operators, the scattering operator can be defined as $\hat{S} := \hat{\Omega}_-^\dagger\hat{\Omega}_+$, meaning that it describes the scattered state $|\varphi\rangle$ as evolving form an asymptotically free state back into an asymptotically free state after the scattering process. More precisely, if there is no interaction $\hat{H}_{\text{int}}$, we have $\hat{S} = \mathbb{1}$. This is a good reason to introduce the *transition operator*

$$\imath\hat{T} := \hat{S} - \mathbb{1} \tag{30}$$

the squares of whose matrix elements for free states describe transition probabilities (the imaginary unit is written out in front for convenience).

[11] The limits in these operators are not generally well-defined mathematically; but they can be made sense of physically nonetheless [11].

Using these insights (and a few 'tricks'), one can demonstrate [66] that

$$\langle \chi_{\boldsymbol{x}'} | \chi_{\boldsymbol{x}} \rangle = \langle \chi_i | \hat{S}^\dagger_{\boldsymbol{x}'} \hat{S}_{\boldsymbol{x}} | \chi_i \rangle = \ldots$$
$$\ldots = 1 - \int \mathrm{d}^3\boldsymbol{q}\, \mu(\boldsymbol{q}) \frac{h^3}{V} \int \mathrm{d}^3\boldsymbol{q}' \left(1 - e^{\frac{i}{\hbar}(\boldsymbol{x}-\boldsymbol{x}')(\boldsymbol{q}-\boldsymbol{q}')}\right) |\langle \boldsymbol{q} | \hat{T} | \boldsymbol{q}' \rangle|^2, \tag{31}$$

with $\int \mathrm{d}^3\boldsymbol{q}\, \mu(\boldsymbol{q}) = 1$ ($\mu(\boldsymbol{q})$ a momentum space density), and $\{|\boldsymbol{q}\rangle\}_{q\in Q}$ a basis of definite momentum states, box-normalized in V. Thus, the vanishing of interference (off-diagonal) terms in the density matrix of the post-scattering dust grain, signifying the loss of measurability of quantum superposition, crucially depends on the transition probability $|\langle \boldsymbol{q}|\hat{T}|\boldsymbol{q}'\rangle|^2$. By considering an expansion of $\hat{S}$ [11], one can motivate that $\langle \boldsymbol{q}|\hat{T}|\boldsymbol{q}'\rangle$ should take on the form

$$\langle \boldsymbol{q}|\hat{T}|\boldsymbol{q}'\rangle = \frac{i}{mh} \delta(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}}) f(\boldsymbol{q}, \boldsymbol{q}'), \tag{32}$$

where $f(\boldsymbol{q}, \boldsymbol{q}')$ is called the *scattering amplitude* (h is Planck's constant). In virtue of the appearance of $\delta^2(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}})$ in the squared modulus of (32), the evaluation of the transition probability requires (again) a little trick, namely, interpreting a mathematical limit as a very large value. This is done by appeal to the integral representation of the Dirac delta:

$$\delta^2(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}}) = \delta(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}}) \lim_{\tau\to\infty} \frac{1}{h} \int_{-\tau/2}^{\tau/2} \mathrm{d}t\, e^{i(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}})t} =$$
$$= \delta(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}}) \lim_{\tau\to\infty} \frac{1}{h} \int_{-\tau/2}^{\tau/2} \mathrm{d}t \approx \delta(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}}) \frac{\tau}{h} \qquad (\tau \gg \tau_0), \tag{33}$$

where we have introduced a typical time τ_0 required for a single scattering event as reference, and used that $\delta(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}}) = 0$ for $E_{\boldsymbol{q}'} \neq E_{\boldsymbol{q}}$. The expression thus derived can be interpreted as the transition probability growing linearly in time for sufficiently large τ.

With $E_{\boldsymbol{q}} = \frac{\boldsymbol{q}^2}{2m} = \frac{q^2}{2m}$, we can construe the remaining Dirac delta $\delta(E_{\boldsymbol{q}'} - E_{\boldsymbol{q}}) = \delta(q' - q)\frac{m}{q}$ as enforcing momentum conservation. Setting also

$$\mu(\boldsymbol{q})\, \mathrm{d}^3\boldsymbol{q} \equiv \frac{1}{N/V} \varrho(q) \frac{\mathrm{d}\Omega\, \mathrm{d}q}{4\pi}, \tag{34}$$

with $\varrho(q)$ a number density for (incoming) particles with magnitude of momentum q, N/V the total number of environmental particles in V, and $\mathrm{d}\Omega$ the differential solid angle in momentum space, we can now rewrite (31) as

$$\langle \chi_{\boldsymbol{x}'} | \chi_{\boldsymbol{x}} \rangle = 1 - \frac{\tau}{N} \int \mathrm{d}q\, \varrho(q) \frac{q}{m} \int\int \frac{\mathrm{d}\Omega\, \mathrm{d}\Omega'}{4\pi} \left(1 - e^{\frac{i}{\hbar} q(\boldsymbol{n}_q - \boldsymbol{n}_{q'})(\boldsymbol{x}-\boldsymbol{x}')}\right) |f(q\boldsymbol{n}_q, q\boldsymbol{n}_{q'})|^2. \tag{35}$$

This describes the situation for a single scattering event though, which is unjustified in virtue of the appeal to $\tau \gg \tau_0$, whence $\mu(\boldsymbol{q})$ should be renormalized to N (getting us rid of the factor $1/N$). However, recall that $\rho(\boldsymbol{x}, \boldsymbol{x}'; t = 0) \langle \chi_{\boldsymbol{x}'} | \chi_{\boldsymbol{x}} \rangle = \rho(\boldsymbol{x}, \boldsymbol{x}'; t = \tau)$. Thus,

$$\rho(\boldsymbol{x}, \boldsymbol{x}'; \tau) - \rho(\boldsymbol{x}, \boldsymbol{x}'; 0) =$$

$$= -\rho(\boldsymbol{x},\boldsymbol{x}';0)\tau \int \mathrm{d}q\,\varrho(q)\frac{q}{m}\int\int \frac{\mathrm{d}\Omega\,\mathrm{d}\Omega'}{4\pi}\left(1 - e^{\frac{i}{\hbar}q(\boldsymbol{n}_q-\boldsymbol{n}_{q'})(\boldsymbol{x}-\boldsymbol{x}')}\right)|f(q\boldsymbol{n}_q, q\boldsymbol{n}_{q'})|^2$$

$$\Leftrightarrow \frac{\rho(\boldsymbol{x},\boldsymbol{x}';\tau)-\rho(\boldsymbol{x},\boldsymbol{x}';0)}{\tau} = -F(\boldsymbol{x}-\boldsymbol{x}')\rho(\boldsymbol{x},\boldsymbol{x}';0), \tag{36}$$

where we have introduced the *decoherence factor*

$$F(\boldsymbol{x}-\boldsymbol{x}') := \int \mathrm{d}q\,\varrho(q)\frac{q}{m}\int\int \frac{\mathrm{d}\Omega\,\mathrm{d}\Omega'}{4\pi}\left(1 - e^{\frac{i}{\hbar}q(\boldsymbol{n}_q-\boldsymbol{n}_{q'})(\boldsymbol{x}-\boldsymbol{x}')}\right)|f(q\boldsymbol{n}_q, q\boldsymbol{n}_{q'})|^2. \tag{37}$$

Taking the limit $\tau \longrightarrow 0$, we obtain the DEQ

$$\frac{\partial}{\partial t}\rho(\boldsymbol{x},\boldsymbol{x}';t) = -F(\boldsymbol{x}-\boldsymbol{x}')\rho(\boldsymbol{x},\boldsymbol{x}';0), \tag{38}$$

which is solved by an exponential in $F(\boldsymbol{x}-\boldsymbol{x}')$, i.e.:

$$\rho(\boldsymbol{x},\boldsymbol{x}';t) = \rho(\boldsymbol{x},\boldsymbol{x}';0)e^{-F(\boldsymbol{x}-\boldsymbol{x}')t}. \tag{39}$$

For particular physical interactions, $F(\boldsymbol{x}-\boldsymbol{x}')$ can also be given a more concrete form. For instance, in the limit of large (de Broglie) wavelengths $\lambda \gg |\boldsymbol{x}-\boldsymbol{x}'|$ of the scattered particles, where $|\boldsymbol{x}-\boldsymbol{x}'|$ describes the distance over which the dust grain may be in a coherent superposition in spite of the interaction, $F(\boldsymbol{x}-\boldsymbol{x}')$ can be derived to be of the form $\Lambda(\boldsymbol{x}-\boldsymbol{x}')^2$, where Λ is a constant calculated in virtue of the effective total scattering cross-section of the interaction. Then $\tau_D := \frac{1}{\Lambda(\boldsymbol{x}-\boldsymbol{x}')^2}$ defines a *decoherence time*, so that off-diagonal terms vanish exponentially for times $t \gg \tau_D$ [46, 66]. These times are typically very small for all macro- or mesoscopic objects; our dust grain, for instance, would decohere within a second due to the cosmic microwave background alone, if $|\boldsymbol{x}-\boldsymbol{x}'|$ is in the order of 10 µm [66].

A crucial step in the above derivation was to take the limit $\tau \longrightarrow 0$, which may seem inappropriate, given that we have required $\tau \gg \tau_0$. But we can reasonably stipulate that we are interested in timescales much larger than τ, whence the limit is still physically meaningful. Granted, some aspects of the derivation may appear a little hand-waving, but decoherence theory has produced multiple successfully confirmed predictions, notably the works by Haroche et. al. (e.g. [13]) that were even regarded as worthy of a Nobel Prize, and the equally impressive experiments by Zeilinger et. al. [3], mentioned before.

Since decoherence suppresses interference terms for sufficiently large systems embedded in an environment, it effectively leads to a *superselection rule*. To recall, two vectors $|\psi\rangle, |\phi\rangle$ are said to be separated by a *selection rule* if $\langle\psi|\hat{H}|\phi\rangle = 0$ ($\hat{H}$ the Hamiltonian), i.e. transitions from one state to the other are inhibited by the dynamical evolution as given by $\hat{H}$. They are said to be separated by a *super*selection rule, if for *all physically realizable* observables A with operator $\hat{A}$, it holds that $\langle\psi|\hat{A}|\phi\rangle = 0$. By appeal to the density matrix, it becomes clear that this implies that no superpositions of the form $|\xi\rangle = \alpha|\psi\rangle + \beta|\phi\rangle$ exist (at least not in any observable fashion): With $\langle\psi|\hat{A}|\phi\rangle = 0$, one has $\langle\xi|\hat{A}|\xi\rangle = |\alpha|^2\langle\psi|\hat{A}|\psi\rangle + |\beta|^2\langle\phi|\hat{A}|\phi\rangle$ which would equally result from $\mathrm{Tr}(\hat{\rho}\hat{A})$ for $\hat{\rho} = |\alpha|^2|\psi\rangle\langle\psi| + |\beta|^2|\phi\rangle\langle\phi|$ a proper mixture, so that actual coherent superpositions of $|\psi\rangle$ and $|\phi\rangle$ cannot be observed [32].

Decoherence now imposes what Zurek [84] has coined an *environment-induced superselection rule* (or ‘einselection-rule’), since “the interaction with the environment forces the system to be in one of the eigenstates of the pointer observable, rather than in some arbitrary superposition of such eigenstates."

We have hence found reasons, from within QM, as to why superpositions of macroscopic objects are never observed in practice (i.e., no proper 'Schrödinger-cats'), why there is a kind of 'preferred basis', i.e., a basis of states representing more or less definite values in one of which the system will be found, and we have seen that even small objects such as dust grains end up quickly in quite definitely localized states, not in 'superpositions of different positions', i.e., not 'fully localized here and there at once'.[12]

Still, decoherence only achieves that—it does *not* effect a change of type (12), leading to *one* particular, definite outcome. There appears to be no transition from the coexistence of multiple alternatives to the existence of merely one of them, no transition from an 'and' to an 'or' [10]. In fact, the change mimics one from an arbitrary quantum state to a mixture of eigenstates of a certain observable, which was equally studied by von Neumann [77]. However, the final state is an *improper* mixture, which does not arise merely from epistemic uncertainty about the consequences of a preparation method, but rather from disregarding a part of the total system $(\mathcal{S} * \mathcal{M} * \mathcal{E})$ [20]. In summary, decoherence *alone* does not solve the MP, but it certainly aids to free the MWI, upon its assumtion, of a bunch of its problems.

In fact, decoherence straightforwardly accounts for the existence of the desired independent branches or worlds, since if interference terms between components of a global wavefunction vanish, then these can indeed be thought of as independent branches. More illuminatingly, this is sometimes explained by appeal to the *decoherent histories*-formalism [30]. To wit, one can define a time-dependent PVM $\{\hat{P}_k(t_j)\}_{k\in J}$, in virtue of the Heisenberg picture, so that every element of the PVM is of the form

$$\hat{P}^j_k \equiv \hat{P}_k(t_j) := e^{\frac{i}{\hbar}\hat{H}t_j}\hat{P}_k e^{-\frac{i}{\hbar}\hat{H}t_j}, \tag{40}$$

(we ignore possible time-dependence of $\hat{H}$ for convenience here). Then a sequence $\hat{P}^j_k\hat{P}^{j-1}_m \ldots \hat{P}^0_\ell$ of such projectors from $j+1$ respective time-dependent PVMs (with times decreasing from left to right) defines a *history operator* $\hat{C}_\alpha$ for the history α of a system described by some quantum state $|\psi\rangle$ at $t = 0$. Applying this operator to $|\psi\rangle$ leads to

$$\hat{C}_\alpha |\psi\rangle = e^{\frac{i}{\hbar}\hat{H}t_j}\hat{P}_k e^{\frac{i}{\hbar}\hat{H}(t_{j-1}-t_j)}\hat{P}_m e^{\frac{i}{\hbar}\hat{H}(t_{j-2}-t_{j-1})} \ldots e^{\frac{i}{\hbar}\hat{H}(t_0-t_1)}\hat{P}_\ell e^{-\frac{i}{\hbar}\hat{H}t_0} |\psi\rangle . \tag{41}$$

In the standard formalism, this could be interpreted as the system being successively projected (by suitable measurements) onto states $|\phi_k\rangle, |\phi_m\rangle, \ldots$, and evolving unitarily for time intervals $\Delta t_n = t_n - t_{n-1}$ in between. Two histories α,β can then be said to be *branching* in case for any given $|\psi\rangle$ and time t_j, $\hat{P}^j_k\hat{P}^{j-1}_m |\psi\rangle \neq 0$ and $\hat{P}^j_k\hat{P}^{j-1}_\ell |\psi\rangle \neq 0$ implies that $\hat{P}^{j-1}_m = \hat{P}^{j-1}_\ell$, or in words: "there is a unique way to connect projectors at later times to projectors at earlier times [...]." [81].

Effectively this means that the histories α,β described by $\hat{C}_\alpha, \hat{C}_\beta$ will agree on the past up to time t_j. The expression $D(\alpha,\beta) = \langle\psi|\hat{C}^\dagger_\alpha\hat{C}_\beta|\psi\rangle$ then defines a *decoherence functional*, and a set H of histories can be said to be *decoherent* (relative to $|\psi\rangle$) if any two (classically) incompatible

[12] Note also that a not-too-narrow Gaussian in position space of course Fourier-transforms into a comparatively narrow Gaussian in momentum space. So both, momentum and position, will be quite definite.

histories $\gamma, \delta \in H$ satisfy $D(\gamma, \delta) = 0$.[13] Crucially now, there is a theorem which, by appeal to these quantities, establishes "that branching entails decoherence and (up to possible coarse-grainings) vice versa [...]." [81] So the vanishing of interference terms leads to a structure which 'looks like', or can be interpreted as, a set of histories of the universe, evolving independently from one another, from a certain (branching-) point on.

These results are certainly impressive in a sense, and they give some credibility to the MWI. Whether decoherence *should* indeed be interpreted as sanctioning the validity of the MWI though, depends on the MWI's *total* plausibility – and that also strongly depends on its capability do deal with problem (b) from section 2, the need to reproduce the empirical predictions of the standard formulation that follow in virtue of its *statistical* content.

2.2 The Problem of Probabilities in the MWI

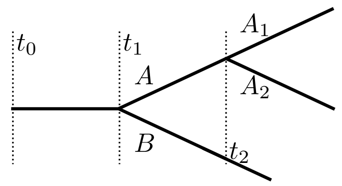

Figure 1: Branching structure with unequal distribution of an event over branches at different times.

The problems assessed in the previous section and allegedly solved by decoherence are hence not the only difficulty for the MWI. Possibly the deeper problem is that of the meaning and measure of *probability* in the MWI. Most importantly, decoherence cannot possibly aid in deriving Born's rule from general properties of QM, because partial tracing, the core ingredient of decoherence theory, is employed exactly because it is the unique function of a density matrix that preserves all statistical averages for operators pertaining only to the system not traced out, i.e., operators of the form $\hat{O}_{\mathcal{A}} \otimes \mathbb{1}_{\mathcal{B}}$ for a density matrix $\hat{\rho}_{\mathcal{A}*\mathcal{B}}$ traced over $\mathcal{B}$ [58]. Hence, partial tracing is crucially motivated by the probabilistic content of QM, or, in other words: *presuposes* the Born rule, thus making any proof of the rule that appeals to partial tracing circular.

It is clearly tempting to accuse the MWI straightforwardly of incompatibility with the Born probabilities, since each of the worlds that emerges out of the decoherence process is treated as equally 'real', i.e., as plainly existent. Then all probabilities at all times should be uniform over the worlds or branches thus defined, and hence the Born rule should be wrong in cases where the expansion coefficients are unequal. The most straightforward safeguard against this objection would be to interpret the amplitudes as representing *multiplicities*, i.e., the fractions of worlds in which a given value is realized in a branching process. But it has been found, by David Wallace in particular [81], that there is something *fundamentally* wrong with a simple counting of branches. For consider a 'branching-structure' such as in figure 1, where in all the A-worlds, one has obtained a coin, say, at t_1, and in the B world this is not so. If one identifies the probability[14] $p(\text{coin}_{t_1 \le t < t_2})$ of having a coin for the specified times with the fraction of branches

[13] In fact, one may distinguish different degrees of decoherence here, since $D(\alpha, \beta)$ is complex valued, so that one can make a difference between the whole functional vanishing or only the real or imaginary part [46]. These details do not matter for the present context though.

[14] Here and in what follows, we use p to denote an arbitrary probability assignment; the indexed Pr-symbol intro-

on which one has the coin, then this would clearly be $\frac{1}{2}$. However, in all A-worlds, the having of the coin remains constant, so that for times $t \geq t_2$, this probability becomes $\frac{2}{3}$. But by the law of total probability, one would have

$$p(\text{coin}_{t_2\leq t}) = p(\text{coin}_{t_2\leq t}|\text{coin}_{t_1\leq t<t_2}) \cdot p(\text{coin}_{t_1\leq t<t_2}) + \\ + p(\text{coin}_{t_2\leq t}|\text{no coin}_{t_1\leq t<t_2}) \cdot p(\text{no coin}_{t_1\leq t<t_2}) = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} \neq \frac{2}{3}. \quad ↯ \tag{42}$$

In principle one could attempt to solve this difficulty by introducing different probabilistic relations for branching structures such as the Everettian multiverse. The reply of certain Everettians [81], however, is that there is not even a 'natural grain' of worlds, and hence no well-defined basis for counting them at all, or, as Wallace puts it:

> the branching structure is given by decoherence, and decoherence does not deliver a structure with a well-defined notion of branch count. Very small changes in how the decoherence basis is defined, or the fineness of grain that is chosen for that basis, will lead to wild swings in the branch count. Insofar as a particular mathematical formalism for decoherence does deliver something that looks like a branch count (and many do not), that something is a mathematical artefact of no physical significance. [81]

The problem with this is elucidated even more clearly by Dawid and Thébault:

> Since there is no unique way to specify at which stage two branches have fully decoupled and therefore must be counted separately, it is impossible to specify one definitive branching structure for a quantum process. This in turn implies that no definitive probabilistic conclusions can be drawn from branch counting since the number of branches is inherently indeterminate. [18]

It is clear that a more sophisticated treatment of probabilities is required for the MWI. In fact, one could raise worries about the very *meaning* of 'probability' in the deterministic universe that the Everettian envisions (we will call this 'the conceptual problem'). It is tempting to answer this worry (and to some extent this has been done [81]) by replying that the meaning of 'probability' is *generally* unclear. Formally, there is no such problem, and the notion of probability can be accounted for by the *Kolmogorov axioms*. Probability is here defined as a function p that maps from an algebra (sometimes also called a field) $\mathcal{A}$ over a set Ω, the *samples space*, into the interval $[0, 1]$. The algebra is a collection $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ of subsets of Ω which includes the empty set $\emptyset$, Ω itself, and is closed under finite or countable unions[15] and under taking complements. The function $p : \mathcal{A} \to [0, 1]$ is then called a *probability function* or *probability measure* in case it satisfies

(i) $p(A) \geq 0, \ \forall A \in \mathcal{A}$,

(ii) $p(\Omega) = 1$,

duced before is reserved for the quantum probabilities.

[15] In the latter case one speaks of a σ-algebra [62].

(iii) $p(\bigcup_{j=1}^{N} A_j) = \sum_{j=1}^{N} p(A_j)$, $\forall A_j \in \mathcal{A}$ in case $A_j \cap A_k = 0$ for $i \neq k$.

In (iii), we have either $N \in \mathbb{N}$ or '$N = \infty$' (in the latter case, (iii) is called 'σ-additivity') [63]. This more or less covers the formal basis of probability theory. But this formal function or measure p is of course used by scientists to *represent* something; the $A \in \mathcal{A}$ are usually construed as representing *events*, and there is a whole host of differing views of what that 'something' is which p represents.

Possibly the broadest dichotomy is that between *epistemic* and *objective* probability. Objective conceptions of probability include *relative frequencies* of occurrences of certain types of events, *limits* of relative frequencies in random sequences over infinite domains, long-run *propensities* as tendencies of types of experiment to bring about a certain outcome, or *single-case* propensities as tendencies of *individual* experiments or events to bring about a certain outcome. Epistemic conceptions of probability include degrees of *belief* (of a real or ideal agent) in some hypothesis, and degrees of *confirmation* of a hypothesis [70].

But none of these views is entirely free of problems. To give an example: probabilities as limits of relative frequencies require an account of what counts as a *random* sequence, because otherwise all kinds of probabilities could be fabricated by a suitable ordering. Such an account was, indeed, worked out by von Mises [75]; but von Mises crucially had to rely on the existence of a *ground sequence* s from which general random sequences would be defined in terms of 'admissible' place selections. And it has (among other things) been objected that "there is not just only one physically possible infinite random sequence; there are many, indeed uncountably many. It seems arbitrary to pick out one of them and declare it to be the 'ground sequence.'" [70] Moreover, it is worthy of question how an ideal mathematical concept such as the limit value of some sequence connects to actual experience, or whether it even connects to reality at all in some cases [*ibid.*].

A similar review of problems could be given w.r.t. other conceptions of probability as well. But we are here concerned with probabilities in the MWI and why there is a *special* conceptual problem there. This special problem is that probability always reflects a kind of subjective *uncertainty* or objective *indeterminateness*, and *even if* we *had* a suitable notion of probability for a *classical* setting, this would not help for the MWI, since the quantum state could (in principle) be perfectly known and then there could be no uncertainty at all (nor would there be indeterminateness, of course) as to what outcome would occur, or with what frequency; *they would all certainly occur with frequency 1*.

In what follows, our primary concern will first be with the problem of connecting the numerical *values* predicted by Born's rule, and so successfully confirmed in experimental practice to the MWI (call this 'the quantitative problem'). To this end, we will consider two influential approaches to deriving the Born rule from general principles (of QM and otherwise), and without the circularity involved in decoherence-based approaches. Subsequently, the discussion will elaborate on the conceptual problem to the extent in that it is associated with both derivations.

3 Deutsch & Wallace: Probabilities from Decision Theory

A purported proof of the Born rule that has drawn much attention among philosophers of physics is that first discovered by Deutsch [21]. The strategy here is to demonstrate that rational decision makers in an Everettian multiverse will behave *as if* the Born rule was true. Deutsch's approach has subsequently been analyzed, extended, and modified by Wallace [78–81]. More precisely, the general aim of all versions is to show that the Born rule provides a *unique* measure over worlds or branches on which a decision maker, aware of QM, should base his decisions to count as rational. It remains open to debate whether Wallace's (and *a fortiori* Deutsch's) reading of decision theory is plausible after all, and can hence support the inclusion of probabilities into the MWI on decision theoretic grounds.

3.1 Decision Theory: Basic Notions & Concepts

Since all versions of the proof (Deutsch's and Wallace's) are based on decision theory, this requires us to review a few basic concepts of the latter. We will here heavily rely on the exposition of Wallace [78], where the crucial concepts established by von Neumann and Morgenstern [76], Savage [65], and others have been preselected according to their relevance to the debate.

Crucially, the setup relies on a set $\mathcal{A}$ of *acts* that can be thought of as pairs $a = (M, \pi)$, where M is a *chance setup*, a situation where any out of a given set of events might occur but it is impossible to tell which (e.g., the rolling of dice), and π is a *payoff function* that maps from a set $\mathcal{S}_M$ of *states* compatible with M (the 'possible results in M') into their *consequences* $C \supseteq \mathrm{ran}(\pi)$.[16] The events whose occurrence is uncertain are regarded as subsets of $\mathcal{S}_M$, whence the power set $\mathcal{P}(\mathcal{S}_M) := \{E | E \subseteq \mathcal{S}_M\}$ of $\mathcal{S}_M$ constitutes the *event space*. π is typically (though not necessarily) made sense of in terms of the rewards of a *bet*, so that, e.g., $\pi(\text{two sixes}) =$ reward X€ (X some precise amount). But this 'bet' may equally be any kind of situation that includes decisions under uncertainty, and the reward need not be monetary. It is also assumed that the acts can be *ordered* according to the preferences of a decision maker, so that for $a, b \in \mathcal{A}$, $a \prec b$ *iff* the agent (decision maker) would prefer b over a. This may include indifference, $\simeq$, as to certain acts, so that $\preceq$ defines a weak total order on $\mathcal{A}$ (i.e., it holds for all $a, b \in \mathcal{A}$ in some order or in both, is *transitive*: $a \preceq b \;\&\; b \preceq c \Rightarrow a \preceq c$, *reflexive*: $a \preceq a$, and *antisymmetric*: $a \preceq b \;\&\; b \preceq a \Rightarrow a \simeq b$).

A requirement for an agent to be *rational* then usually is that she should prefer an act a to an act b if (for her) the (expected) *utilities* u of these two acts satisfy $u(a) > u(b)$. Utility of course needs a definition, which may be given as

$$u(a) := \sum_{x \in \mathcal{S}_M} p(x|M)\pi(x), \qquad \text{where } a = (M, \pi). \tag{43}$$

u is construed as a measure of how useful the consequences of a are for the agent; $p(\cdot|\cdot)$ is just some conditional probability measure, or, more carefully, some 'decision weight'. So the utility of playing a game of dice, say, is evaluated by considering how probable any given result of a throw is, and what the consequences of the occurrence of the particular results are

[16] We here take it that π is real valued, for convenience, i.e., that $C \subseteq \mathbb{R}$. But one could also introduce an additional map that assigns real values to the consequences [78].

(money received, joy of winning, loss of money...). The average of these consequence-values or 'payoffs' w.r.t. the probability measure then yields a utility value for the game. But of course the theory should also be transferable to situations where the reward is not money, and the situation is more than just a game (cf. [78] for examples).

In this definition, however, there are two 'variables' to be filled, namely the occurring probability measure, and the utility function itself. However, it makes sense to connect the two by *defining either through the other*. This is sometimes accomplished by considerations of betting behavior, so that p is supposed to reflect a *degree of belief* or *confidence* in a given outcome. But one might also take the definition of u as a direct indicator as to the role of p. What the Deutsch-Wallace proof provides, in the version that we are considering here, is a *uniqueness theorem* that p is the Born probability if (M, π) is a 'quantum game' (defined below), given some decision theoretic axioms effectively constraining u.[17] It is thereby shown that rational agents competent in QM *behave as if* the Born rule was true. This is viewed, in turn, as a justification of the Born rule that is compatible with the MWI.

3.2 The Deutsch-Wallace Proof

To derive the Born rule with the aid of decision theoretic principles, it is of course necessary to introduce QM notions into the language of decision theory. In fact, the chance setup M can be identified with the measurement of some observable O with operator $\hat{O}$, and $\mathcal{S}_M$ with the set of states of the measuring apparatus, i.e., pointer positions, light bulb configurations, ... and so forth, which can be mapped onto (identified with) eigenvalues o_j of $\hat{O}$. The measurement is performed on some quantum state $|\psi\rangle$ and the o_j will in turn be mapped onto consequences (payoffs) by a suitable π. Thus the ordered triple $(|\psi\rangle, \hat{O}, \pi)$ formally defines a *quantum game g*, the analogue of an act. Of course such games are abstract formal entities that may be implemented physically in various ways (for instance, spin experiments can typically be used in analogy to polarization experiment). Agents will be considered *indifferent* between two different physical realizations of the same triple $(|\psi\rangle, \hat{O}, \pi)$, as is postulated as an axiom by Wallace [78].

Indifference, $\simeq$, in fact defines an *equivalence relation*, as can be easily seen from the properties of $\preceq$ as introduced above. Using this insight, Wallace [78, 79] proves the following two equivalences:[18]

Theorem 3.1 (*Equivalence theorem*)**.**

1. $(|\psi\rangle, \hat{O}, \pi) \simeq (|\psi\rangle, f(\hat{O}), \pi \circ f^{-1})$ ('payoff equivalence'; PE),
   where f maps the eigenvalues $\sigma(\hat{O})$ of $\hat{O}$ into $\mathbb{R}$.

2. $(|\psi\rangle, \hat{O}, \pi) \simeq (\hat{U}|\psi\rangle, \hat{O}', \pi')$ ('measurement equivalence'; ME),
   where $\hat{U}$ is a unitary operator, $\hat{U}\hat{O} = \hat{O}'\hat{U}$, and π and π' give the same payoffs for the

[17] Wallace [78] indeed argues that it is "foundationally [...] far more satisfactory" to first derive the uniqueness of p and then fix u. He has also presented various alternative ways of deriving the Born rule from different sets of decision theoretic axioms [79–81]. We will focus on the direction $u \Rightarrow p$ and a version of the proof as originally presented by Deutsch [21], since this is the most straightforward way and fully sufficient for our discussion.

[18] Wallace [79] actually proves three equivalences, but we do not need the third one, and it implies a questionable appeal to 'Born-weights'.

eigenvalues $\sigma(\hat{O})$ and $\sigma(\hat{O}')$ of $\hat{O}$ and $\hat{O}'$ respectively.[19]

The proofs of 1. and 2. both essentially boil down to analyzing the physical processes involved, and we will omit full proofs for either of them (see [79] for further reference). To briefly sketch the ideas: In 1. it is exploited that $f(\hat{O})$ maps the eigenvalues of $\hat{O}$ to some interval of $\mathbb{R}$, and if the payoff is 'shifted back' accordingly ($\pi \mapsto \pi \circ f^{-1}$) then this is equal to a simple *relabeling* of the outcome states, without any physical change. In 2., the two states $|\psi\rangle$, $\hat{U}|\psi\rangle$ are construed to be elements of different spaces $\mathcal{H}, \mathcal{H}'$ repsectively, and it is demonstrated that one can repeat the same physical process (with identical payoffs) utilizing states from either $\mathcal{H}$ or $\mathcal{H}'$ as states of an auxiliary system (i.e., using $\mathcal{H}\otimes\mathcal{H}'$), so that neither is physically preferred (or rather: both can be viewed as physically identical), in virtue of the possibility of reordering the tensor product or a mere relabeling. (□)

Crucial to the actual proof of the theorem is also the *composition* of games into more complex games. This is done [21, 79] by letting consequences of games be further games, so that, "[f]or instance, we might measure the spin of a spin-half particle, and play one of two possible games according to which spin we obtained." [79]

We mentioned at the conclusion of the last section that the utility function u needs to be suitably constrained in order to derive the Born rule (prove its uniqueness). The constraints that are used are the following:[20]

Dominance (DOM)**.** If for all $o \in \sigma(\hat{O}), \pi(o) \geq \pi'(o)$ then $u(|\psi\rangle, \hat{O}, \pi) \geq u(|\psi\rangle, \hat{O}, \pi')$.

Substitutivity (SUB)**.** If g_{comp} is a compound game formed from g by substituting some of g's payoffs $c_j, \ldots c_{j+k}$ by games $g_j \ldots g_{j+k}$ such that $u(g_i) = c_i$, $j \leq i \leq j+k$, then $u(g_{\text{comp}}) = u(g)$.

Additivity (ADD)**.** For $k \in \mathbb{R}$ it holds that $u(|\psi\rangle, \hat{O}, \pi + k) = u(|\psi\rangle, \hat{O}, \pi) + k$.

Zero-Sum (ZS)**.** For some given π, let $\bar{\pi}(o) = -\pi(o), \forall o \in \text{dom}(\pi)$. Then $u(|\psi\rangle, \hat{O}, \bar{\pi}) = -u(|\psi\rangle, \hat{O}, \pi)$.

DOM and ZS should be easily acceptable. Of course a game with greater payoff is more useful to an agent, where we remember that 'payoff' should not be taken to literal (in the sense of money), and hence may depend on the specific agent to the same degree as does utility. And zero-sum basically defines negative utility; so if an agent looses something she considers valuable (the payoff), then the act leading to the loss will do her harm/have negative use for her. To SUB one might object that, in real life, the playing of the subsequent games $g_j \ldots$ comes with extra effort for the agent (instead of receiving the payoffs directly), so substitutivity is at least not 'self evident'. But of course, the payoff may be so adapted that it compensates the additional cost; e.g., if one plays a game of dice and instead of receiving a reward of 5€, one is offered to play a card game where the potential reward is 10€ instead, this might be considered equally worthwhile. ADD basically says that a certain additional payoff to a given game is as useful as another game in which the payoff is always higher. This requires, of course, that payoffs and

[19] Note that $\hat{U}$ may connect different Hilbert spaces, and in case it maps a Hilbert space onto itself, the condition simplifies to $(|\psi\rangle, \hat{O}, \pi) \simeq (\hat{U}|\psi\rangle, \hat{U}\hat{O}\hat{U}^\dagger, \pi')$ [79].

[20] We again loosely follow [79] with terminology and formulation.

utilities can be straightforwardly compared or equated; but this is plausible, since utilities are computed as a weighted sum of payoffs, so they should be quantified in a comparable fashion.[21]

There is no use in fixating on these assumptions in particular though, since (as mentioned before) Wallace has offered proofs from different (arguably weaker) assumptions. We shall hence later assess more general concerns about the decision theoretic approach. Assuming also that two equivalent games $g \simeq g'$ have the same utility (Wallace calls this 'physicality'), which constitutes basically a conceptual/definitional link between '$\simeq$' and 'u', one can now prove the following theorem:

Theorem 3.2. For a value function u which is additive, substitutive, and satisfies dominance and zero-sum and a quantum game $g = (|\psi\rangle, \hat{O}, \pi)$, $u(g)$ is uniquely given by

$$u(g) = \sum_{o \in \sigma(\hat{O})} \langle\psi|\hat{P}_o|\psi\rangle\, \pi(o) = \sum_{o \in \sigma(\hat{O})} \mathrm{Pr}^{\psi}_{O}(o)\pi(o).$$

Proof. As a gross simplification, we generically set $\pi \equiv \mathrm{id}_{\hat{O}}$, i.e., the payoff is the identity map on $\sigma(\hat{O})$. This is possible without loss of generality, since by PE, we can map this situation onto any other by a mere 'relabeling'. The proof can then be given in six stages (i-vi) [79] (we mark the conclusion of each stage by '$\diamond$').

(i) Consider a game where $|\psi\rangle = \frac{1}{\sqrt{2}}(|o_1\rangle + |o_2\rangle)$. Then from ADD we have

$$u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) + k = u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}} + k).$$

Defining $f(o) = o + k$, we have $f^{-1}(\tilde{o}) = \tilde{o} - k$ ($\tilde{o} \in \mathrm{ran}(f)$), and according to PE (and 'physicality'), we thus get

$$u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) + k = u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}} + k) = u(|\psi\rangle, \hat{O} + k, \mathrm{id}_{\hat{O}}),$$

where $\hat{O} + k$ is understood as defined on the values in $\sigma(\hat{O})$. Using PE in the same fashion in combination with ZS yields

$$u(|\psi\rangle, -\hat{O}, \mathrm{id}_{\hat{O}}) = u(|\psi\rangle, \hat{O}, \bar{\mathrm{id}}_{\hat{O}}) = -u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}).$$

Thus overall we have

$$u(|\psi\rangle, -\hat{O} + k, \mathrm{id}_{\hat{O}}) = -u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) + k. \tag{$\circ$}$$

Now let $f(o) := -o + o_1 + o_2$ and distinguish the following cases:

(a) $\hat{O}$ is non-degenerate and $\sigma(\hat{O})$ is invariant under f (e.g., $\sigma(\hat{O}) = \{\pm\frac{1}{2}\}$). Then there is a $\hat{U}_f$ such that $\hat{U}_f\hat{O}\hat{U}_f^\dagger = f(\hat{O})$ (since f is invertible and $\hat{O}$ is not 'rescaled'), and $|\psi\rangle = \frac{1}{\sqrt{2}}(|o_1\rangle + |o_2\rangle)$ is left invariant by $\hat{U}_f$. But then by ME we have

$$u(|\psi\rangle, -\hat{O} + o_1 + o_2, \mathrm{id}_{\hat{O}}) = u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}).$$

[21] In [21], this version of the additivity is proved from an additivity assumption on the payoffs. No harm comes from assuming the additivtiy for u though.

Thus, using (∘) with $k = o_1 + o_2$, we have

$$u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) = -u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) + o_1 + o_2 \Leftrightarrow u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) = \\ = \frac{1}{2}(o_1 + o_2) = \sum_{o \in \sigma(\hat{O})} \langle\psi|\hat{P}_o|\psi\rangle\, \pi(o).$$

(b) $\hat{O}$ is degenerate or $\sigma(\hat{O})$ is non-invariant under f. Then by ME one obtains an equivalent game through some transformation $\hat{O} \mapsto \hat{U}\hat{O}\hat{U}^\dagger =: \hat{O}'$, s.t. $\hat{O}'$ agrees with $\hat{O}$ on the span of $\{|o_1\rangle, |o_2\rangle\}$ and is $\frac{1}{2}(o_1+o_2)\mathbb{1}$ otherwise. The result then follows analogously for the equivalent game (except if $o_1 = o_2$, in which case the result follows trivially). ⋄

(ii) Let $N = 2^n, n \in \mathbb{N} \setminus \{0\}$, $|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^N |o_j\rangle$. We prove by induction on $n \geq 2$ ($n = 1$ is covered by (i)) that $u(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) = \frac{1}{N}(\sum_{j=1}^N o_j)$.

- *Base step*: for $n = 2$, we have $|\psi\rangle = \frac{1}{2}\sum_{j=1}^4 |o_j\rangle$. Now define

$$|A\rangle := \frac{1}{\sqrt{2}}(|o_1\rangle + |o_2\rangle), \quad |B\rangle := \frac{1}{\sqrt{2}}(|o_3\rangle + |o_4\rangle), \quad \hat{Q} := \frac{o_1+o_2}{2}|A\rangle\langle A| + \frac{o_3+o_4}{2}|B\rangle\langle B|.$$

Then $|\psi\rangle = \frac{1}{\sqrt{2}}(|A\rangle + |B\rangle)$, and $g_{\hat{Q}} := (|\psi\rangle, \hat{Q}, \mathrm{id}_{\hat{Q}})$ can easily be seen to have value $\frac{1}{4}\sum_{j=1}^4 o_j$ by (i). We now distinguish two cases:

(a) Either the observer finds himself in the A-branch, and receives $\frac{o_1+o_2}{2}$. By SUB, the observer is now indifferent between immediately receiving $\frac{o_1+o_2}{2}$ and playing $g_A := (|A\rangle, \hat{O}, \mathrm{id}_{\hat{O}})$, since by (i) and definition of $|A\rangle$, $u(g_A) = \frac{1}{2}o_1 + o_2$ which is the payoff.

(b) The same applies in the B-branch with $g_B := (|B\rangle, \hat{O}, \mathrm{id}_{\hat{O}})$ and $\frac{o_3+o_4}{2}$.

The total value of playing $g_{\hat{Q}}$ with g_A and g_B substituted for payoffs is thus equally $\frac{1}{4}\sum_{j=1}^4 o_j$. But this game is physically realized by the same process as $(|\psi\rangle, \hat{O}, \mathrm{id}_{\hat{O}})$.

- *Induction hypothesis*: Let, for some n, $u(|\psi_n\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) = \frac{1}{2^n}(\sum_{j=1}^{2^n} o_j)$, where $|\psi_n\rangle = \frac{1}{\sqrt{2^n}}\sum_{j=1}^{2^n}|o_j\rangle$.

- *Inductive step*: We repeat the procedure of the base step for $|\psi_{n+1}\rangle = \frac{1}{\sqrt{2^{n+1}}}\sum_{j=1}^{2^{n+1}}|o_j\rangle$, by defining a suitable observable $\hat{Q}_{n+1} = \frac{\sum_{j=1}^{2^n} o_j}{2^n}|A\rangle\langle A| + \frac{\sum_{j=2^n+1}^{2^{n+1}} o_j}{2^n}|B\rangle\langle B|$ with

$$|A\rangle := \frac{1}{\sqrt{2^n}}\sum_{j=1}^{2^n}|o_j\rangle, \quad |B\rangle := \frac{1}{\sqrt{2^n}}\sum_{j=2^n+1}^{2^{n+1}}|o_j\rangle, \quad |\psi_{n+1}\rangle = \frac{1}{\sqrt{2}}(|A\rangle + |B\rangle),$$

and using the i.h. and (i) and SUB again. ⋄

(iii) Now let $N = 2^n, n, a_1, a_2 \in \mathbb{N}\setminus\{0\}$, where $a_1 + a_2 = N$, and $|\psi\rangle = \frac{1}{\sqrt{N}}(\sqrt{a_1}|o_1\rangle + \sqrt{a_2}|o_2\rangle)$.

Then define $\hat{Q} := \sum_{j=1}^{N} q_j \left|q_j\right\rangle\!\left\langle q_j\right|$ over an N-dimensional Hilbert space $\mathcal{H}_N$, and

$$f(q_j) = \begin{cases} o_1 \text{ if } 1 \leq j \leq a_1, \\ o_2 \text{ else.} \end{cases}$$

Also, let (without loss of generality) $\mathrm{dom}(\hat{O}) = \mathrm{span}\{\left|o_1\right\rangle, \left|o_2\right\rangle\}$ and $\hat{V} : \mathrm{dom}(\hat{O}) \to \mathcal{H}_N$ be defined s. t.

$$\hat{V}\left|o_1\right\rangle = \frac{1}{\sqrt{a_1}} \sum_{j=1}^{a_1} \left|q_j\right\rangle, \quad \hat{V}\left|o_2\right\rangle = \frac{1}{\sqrt{a_2}} \sum_{j=a_1+1}^{a_2} \left|q_j\right\rangle.$$

Then we have $f(\hat{Q})\hat{V} = \hat{V}\hat{O}$, and from PE, ME, we get $(\left|\psi\right\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) \simeq (\hat{V}\left|\psi\right\rangle, f(\hat{Q}), \mathrm{id}_{f(\hat{Q})}) \simeq (\hat{V}\left|\psi\right\rangle, \hat{Q}, f \circ \mathrm{id}_{\hat{Q}})$. Since $\hat{V}\left|\psi\right\rangle$ is an even superposition and we receive $a_1 \cdot o_1$ and $a_2 \cdot o_2$ in the last game, by definition of f, this yields $u(\left|\psi\right\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) = \frac{1}{N}(a_1 o_1 + a_2 o_2)$. ⋄

(iv) Next, we let $1 > a \in \mathbb{R}^+$, $\left|\psi\right\rangle = \sqrt{a}\left|o_1\right\rangle + \sqrt{1-a}\left|o_2\right\rangle$. Let also $a = \lim_{n\to\infty} a_n$ for decreasing $a_n =: \frac{A_n}{2^n}, A_n \in \mathbb{N} \setminus \{0\}, \forall n \in \mathbb{N}$ (i.e., we approximate a by a series of numbers that are dense in $\mathbb{R}^+$). We then define, for each n,

$$\left|\psi_n\right\rangle := \sqrt{a_n}\left|o_1\right\rangle + \sqrt{1-a_n}\left|o_2\right\rangle, \quad \left|\phi_n\right\rangle := \sqrt{\frac{a}{a_n}}\left|o_1\right\rangle + \sqrt{1-\frac{a}{a_n}}\left|o_2\right\rangle,$$

and $g_n := (\left|\psi_n\right\rangle, \hat{O}, \mathrm{id}_{\hat{O}})$, $g'_n := (\left|\phi_n\right\rangle, \hat{O}, \mathrm{id}_{\hat{O}})$; g is then just the same game on $\left|\psi\right\rangle$. Since $a_n \in \mathbb{Q}, \forall n \in \mathbb{N}$, we get from (iii) that $u(g_n) = a_n o_1 + (1-a_n) o_2$. Without loss of generality we set $o_1 \leq o_2$, so that by DOM we can say that $u(g'_n) \geq o_1$. Call g_n^- the game whose utility is exactly o_1. Then by SUB we receive a game equivalent to g_n by substituting g_n^- for o_1 in g_n and we then know by DOM that utility of the game resulting from substituting g'_n instead is at least $u(g_n)$. But substituting $\left|\phi_n\right\rangle$ for $\left|o_1\right\rangle$ in $\left|\psi_n\right\rangle$, we obtain $\sqrt{a}\left|o_1\right\rangle + (\sqrt{a_n - a} + \sqrt{1-a_n})\left|o_2\right\rangle$ which, in the $n \longrightarrow \infty$-limit is just $\left|\psi\right\rangle$. So we can see that $u(g) \geq u(g_n), \forall n \in \mathbb{N}$, that is, $u(g) \geq a o_1 + (1-a) o_2$. Through an analogous argument on an increasing series of a_n, we can establish that $u(g) \leq a o_1 + (1-a) o_2$, whence $u(g) = a o_1 + (1-a) o_2$. ⋄

(v) For $\alpha, \beta \in \mathbb{C}$, so that $|\alpha|^2 + |\beta|^2 = 1$, and $\left|\psi\right\rangle = \alpha\left|o_1\right\rangle + \beta\left|o_2\right\rangle$, we write $\alpha = |\alpha| e^{\imath\varphi_1}, \beta = |\beta| e^{\imath\varphi_2}$, and transform $\hat{O}$ and $\left|\psi\right\rangle$ by $\hat{U} = e^{-\imath\varphi_1}\left|o_1\right\rangle\!\left\langle o_1\right| + e^{-\imath\varphi_2}\left|o_2\right\rangle\!\left\langle o_2\right|$. Then $\hat{U}\left|\psi\right\rangle = |\alpha|\left|o_1\right\rangle + |\beta|\left|o_2\right\rangle$, and $\hat{U}\hat{O}\hat{U}^\dagger = \hat{O}$. But then by ME, we have $g := (\left|\psi\right\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) \simeq (\hat{U}\left|\psi\right\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) =: \tilde{g}$, and $u(\tilde{g})$ is just $|\alpha|^2 o_1 + |\beta|^2 o_2$ by (iv). So $u(g) = |\alpha|^2 o_1 + |\beta|^2 o_2$. ⋄

(vi) We can generalize (v) to $\left|\psi\right\rangle = \sum_j \alpha_j \left|o_j\right\rangle$ and $u(\left|\psi\right\rangle, \hat{O}, \mathrm{id}_{\hat{O}}) = \sum_j |\alpha_j|^2 o_j$, in analogy to the generalization from (i) to (ii). □

3.3 Discussion (i)

The proof presented above is rather seamless, and we take it that this is a deductively valid argument from the decision-theoretic premises to its conclusion. The same has been judged (by critics [53]) about Wallace's other version(s) of the proof. But that an argument is deductively valid entails nothing about the truth of its premises. That is to say, if one finds the conclusion of the theorem to be highly implausible for good reasons, this merely implies that at least one of the premises from which the conclusion follows must have been wrong all along. As philosophers sometimes put it: one's *modus ponens* is another's *modus tollens*.

So how reasonable is it that believers in the MWI will apply the Born rule to guide their actions? Harsh criticism has in fact been advanced against such a conclusion. The first thing to notice here (which also lies at the heart of the criticism) is the rather intricate connection between the different 'observers' in a branching process. For if an observer $\mathcal{O}$ makes an x-spin measurement on a system prepared as $|\uparrow_z\rangle$, then she will end up, according to the MWI, entangled with the state of the spin. The precise form of the entangled state will be determined by the dynamical properties of the interaction (by decoherence), but of course the subjective experience of $\mathcal{O}$ will inevitably be such that she either experiences measuring $|\uparrow_x\rangle$ or $|\downarrow_x\rangle$, as is also assumed by Deutsch and Wallace. However, there will then be *two* observers that emerge out of this entangling process, which we will call $\mathcal{O}_{\uparrow_x}$ and $\mathcal{O}_{\downarrow_x}$. Then *which* of the two observers *is* $\mathcal{O}$ after the measurement-interaction? Which is the 'rightful heir' of $\mathcal{O}$'s conscious memory?

The answer clearly must be either 'none' or 'both', because there is no sense in which either $\mathcal{O}_{\uparrow_x}$ or $\mathcal{O}_{\downarrow_x}$ is preferred (in the QM formalism at least).[22] This ultimately means that there is no clear sense in which any of the two can count as the successor of $\mathcal{O}$, and that hence $\mathcal{O}$ has no good reason to identify her wishes and desires with those of either $\mathcal{O}_{\uparrow_x}$ or $\mathcal{O}_{\downarrow_x}$ [7].

The problem that is dawning here is that $\mathcal{O}$ cannot even evaluate the utility of measuring either $|\uparrow_x\rangle$ or $|\downarrow_x\rangle$, because $\mathcal{O}$ will not measure anything; $\mathcal{O}_{\uparrow_x}$ and $\mathcal{O}_{\downarrow_x}$ will, and there is no clear sense in which either of them is 'really' $\mathcal{O}$. More precisely, $\mathcal{O}$ cannot have any *uncertainty* about whether she will become $\mathcal{O}_{\uparrow_x}$ or $\mathcal{O}_{\downarrow_x}$ (because she will become both/neither), so the squared amplitude must be reinterpreted in some way, to make sense in the decision theoretic context.[23] It has been suggested by L. Vaidman [74] and H. Greaves [34, 35] to interpret the Born rule-decision weights as a *caring measure*, a measure of how much one cares about what happens to any of one's future selves.

Greaves [35] has argued that this is even feasible for the context of confirmation of scientific theories, since a caring measure satisfies, by some standards, the same requirements for confirmation that probability does. In particular, Greaves appeals to *Bayesian* confirmation theory, according to which *rational* agents must update their beliefs in the truth of a hypothesis H in the form of probabilistic *conditionalizing* on new evidence. Thus, say a hypothesis H is believed to some degree $p(H)$ prior to learning evidence E. This may be understood in terms of betting behavior, as was indicated in section 3.1. Then if an agent updates her believes at all, but does not update them by conditionalizing, i.e., according to

$$p_E(H) = p(H|E) = \frac{p(E|H)}{p(E)}p(H),^{24} \tag{44}$$

[22] An alternative is to associate a *multitude* of 'minds' or consciousnesses with any system that is physically capable of carrying consciousness at all (complex biological organisms, according to current knowledge), which also opens up an alternative way of associating probabilities with measurement outcomes [2]. But then one still needs advanced dynamics for the interrelations between the successive multitudes of consciousnesses that are spawned off in measurement-like situations [50].

[23] For a diverging opinion see [64]. The arguments presented therein have, however, been countered by Greaves [34] in a way that we here hold to be satisfactory.

[24] This is a version of *Bayes' theorem*, which follows straightforwardly from the definition of conditional proba-

then a 'diachronic Dutch book' can be made against her, i.e., one can devise a series of bets that she is certain to loose, in virtue of her updating rule diverging from (44).

Greaves then argues that the Born rule-caring measure can do the same in a branching multiverse: it offers a unique means for determining betting quotients at which one would be willing to accept a certain bet without the possibility of a 'Dutch book', with the prospect now being sure losses to *all* of ones future selves, on all branches. I.e., the 'Born probability' quantifies to what degree $\mathcal{O}$ cares about the gains and losses of $\mathcal{O}_{\uparrow_x}$ and $\mathcal{O}_{\downarrow_x}$ in certain bets whose payoffs depend on observing either $|\uparrow_x\rangle$ or $|\downarrow_x\rangle$. And it does so in the unique way that avoids sure losses to all future selves ($\mathcal{O}_{\uparrow_x}$ and $\mathcal{O}_{\downarrow_x}$).

But regardless of whether Greaves's analysis is correct, it has been argued that agents could – and indeed should – equally well care about their future selves in fashion that defies the Born rule as a suitable 'caring measure', in the multiverse. The alternative caring measures suggested are typically intended as a *reductio ad absurdum*, i.e., as a mere counterexample, demonstrating that rationality does clearly not dictate the Born probabilities (and therefore, at least one of the decision-theoretic assumptions must be wrong, or rather: does not reflect rationality). Albert [1], for instance, has argued that one could equally use a 'fatness measure' to quantify the future well-being of the future versions of one's self, i.e., that one could care more greatly about branches in which there is 'more of oneself'. A measure that takes notice of that could be established – in a fashion that avoids quantifying how many future selves exactly there are, i.e., avoids the need for a problematic *counting* strategy – by multiplying the Born weights with a measure of one's own 'fatness' (e.g. ones *mass* [81]). As Albert argues, it does not seem irrational to care more about those branches where there is 'more of oneself'.

It is doubtworthy that caring more about more mass is indeed rational. But one need not even invoke such cynical means as the 'fatness measure' to see that something must be wrong with the decision-theoretic approach. A key issue that points in this direction is that the Everettian, aware of the branching, could also well care about some outcome of a game that concerns *all of the future selves*. Such a possibility is discussed by Wallace [81] and Maudlin [53], but with unequal conclusions. Let us imagine [53, 81] a student who would really like to study physics *and* history, but only has the (financial, cognitive) capacities to study either of the two 'in any one branch'. Let us also say that she is a firm Everettian and has reasons to settle for physics, but that she meets an experimental physicist who offers her (for 10€, say) to make an x-spin measurement on $|\uparrow_z\rangle$, based on whose outcome she can reconsider (i.e.: stick to physics if $|\uparrow_x\rangle$ is observed, and switch to history if $|\downarrow_x\rangle$ is observed). From an Everettian standpoint, it seems perfectly rational for the student to accept this bargain, since she will become $\mathcal{O}_{\uparrow_x}$ *and* $\mathcal{O}_{\downarrow_x}$, at least in two separate branches, and hence come as close as it gets to fulfilling her dream of studying both subjects. Granted, $\mathcal{O}_{\uparrow_x}$ and $\mathcal{O}_{\downarrow_x}$ do not share a joint consciousness, so this is not exactly the situation desired. But we have argued that there is no clear reason for the student before the measurement to identify with either $\mathcal{O}_{\uparrow_x}$ or $\mathcal{O}_{\downarrow_x}$ on the MWI. Thus in virtue of her Everettian beliefs, this may indeed count as 'as close as it gets'.

But if she would indeed take this deal, then she would *not* use the branch weights or the Born rule-'caring measure' to base her decision on. Thus there are good reasons to believe that the behavior of rational Everettians does *not* always conform to the Born rule. It is also clear that the student would be grossly irrational in a non-Everettian universe, since she would either pay

bility as $p(X|Y) = p(X,Y)/p(Y)$ [70].

10€ for reaffirmation of a decision she has already made, or simply loose 10€ to decide for something she had deemed slightly less desirable.

The problems raised here become most pressing when transferred to matters of life an death. For consider a malfunctioning nuclear power plant, and the staff faced with two options: (a) send 12 workers to certain death to prevent a meltdown or (b) perform a procedure that depends on a quantum mechanical process with a Born probability of 0.9999 of avoiding disaster and 0.0001 of sending everyone (1000 people, say) to death. If one is not an Everettian, it seems perfectly reasonable to rely on the Born rule and chose (b). But for Everettians, option (a) indeed seems much more reasonable, because upon choosing (b), there *will be* branches that include the deaths of 1000 people [53].

Assuming that it is the case that the MWI is true, we may have to adapt our intuitions about such moral dilemmas, and find actions to be reasonable that we would not have found reasonable before. But that does not change the fact that there are cases where it seems decisively rational for Everettians not to rely on the Born rule to guide their actions (as the student-example straightforwardly shows). So one of the premises of the Deutsch-Wallace proof must clearly be false.

Maudlin [53], indeed, has also put the finger on the relevant premise, a prerequisite not made explicit but used in multiple instances in the above proof, and stated explicitly in Wallace's advanced version of it [81]. Namely, outcomes are still treated as *mutually exclusive* in the whole decision-theoretic approach to deriving the Born rule. But there is a clear sense in which this assumption is not warranted in the MWI anymore, whence, as Maudlin has piercingly put it: "Wallace's [and Deutsch's – FB] decision theory has been set up *in its axioms* to rule out a rational agent acting in a way that takes Everettian quantum mechanics seriously." (The emphasis is Maudlin's.)

For completeness's sake, we note that Wallace [81] has responded to the problem posed by the student-example by invoking a principle he calls 'diachronic consistency', according to which (roughly) an agent will not suddenly change his preferences after a branching event, i.e., maintain consistent values for acts over time. Wallace now argues that choosing to 'split' can violate diachronic consistency since the version of the student who would end up studying history could always switch back without hurting his 'counterpart' in the other branch. But it is hard to see how this response applies, since if the goal is to have two versions of oneself, one who studies history and one who studies physics, then this would clearly forestall switching back [53].

A second reply by Wallace is that this is not even a counterexample because the decision-problem is now a different one, and, in case one does not switch back, "the utility you are assigning to (history-after-process) is higher than the utility you assign to (physics-against-process), and indeed higher than (physics-without-process). The different situations in which you end up doing history count as different rewards." [81] Against this, Maudlin has (concisely) countered that it misses the point, since "[i]t is the *outcome*, not the *process*, that matters. In the relevant sense, a 'split life' is a *different* outcome than either non-split life would be, and the agent simply prefers this outcome and would pay a premium for it. Such behavior is a 'counterexample to the Born rule', i.e. to using the branch amplitudes as weights in decision theory for calculating an expected utility." [53] (The emphasis is Maudlin's.)

4 Zurek: Probabilities from QM Alone?

The forgoing discussion has highlighted difficulties of the decision-theoretic approach to justifying (or formally deriving) the Born rule. An intriguing alternative was opened up by Zurek [86], who found a method of deriving the Born rule from the remainder of QM and without direct appeal to decision theory or the like. However, close scrutiny of the actual proof shows that Zurek also has to assume more than meets the eye, in particular about probabilities and how they enter into the formalism. Moreover, no clear view of how to interpret the probabilities is offered, and this, in fact, again raises potential trouble for the MWI. Below, we will first analyze the derivation itself in the detailed form presented in [87], and then discuss objections to the approach and its purported implications.

4.1 Introducing Envariance

Zurek's proof crucially rest on a feature he calls 'environment-assisted invariance', or short: *envariance*, a feature of entangled states that is defined as follows [86].[25]

Definition 4.1 (*Envariance*)**.** Let $|\psi_{\mathcal{S}*\mathcal{E}}\rangle \in \mathcal{H} = \mathcal{H}_{\mathcal{S}} \otimes \mathcal{H}_{\mathcal{E}}$ be a joint state of a system of interest $\mathcal{S}$ and its environment $\mathcal{E}$, and let $\hat{U}_{\mathcal{S}} = \hat{u}_{\mathcal{S}} \otimes \mathbb{1}_{\mathcal{E}}$ be a unitary operator acting solely on the space $\mathcal{H}_{\mathcal{S}}$. Then $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ is called *envariant* under $\hat{U}_{\mathcal{S}}$ *iff* there is an $\hat{U}_{\mathcal{E}} = \mathbb{1}_{\mathcal{S}} \otimes \hat{u}_{\mathcal{E}}$, acting solely on $\mathcal{H}_{\mathcal{E}}$, s. t.

$$\hat{U}_{\mathcal{E}}\hat{U}_{\mathcal{S}}\,|\psi_{\mathcal{S}*\mathcal{E}}\rangle = |\psi_{\mathcal{S}*\mathcal{E}}\rangle\,.$$

We will follow Zurek's practice in calling *operators* envariant on given states if all the states are envariant under them. For now we also assume all Hilbert spaces to be of finite dimension, so that one should think of summation indices as ranging from 1 to some suitable $N \in \mathbb{N}$. The infinite and continuous cases will be considered subsequently.

Informally, envariance means that operations performed on the system can entirely be undone by operations performed on the environment alone. This feature is quite peculiar to QM, and, as mentioned above, it occurs in virtue of entangled states. To see this, consider some state

$$|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j \alpha_j\,|\sigma_j\rangle|\varepsilon_j\rangle \tag{45}$$

in Schmidt decomposition (cf. footnote 8), with the $|\sigma_j\rangle$ $\mathcal{S}$'s states and the $|\varepsilon_j\rangle$ $\mathcal{E}$'s. Now $\hat{U}_{\mathcal{S}}$ may have the effect

$$\hat{u}_{\mathcal{S}}\,|\sigma_j\rangle = e^{\imath\varphi_j}\,|\sigma_j\rangle\,, \tag{46}$$

on each $|\sigma_j\rangle$, i.e., of inducing a particular phase φ_j on each of the states respectively. This can be completely undone by

$$\hat{u}_{\mathcal{E}}\,|\varepsilon_j\rangle = e^{-\imath(\varphi_j+2\pi m_j)}\,|\varepsilon_j\rangle\,, \tag{47}$$

for $m_j \in \mathbb{Z}, \forall j$, since the phases φ cancel, and $e^{-\imath 2\pi m_j} = 1$, for the given choices of m_j [86, 87]. In fact there is the following lemma about the form of such envariant operators.

[25] Even though the term is introduced more 'casually' by Zurek, we take it that he intends an explicit definition.

Lemma 4.2 (*Codiagonalizability*)**.** An operator $\hat{U}_{\mathcal{S}} = \hat{u}_{\mathcal{S}} \otimes \mathbb{1}_{\mathcal{E}}$ is envariant on a state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j \alpha_j |\sigma_j\rangle|\varepsilon_j\rangle$ (where $|\sigma_j\rangle \in \mathcal{H}_{\mathcal{S}}$, $|\varepsilon_j\rangle \in \mathcal{H}_{\mathcal{E}}, \forall j$) *iff* $\hat{u}_{\mathcal{S}}$ can be given in the (codiagonal) form

$$\hat{u}_{\mathcal{S}} = \sum_j \exp\!\big(\imath\varphi_j\big) |\sigma_j\rangle\langle\sigma_j| \,. \tag{†}$$

Proof.
Direction "⇒": If $\hat{u}_{\mathcal{S}}$ is of the form (†), $\hat{U}_{\mathcal{S}}$ can always be undone by some $\hat{U}_{\mathcal{E}} = \mathbb{1}_{\mathcal{S}} \otimes \hat{u}_{\mathcal{E}}$ with

$$\hat{u}_{\mathcal{E}} = \sum_j \exp\!\big(-\imath(\varphi_j + 2\pi m_j)\big) |\varepsilon_j\rangle\langle\varepsilon_j| \qquad m_j \in \mathbb{Z}, \forall j. \tag{⋆}$$

Direction "⇐": Assume that $\hat{u}_{\mathcal{S}}$ cannot be given as (†). Then

$$\hat{U}_{\mathcal{S}} |\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j \alpha_j |\hat{u}_{\mathcal{S}}\sigma_j\rangle|\varepsilon_j\rangle = \sum_j \alpha_j |\tilde{\sigma}_j\rangle|\varepsilon_j\rangle =: |\eta_{\mathcal{S}*\mathcal{E}}\rangle \,,$$

where by assumption $\exists j \forall \varphi : e^{\imath\varphi} |\tilde{\sigma}_j\rangle \notin \{|\sigma_j\rangle\}_{j\in J}$. Since $\hat{U}_{\mathcal{S}}$ is supposed to be envariant on $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$, there must be a $\hat{U}_{\mathcal{E}} = \mathbb{1}_{\mathcal{S}} \otimes \hat{u}_{\mathcal{E}}$ s. t.

$$\hat{U}_{\mathcal{E}} |\eta_{\mathcal{S}*\mathcal{E}}\rangle = |\psi_{\mathcal{S}*\mathcal{E}}\rangle \,.$$

But $\hat{U}_{\mathcal{E}}$ only has an effect on $\mathcal{H}_{\mathcal{E}}$, so it cannot effect a mapping $|\tilde{\sigma}_j\rangle \mapsto |\sigma_j\rangle$ for the $|\tilde{\sigma}_j\rangle$ not contained (up to phase) in $\{|\sigma_j\rangle\}_{j\in J}$. Contradiction. □

Moreover, one can easily see that the environment $\mathcal{E}$ is not uniquely defined, since plugging in an apparatus $\mathcal{M}$, the state

$$|\psi_{\mathcal{S}*\mathcal{M}*\mathcal{E}}\rangle = \sum_j \alpha_j |\sigma_j\rangle|\mu_j\rangle|\varepsilon_j\rangle \tag{48}$$

is obviously envariant in the same sense for manipulations

$$\hat{U}_{\mathcal{M}} = \mathbb{1}_{\mathcal{S}} \otimes \Big(\sum_j \exp\!\big(-\imath(\varphi_j + 2\pi m_j)\big) |\mu_j\rangle\langle\mu_j|\Big) \otimes \mathbb{1}_{\mathcal{E}} \tag{49}$$

of the apparatus states $|\mu_j\rangle$; and in principle, a larger environment can always be defined (depending on the resources in the universe of course) on which the undoing operation is performed [87].

In principle, the phases can be understood as generated by the eigenvalues $\omega_j = \frac{E_j}{\hbar}$ up to $\hbar$ of some Hamiltonian whose eigenstates are those of the Schmidt decomposition, multiplied by duration t of the evolution described by the Hamiltonian. I.e., $\varphi_j = \omega_j^{\mathcal{S}} \cdot t_{\mathcal{S}}$ and $-\varphi_j - 2\pi m_j = \omega_j^{\mathcal{E}} \cdot t_{\mathcal{E}}$, where the eigenvalues are essentially arbitrary, up to their interrelatedness on $\mathcal{S}$ and $\mathcal{E}$ [86].

However, for certain states, there is an interesting subclass of envariant operations:

Lemma 4.3 (*Envariance under swaps*)**.** A state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j \alpha_j |\sigma_j\rangle|\varepsilon_j\rangle$ is envariant under a *swap*

$$\hat{u}_{\mathcal{S}}(i \leftrightharpoons k) \otimes \mathbb{1}_{\mathcal{E}} = \Big(e^{\imath\varphi_{i,k}} |\sigma_i\rangle\langle\sigma_k| + e^{-\imath\varphi_{i,k}} |\sigma_k\rangle\langle\sigma_i| + \sum_{i \neq j \neq k} |\sigma_j\rangle\langle\sigma_j|\Big) \otimes \mathbb{1}_{\mathcal{E}} \tag{$\diamond$}$$

iff it holds that $|\alpha_i| = |\alpha_k|$.

Proof.
Direction "$\Rightarrow$": The swap is 'undone' by a counterswap on $\mathcal{H}_{\mathcal{E}}$:

$$\hat{u}_{\mathcal{E}}(i \leftrightharpoons k) = e^{-\imath(\varphi_{i,k}+\varphi_i-\varphi_k+2\pi m_{ik})} |\varepsilon_i\rangle\langle\varepsilon_k| + e^{\imath(\varphi_{i,k}+\varphi_i-\varphi_k+2\pi m_{ik})} |\varepsilon_k\rangle\langle\varepsilon_i| + \sum_{i \neq j \neq k} |\varepsilon_j\rangle\langle\varepsilon_j|,$$
$$m_{ik} \in \mathbb{Z}. \tag{50}$$

Direction "$\Leftarrow$": Assume that $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ is envariant under ($\diamond$) and $|\alpha_i| \neq |\alpha_k|$. Then we get states $\alpha_k |\sigma_i\rangle \neq |\sigma_i\rangle, \alpha_i |\sigma_k\rangle \neq |\sigma_k\rangle$ whence we cannot restore the original state by merely acting unitarily on $|\varepsilon_i\rangle, |\varepsilon_k\rangle$. Contradiction. □

The existence of swaps is not in contradiction to the codiagonalizability lemma 4.2, since one can equally prove the $\Rightarrow$-part by appeal to the Hadamard basis of the subspace of two states being swapped, i.e., the basis of equal- and opposite-sign superposition with equal coefficients, in which the swap is diagonal and hence effectively of the form exposed in (†) [87].

Swaps of course exchange the i-th and k-th sate of $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$. If $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ is of the 'even' form

$$|\overline{\psi}_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j |\alpha| e^{\imath\varphi_j} |\sigma_j\rangle|\varepsilon_j\rangle, \tag{51}$$

the state is envariant under arbitrary and simultaneous (or 'total') swaps of all system states.

We note at this point that the restriction to *unitary* operators figures crucially in the proofs of the two lemmas, since an exchange of two states $|\sigma_i\rangle, |\sigma_k\rangle$ that implies changing the moduli as $|\alpha_i| \mapsto |\alpha_k|, |\alpha_k| \mapsto |\alpha_i|$ could be undone by a non-norm preserving transformation that exchanges the environment states $|\varepsilon_i\rangle, |\varepsilon_k\rangle$ and simultaneously reduces their norms as $|\alpha_k| \mapsto |\alpha_k| \cdot |\alpha_i|/|\alpha_k|$ and $|\alpha_i| \mapsto |\alpha_i| \cdot |\alpha_k|/|\alpha_i|$ respectively.

4.2 Deriving the Born Probabilities

The last insight of the previous section, that entangled states with equal moduli for (some) all terms are envariantly swappable, is considered the "cornerstone" of his derivation by Zurek [87]. Next to this, he assumes three premises, which he thinks of as interpretation-neutral "facts", and which he believes only to clarify the notions of "state" and "(sub)system", as they are used ubiquitously in the debate [*ibid.*]:[26]

Premise 1 (UAct)**.** To represent the alteration of the state of a system $\mathcal{S}$ by a unitary operator $\hat{u}$, $\hat{u}$ must act on the Hilbert space $\mathcal{H}_{\mathcal{S}}$ of that system.

[26]Note that we give paraphrases of these, not quotes, as we shall do with other ingredients of Zurek's proof. 'UAct' may stand for the need of an *act*ion of a *u*nitary on a system's space to alter its state, 'SEnc' for properties and probabilities being *enc*oded into a system's *s*tate, and 'SInc' for the *inc*lusion of them system's *s*tate in the joint state with the environment.

Premise 2 (SEnc)**.** All measurable quantities pertaining to a system $\mathcal{S}$ *and their respective probabilities* are fully and exclusively specified by $\mathcal{S}$'s state.

Premise 3 (SInc)**.** The state of a subsystem $\mathcal{S}_j$ included in a larger system $\mathcal{S} = \mathcal{S}_1 * \ldots * \mathcal{S}_j * \ldots * \mathcal{S}_N$ is fully and exclusively specified by the state of $\mathcal{S}$.

We have given these premises names for further reference and in virtue of their importance for all subsequent steps. They clearly include presuppositions about probabilities and states and how they relate to QM. To some degree though, they merely capture certain intuitions about the meanings of such terms as "state" or "(sub)system" as they are used in QM. But we should later still assess how plausible the premises of Zurek's proof are, and this will include an indirect assessment of the three.

Before proceeding to the Born rule now, Zurek first proves the following theorem:

Theorem 4.4 (*Phase independence*)**.** For an entangled global state of system $\mathcal{S}$ and environment $\mathcal{E}$, all measurable quantities pertaining to $\mathcal{S}$ and their probabilities are independent of the phases φ_j in the Schmidt representation $|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j e^{\imath\varphi_j}|\alpha_j|\,|\sigma_j\rangle|\varepsilon_j\rangle$.

Proof. By premise 1, a phase transformation must be a unitary operator $\hat{U}_\mathcal{S}$ acting on $\mathcal{H}_\mathcal{S}$ of $\mathcal{S}$. In Schmidt representation, all such phase transformations are of the form (†), and hence envariant. By that token, the phase transformation can be undone by an operation $\hat{U}_\mathcal{E} = \mathbb{1}_\mathcal{S} \otimes \hat{u}_\mathcal{E}$ as in (⋆), acting on $\mathcal{H}_\mathcal{E}$, which, again by premise 1, cannot affect the state of $\mathcal{S}$.

Since this restores the state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$, however, by premise 3 the state of $\mathcal{S}$ must be restored as well. Once more by premise 1, this means that $\mathcal{S}$ must have remained in the same state throughout the operation $\hat{U}_\mathcal{E}\hat{U}_\mathcal{S}$. Thus, by premise 2, the statement holds. □

The gist of this result is that $\mathcal{S}$'s state and the probabilities in question must be fully determined by the set of pairs $\{(|\alpha_j|, |\sigma_j\rangle)\}_{j\in J}$. This is an important conceptual step, and the proof of the theorem is instructive for the remainder of the discussion.

Subsequently, Zurek establishes the following theorem:

Theorem 4.5 (*Equal probabilities for equal coefficients*)**.** The probabilities for states $|\sigma_k\rangle$, $|\sigma_\ell\rangle$ in a Schmidt state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j \alpha_j |\sigma_j\rangle|\varepsilon_j\rangle$ of a system $\mathcal{S}$ and its environment $\mathcal{E}$ which have coefficients of equal modulus $|\alpha_k| = |\alpha_\ell| = |\alpha|$ are all equal.

The proof of this theorem requires, however, either of the three following additional premises, which all function as *bridging principles* for establishing (or connecting) *values* of probabilities:

Premise 4(a). If swaps of two orthonormal states in a joint state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ leave the state of $\mathcal{S}$ unchanged, the probabilities for the outcomes associated with these states must be equal.

Premise 4(b). If the state of $\mathcal{S}$ in $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ is left unchanged by all conceivable unitary transformations acting on a subspace $\tilde{\mathcal{H}}_\mathcal{S} \subset \mathcal{H}_\mathcal{S}$, then the probabilities for all outcomes of a measurement associated with the states in an ONB of $\tilde{\mathcal{H}}_\mathcal{S}$ are equal.

Premise 4(c). The outcomes associated with states in a tensor term of some Schmidt decomposition are perfectly correlated, i.e., if σ_j is measured on $|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j \alpha_j |\sigma_j\rangle|\varepsilon_j\rangle$, ε_j will be measured with probability 1 as well (and *vice versa*).

The proof from 4(a) is straightforward:

Proof. *(a)* Theorem 4.5 is implicitly of a conditional form (*if* we have a Schmidt state with equal coefficients, *then*...), so we assume the antecedent (the if-part) and prove the consequent (the then-part). Thus assume a Schmidt state with at least two equal-modulus coefficients. Since swaps are envariant on the subspace of these states (lemma 4.3), we get from an analogous reasoning as in theorem 4.4 that they leave the state unchanged, whence, by *modus ponens* on 4(a), we get the desired consequent. □

The proof from 4(b) is more involved and requires the following lemma:

Lemma 4.6 (*Envariance under partial swaps*)**.** Let $\mathcal{S} := \{\,|\sigma_j\rangle\}_{j=1}^K, \tilde{\mathcal{S}} := \{\,|\tilde{\sigma}_j\rangle\}_{j=1}^K$ be two orthonormal bases that span an even subspace $\overline{\mathcal{H}}_{\mathcal{S}} \subset \mathcal{H}_{\mathcal{S}}$, i.e., a space of states on which the Schmidt state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ has equal moduli for all coefficients. Then $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ is also envariant under swaps of all states in $\{\,|\sigma_j\rangle\}_{j=1}^K$ for those in $\{\,|\tilde{\sigma}_j\rangle\}_{j=1}^K$ ('partial swaps').

Proof. By its definition, a partial swap (restricted to $\overline{\mathcal{H}}_{\mathcal{S}}$) is of the form

$$\hat{u}_{\mathcal{S}}(\mathcal{S} \leftrightharpoons \tilde{\mathcal{S}}) = \sum_{|\sigma_j\rangle\in\overline{\mathcal{H}}_{\mathcal{S}}} |\tilde{\sigma}_j\rangle\langle\sigma_j|\,.$$

The restriction $|\overline{\psi}_{\mathcal{S}*\mathcal{E}}\rangle$ of $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ to $\overline{\mathcal{H}}_{\mathcal{S}}$ satisfies

$$|\overline{\psi}_{\mathcal{S}*\mathcal{E}}\rangle \propto \sum_{j=1}^K e^{\imath\varphi_j}\,|\sigma_j\rangle|\varepsilon_j\rangle\,.$$

Since $\tilde{\mathcal{S}}$ is an ONB for $\overline{\mathcal{H}}_{\mathcal{S}}$, we get that

$$|\overline{\psi}_{\mathcal{S}*\mathcal{E}}\rangle \propto \sum_{j=1}^K e^{\imath\varphi_j}\Big(\sum_{\ell=1}^K |\tilde{\sigma}_\ell\rangle\langle\tilde{\sigma}_\ell|\Big)|\sigma_j\rangle|\varepsilon_j\rangle = \sum_{\ell=1}^K |\tilde{\sigma}_\ell\rangle\underbrace{\Big(\sum_{j=1}^K e^{\imath\varphi_j}\,\langle\tilde{\sigma}_\ell|\sigma_j\rangle|\varepsilon_j\rangle\Big)}_{=:|\tilde{\varepsilon}_\ell\rangle}. \tag{‡}$$

The $|\tilde{\varepsilon}_\ell\rangle$ satisfy

$$\begin{aligned}\langle\tilde{\varepsilon}_i|\tilde{\varepsilon}_k\rangle &= \Big(\sum_{j=1}^K e^{-\imath\varphi_j}\,\langle\sigma_j|\tilde{\sigma}_i\rangle\,\langle\varepsilon_j|\Big)\Big(\sum_{\ell=1}^K e^{\imath\varphi_\ell}\,\langle\tilde{\sigma}_k|\sigma_\ell\rangle\,|\varepsilon_\ell\rangle\Big) = \\ &= \sum_{j,\ell=1}^K e^{\imath(\varphi_\ell-\varphi_j)}\,\langle\tilde{\sigma}_k|\,|\sigma_\ell\rangle\langle\sigma_j|\,|\tilde{\sigma}_i\rangle\,\underbrace{\langle\varepsilon_j|\varepsilon_\ell\rangle}_{=\delta_{j\ell}} = \langle\tilde{\sigma}_k|\tilde{\sigma}_i\rangle = \delta_{ik},\end{aligned}$$

in virtue of $|\overline{\psi}_{\mathcal{S}*\mathcal{E}}\rangle$ being a Schmidt state (i.e., $\{\,|\varepsilon_j\rangle\}_{j=1}^K =: \mathcal{E}$ being an ONB of $\overline{\mathcal{H}}_{\mathcal{E}}$, the $\mathcal{E}$-analogue of $\overline{\mathcal{H}}_{\mathcal{S}}$). Thus, the RHS of (‡) is just the same state rewritten in a different Schmidt decomposition, and

$$\hat{u}_{\mathcal{E}}(\mathcal{E} \leftrightharpoons \tilde{\mathcal{E}}) = \sum_{|\varepsilon_j\rangle \in \overline{\mathcal{H}}_{\mathcal{E}}} e^{-\imath\varphi_j} |\tilde{\varepsilon}_j\rangle\langle\varepsilon_j|$$

defines a partial counterswap on $\overline{\mathcal{H}}_{\mathcal{E}}$, where $\tilde{\mathcal{E}} := \{ |\tilde{\varepsilon}_j\rangle\}_{j=1}^{K}$. □

Since all conceivable unitaries on $\overline{\mathcal{H}}_{\mathcal{S}}$ are in fact such partial swaps – i.e., *changes of basis* of $\overline{\mathcal{H}}_{\mathcal{S}}$ – including 'trivial' ones, that merely rotate the states within some basis, the proof from 4(b) now follows immediately:

Proof. (b) If we again assume the antecedent of theorem 4.5, then the states of equal modulus define a subspace $\overline{\mathcal{H}}_{\mathcal{S}}$ with the properties required for lemma 4.6 to apply. By lemma 4.6, all unitaries on $\overline{\mathcal{H}}_{\mathcal{S}}$ leave $\overline{\mathcal{H}}_{\mathcal{S}}$ unchanged, whence by 4(b), the theorem follows. □

Finally, the proof from 4(c) proceeds in a relatively straightforward manner again:

Proof. (c) From the perfect correlation, we get that

$$\left.\begin{array}{l} p(\sigma_j|\varepsilon_j) = 1 = \frac{p(\sigma_j,\varepsilon_j)}{p(\varepsilon_j)} \Rightarrow p(\sigma_j,\varepsilon_j) = p(\varepsilon_j) \\ p(\varepsilon_j|\sigma_j) = 1 = \frac{p(\varepsilon_j,\sigma_j)}{p(\sigma_j)} \Rightarrow p(\sigma_j,\varepsilon_j) = p(\sigma_j) \end{array}\right\} \Rightarrow p(\varepsilon_j) = p(\sigma_j).$$

Now consider a swap of states $|\sigma_k\rangle$, $|\sigma_\ell\rangle$ in a Schmidt state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j \alpha_j |\sigma_j\rangle|\varepsilon_j\rangle$, and denote the probability for finding outcomes σ_k, σ_ℓ in $|\eta_{\mathcal{S}*\mathcal{E}}\rangle$, which is $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ after the swap, by $\tilde{p}(\sigma_{k/\ell})$ respectively. Then from 4(c) and the transitivity of identity, we also have

$$\tilde{p}(\sigma_{k/\ell}) = p(\varepsilon_{\ell/k}) = p(\sigma_{\ell/k}).$$

For $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ a state with equal-modulus coefficients (we choose these to be the k-th and ℓ-th), as required by the antecedent of theorem 4.5, we can counterswap to restore $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$. Again using UAct and the fact that $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ is restored without consecutive action on $\mathcal{S}$'s Hilbert space after the swap, we can conclude from SInc that $\mathcal{S}$'s state is untouched by the swap. Using now SEnc, we can conclude that

$$\tilde{p}(\sigma_{\ell/k}) = p(\sigma_{\ell/k}),$$

and using the transitivity of identity again, we get

$$p(\sigma_{\ell/k}) = p(\sigma_{k/\ell})$$

as desired. □

For an even state (with all N moduli equal), this of course means that $p(\sigma_k) = 1/N, \forall 1 \leq k \leq N$, assuming that there are no 'hidden' sates that do not figure in the decomposition.

By considering the general case where the moduli of the coefficients need not be equal at all, Zurek proves the Born rule for finite dimensions ($\dim(\mathcal{H}) = N \in \mathbb{N}$).

Theorem 4.7 (*Born's rule for finite dimensions*)**.** For any Schmidt state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_{j=1}^{N} \alpha_j |\sigma_j\rangle|\varepsilon_j\rangle$, $N \in \mathbb{N}$ it holds that $p(\sigma_j) = |\alpha_j|^2$.

Proof. The proof appeals to the fact that the environment is not uniquely defined (cf. section 4.1). It proceeds in two stages/by case distinction:

(i) Consider a state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ where one can rewrite the coefficients α_j as $\sqrt{\frac{m_j}{M}}e^{\imath\varphi_j}$, with $m_j, M \in \mathbb{N}, \forall 1 \leq j \leq N$. Then $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ obtains the form

$$|\psi_{\mathcal{S}*\mathcal{E}}\rangle \propto \sum_{j=1}^{N} \sqrt{m_j}e^{\imath\varphi_j}\,|\sigma_j\rangle|\varepsilon_j\rangle\,.$$

Now we introduce a third system $\mathcal{C}$, either extracted from $\mathcal{E}$ or (by definition of $\mathcal{E}$) not included in it from the outset, and which may be thought of as representing the measuring device. Then the joint state becomes

$$|\psi_{\mathcal{S}*\mathcal{E}*\mathcal{C}}\rangle \propto \sum_{j=1}^{N} \sqrt{m_j}e^{\imath\varphi_j}\,|\sigma_j\rangle|\varepsilon_j\rangle|\zeta_j\rangle\,,$$

$|\zeta_j\rangle$ the $\mathcal{C}$-states. Now additionally assume the space $\mathcal{H}_\mathcal{C}$ to be of sufficient dimensionality so that the $|\zeta_j\rangle$ can be expanded in (or 'fine-grained' into) subspaces of dimension m_j as

$$|\zeta_j\rangle = \frac{1}{\sqrt{m_j}} \sum_{i=r_{j-1}+1}^{r_j} |c_{i_j}\rangle\,,$$

where the r_j are chosen such that $r_j = r_{j-1} + m_j, r_0 = 0$. Also, consider $\mathcal{C}$ to be such that the joint evolution of $\mathcal{C}$ and $\mathcal{E}$ proceeds as

$$|\varepsilon_j\rangle|c_{i_j}\rangle \longrightarrow |e_{i_j}\rangle|c_{i_j}\rangle\,, \tag{52}$$

with $\langle e_{i_j}|e_{\ell_k}\rangle = \delta_{i_j\ell_k}$, i.e., where the (larger) environment 'monitors' the state of $\mathcal{C}$ (the measuring device). Given the expansion of the $|\zeta_j\rangle$, we can hence rewrite $|\psi_{\mathcal{S}*\mathcal{E}*\mathcal{C}}\rangle$ as

$$|\psi_{\mathcal{S}*\mathcal{E}*\mathcal{C}}\rangle \propto \sum_{i_j=1}^{M} e^{\imath\varphi_{i_j}}\,|\sigma_{j(i_j)}\rangle|e_{i_j}\rangle|c_{i_j}\rangle\,, \tag{53}$$

where M now appears as the upper limit of the N intervals of magnitude m_j, as defined by the fine-graining ($M = \sum_{j=1}^{N} m_j$), and $j(i_j) = j$ for every $r_{j-1} < i_j \leq r_j$. It now follows from theorem 4.5 that $p(c_{i_j}) = \frac{1}{M}$, and since this goes for all m_j values on $\mathcal{C}$ for $\sigma_{j(i_j)}$, we get

$$p(\sigma_j) = \frac{m_j}{M} = |\alpha_j|^2, \tag{54}$$

as required by Born's rule.

(ii) In case there are no such $m_j \in \mathbb{N}$, one can always approximate the $|\alpha_j|$ by defining bounds $\sqrt{\frac{\check{m}}{M}} < |\alpha_j| < \sqrt{\frac{\check{m}+1}{M}}$ where $M, \check{m} \in \mathbb{N}$, since $\mathbb{Q}$ is dense in $\mathbb{R}$, and hence get the same results as in (i) by letting $M \longrightarrow \infty$ (i.e., by 'increasing the fineness of the grain' in the above procedure). □

The intermediate assumption in this proof, that $\mathcal{H}_\mathcal{C}$ allows for the required fine-graining, can (for now) be understood in the sense that $\mathcal{C}$ is chosen to be 'rich enough in resources' such as to map the state of the system measured to sufficient accuracy to keep track of frequencies. For instance, no less than the states of a spin-3/2 system with $|s = \frac{3}{2}; m = \frac{3}{2}\rangle \doteq (1, 0, 0, 0)^t, \ldots$ would do to accomplish this task for a single spin-1/2 system with $|\!\uparrow\rangle \doteq \binom{1}{0}$, $|\!\downarrow\rangle \doteq \binom{0}{1}$, in an unequal superposition $\sqrt{\frac{3}{4}}\,|\!\uparrow\rangle + \frac{1}{\sqrt{4}}\,|\!\downarrow\rangle$, say. The continuum limit in case (ii) can then be read as an idealization, in the sense that the only thing 'measurable' in an actual experiment is a relative frequency anyway, and that a sufficiently complex apparatus is needed to deliver frequencies that approximate the real (irrational) numbers predicted by the Born rule to a satisfactory degree.

However, there is of course another limit involved in QM, namely the limit of $\dim(\mathcal{H}) = \aleph_0$ or $\dim(\mathcal{H}) = \aleph_1$, i.e., $|\psi\rangle = \sum_{j=1}^{\infty} \alpha_j\,|\phi_j\rangle$ or $|\psi\rangle = \int \mathrm{d}\lambda\,\alpha(\lambda)\,|\lambda\rangle$.[27] To formally include these cases, Zurek argues as follows. In a (separable) Hilbert space of countable dimension, one can write, for a given value δ and the Hilbert space of $\mathcal{C}$ finite-dimensional, the state of $\mathcal{S} * \mathcal{E} * \mathcal{C}$ as

$$|\psi_{\mathcal{S}*\mathcal{E}*\mathcal{C}}\rangle = \sum_{j=1}^{N_\delta} \alpha_j\,|\sigma_j\rangle|\varepsilon_j\rangle|c_j\rangle + \delta\,|r_{N_\delta+1}\rangle|c_{N_\delta+1}\rangle\,, \tag{55}$$

where the residual state $|r_{N_\delta+1}\rangle$ summarizes all the remaining states of $\mathcal{S} * \mathcal{E}$ from $j = N_\delta + 1$ up to "$N = \infty$". The α_j in both the finite sum and the residual may be assumed as ordered according to the magnitude of $|\alpha_j|$. Then $\delta \longrightarrow 0$ as $N_\delta \longrightarrow \infty$. Since Born's rule holds for the two finite terms, $\sum_{j=1}^{N_\delta} \alpha_j\,|\sigma_j\rangle|\varepsilon_j\rangle|c_j\rangle$ and $\delta\,|r_{N_\delta+1}\rangle|c_{N_\delta+1}\rangle$, and the total probability must add up to one, it follows that the conditional probability of getting one of the outcomes in the first sum is given by

$$p(\sigma_j | j \leq N_\delta) = \frac{|\alpha_j|^2}{1 - |\delta|^2} \xrightarrow{N_\delta \to \infty} |\alpha_j|^2. \tag{56}$$

The continuous case proceeds in a similar manner by approximating

$$\psi(x) \approx \Xi(x) = \sum_k \psi_k \chi_{\Delta_k}(x), \tag{57}$$

which may be a sum or a series, depending on the behavior of ψ, and where $\chi_{\Delta_k}(x)$ is the characteristic function for $\Delta_k = [k\Delta x, k\Delta x + \Delta x)$ (Δx some fixed interval on the real line), i.e., $\chi_{\Delta_k}(x) = 1$ for $x \in \Delta_k$, 0 else. ψ_k is given by

$$\psi_k = \frac{1}{\Delta x} \int_{\Delta_k} \psi(x)\,\mathrm{d}x\,, \tag{58}$$

i.e., $\psi(x)$ averaged over Δ_k. The $\chi_{\Delta_k}(x)$ do not really form a basis of $L^2(\mathbb{R})$, as can be given in terms of Hermite polynomials [45], but they obviously satisfy $\langle \chi_{\Delta_k} | \chi_{\Delta_{k'}} \rangle = \delta_{kk'} \Delta x$, so they can at least be normalized by including a factor $(\Delta x)^{-1/2}$.

The approximation may again be taken as implying the existence of a remainder $r(x) :=$

[27] Strictly speaking, the latter case of course requires a Gelfand triple or 'rigged' Hilbert space, rather than a proper Hilbert space [19].

$\psi(x) - \Xi(x)$, where one immediately sees that

$$\begin{aligned}\langle\Xi|r\rangle &= \int_{\mathbb{R}} \Xi^*(x) r(x)\,\mathrm{d}x = \int_{\mathbb{R}} \Xi^*(x)\psi(x)\,\mathrm{d}x - \langle\Xi|\Xi\rangle = \\ &= \int_{\mathbb{R}} \sum_k \frac{1}{\Delta x} \int_{\Delta_k} \psi^*(x')\,\mathrm{d}x'\,\chi_{\Delta_k}(x)\psi(x)\,\mathrm{d}x - \langle\Xi|\Xi\rangle = \\ &= \sum_k \frac{1}{\Delta x} \int_{\Delta_k} \psi^*(x')\,\mathrm{d}x' \int_{\Delta_k} \psi(x)\,\mathrm{d}x - \langle\Xi|\Xi\rangle = \\ &= \Delta x \sum_k \psi_k^* \psi_k - \langle\Xi|\Xi\rangle = 0. \end{aligned} \tag{59}$$

Realizing that

$$\langle\Xi|\Xi\rangle = \sum_k |\psi_k|^2 \Delta x = \langle\psi|\psi\rangle - \langle r|r\rangle = 1 - |\delta|^2, \tag{60}$$

where $|\delta|^2$ vanishes for $\Delta x \longrightarrow 0$, which justifies considering $|\psi_k|^2\Delta x$ as an approximation to $|\psi(x)|^2\,\mathrm{d}x$, the argument now proceeds in analogy and by appeal to the foregoing results, applied to the 'pseudo-Schmidt' state

$$|\Psi_{\mathcal{S}*\mathcal{E}*\mathcal{C}}\rangle = \sqrt{\Delta x} \sum_k |\psi_k| e^{\imath\varphi_k} \frac{|\chi_{\Delta_k}\rangle}{\sqrt{\Delta x}} |\varepsilon_k\rangle |c_k\rangle\,, \tag{61}$$

so that $|\psi_k|^2\Delta x = p(x \in \Delta_k)$. The Born rule may thus be considered to be proved for this case as well. Moreover, it generally holds for some chosen $k_i < k_f$ and $\Gamma := \Delta_{k_i} \cup \ldots \cup \Delta_{k_f}$, that

$$\sum_{k=k_i}^{k_f} |\psi_k|^2 \Delta x = \int_\Gamma |\psi(x)|^2\,\mathrm{d}x\,, \tag{62}$$

so that Born's rule for expressions of the form $\mathrm{Pr}_x^\psi(x \in \Gamma)$ – which are arguably the only empirically meaningful statements for the continuous case – is justified in this limit even more straightforwardly.[28]

4.3 Discussion (ii)

The foregoing proof of the Born rule is, again, deductively seamless. However, we must wonder whether (A) its premises are suitably justified, and (B) in how far the result really supports the MWI. Regarding (B), we first note that Zurek's derivation strongly requires the applicability of entangled (Schmidt-) states for $\mathcal{S} * \mathcal{E}$ on a formal *and* on a conceptual level, and therefore *prima facie* proceeds within an inherently MWI-friendly framework [6, 17]. As regards (A), there have indeed been multiple critical assessments [6, 17, 57, 67] of Zurek's earlier paper [86] in which the proof was presented in a less explicit manner. We will assess some of the issues below, but first make a few more general observations.

[28] We here leave aside the subtleties briefly discussed in [87] that arise for functions that are not sufficiently smooth over small scales.

A first important observation is the following. The proof of the Born rule requires theorem 4.5, and, following Zurek, we have outlined three possible ways to derive this theorem. It appears from our analysis that the dervation via premise 4(b) is the 'least costly', since here, in fact, no actual use is made of premise 2 (SEnc).[29] But this appearance comes out flawed upon closer inspection, since premise 4(b) is actually the least plausible, and therefore in greatest need of justification. For it is a conditional statement, stating a relation between basis changes in a Hilbert space and probabilities, which appears rather unmotivated.

Strictly speaking, 4(b) tells us that, if we can arbitrarily change the basis in some particular subspace of a system's Hilbert space, then we should think of all states in any given basis of that space – or rather, the outcomes associated with these – as equiprobable. But why exactly should we do such a thing? Note that we are here not merely 'fiddling around' with the states within *one particular* basis that represents meaningfully interpretable outcomes; the *entire basis* is replaced by another one. It is completely elusive why this should imply that the states in the *one* basis that we have found to be meaningful are equiprobable. Only in the special case where one reassigns the coefficients (i.e., effectively the *phases*) among states *in that particular basis* without effecting a change in the total state of the system (as defined in virtue of premises 1 and 3) one could make a case that these basis states should turn up equally likely in a measurement, since the environment should be 'indifferent' as to their occurrence. And that is premise 4(a).

In fact, Zurek construes his 'envariance under swaps' as a kind of *objectivization* of the *principle of indifference*, i.e., the principle of assigning equal probability to any out of a set of mutually exclusive and exhaustive events about which there is nothing known but their names or labels [43]. According to Zurek, the QM probabilities arise not out of a subjective lack of knowledge as to which state of a system actually obtains, but instead out of a precise knowledge about the total state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ and the symmetries implied thereby. It is not really clear though what this means, or what concept of probability is endorsed here. It is meaningless to suppose that the *environment* be literally indifferent as to the outcomes of an experiment, since the environment is not an agent with expectations, so it does not even satisfy necessary requirements for exhibiting indifference. But if this indifference is meant to apply to an observer in virtue of known symmetries, then this still needs further explanation. For if an agent $\mathcal{A}$ is indifferent (say) between sitting on either of two chairs, a and b, put in a particular position, because they are equally comfortable – i.e., because the situation is 'symmetric under chair-swaps' – then this indifference still arises in virtue of two *different* outcomes *for* $\mathcal{A}$.

It has been objected [57] that Zurek claims to elucidate the origin of probabilities in QM, but that he does no such thing and really has to insert probabilities by hand. It is doubtful though that Zurek is truly guilty of that mistake, as should be clear form our analysis of the (revised, more detailed exposition of the) proof, where SEnc is assumed to even associate states of systems with probabilities at all, as well as one of the premises 4(a)-(c) as a bridging principle to arrive at values. Moreover, Schlosshauer and Fine have noted that this is a feature of *any* probabilistic theory: "we cannot derive probabilities from a theory that does not already contain some probabilistic concept; at some stage, we need to 'put probabilities in to get probabilities out.'" [67] So what Zurek establishes – and arguably only claims to establish – is that *if* QM is associated with probabilities in the first place, then these follow the Born rule.

[29] The reader will find that we have well used the other two premises (UAct, SInc) to apply the formal representations of the notions 'state' and 'unchanged' in the proof.

But this in turn means exactly that Zurek's proof does not at all answer the question as to the *meaning* of 'probability' in the MWI, but instead equally requires us to answer it. Upon the above line of reasoning, we are faced with a choice between premises 4(a) and 4(c), so we must assess the plausibility of these by supplying a suitable meaning to the word 'probability' as it occurs in them. Note, however, that the quantitative problem is also at best solved *partially* by the proof, since both of these premises put in *values* by hand—albeit only to the extent that they must be equal.

Intriguingly, Carroll and Sebens [16] have offered a proof of the Born rule which is very close to that of Zurek, but which supplements a more definite meaning to the probabilities. Based on the work of Vaidmann [74], Carroll and Sebens understand the probabilities as the *self-locating uncertainty* an observer has in the short time span τ_{unc} between physical measurement and conscious observation, i.e., between steps 2 and 3 of the measurement process, as we have identified them in section 1.2. They then base their version of the proof on what they call an *epistemic separability principle*, which (in essence) states that the probabilities an observer assigns to measurement values on $\mathcal{S}$ in a Schmidt-state $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ during τ_{unc} should remain invariant under actions $\hat{U}_{\mathcal{E}} = \mathbb{1}_{\mathcal{S}} \otimes \hat{u}_{\mathcal{E}}$ – so essentially on Zurek's premise 1, supplemented with a normative claim about probability assignments for (rational) agents.

Crucially, they also repeat the central step of Zurek's proof for the unequal amplitude case, i.e., multiply the number of states on a part of the environment (the system $\mathcal{C}$ in Zurek's derivation) so that their number will equal the value of the enumerator for amplitudes that are square roots of rational numbers. However, Carroll and Sebens make a much clearer concession as to the meaning of this procedure than does Zurek, namely (for the case where enumerators have the ratio $\sqrt{2}/1$), this procedure "reduces the problem of two branches with unequal amplitudes to that of three branches with equal amplitudes" [16] But here we are back to a form of *branch counting* and its associated problems that we elaborated on in section 2.2. And indeed, Zurek also refers to the system $\mathcal{C}$ he appeals to in the unequal amplitude-case as a counterweight *counter* [87].

It may be, though, that this is merely a *formal* step by Carroll and Sevens, as they later write: "In our approach, the question is [...] how the various future selves into which you will evolve should apportion their credences. Since every one of them should use the Born Rule, it is justified to talk *as if* future measurement outcomes simply occur with the corresponding probability." (The emphasis is theirs.) Here we are back to normative claims about how to assign credences to events instead, and hence ultimately to decision theoretic concerns, the appeal to which we have found clearly wanting above. Thus if the introduction of additional branches to arrive at the Born probabilities is merely a formal trick, then we lack good reasons as to why the Born rule should be our guiding principle to rational credences in the suitable contexts after all. *Only* if there is a well defined number of branches for which the agent can be uncertain as to which one he is on, he can reasonably "apportion his credences" in concert with the Born rule. Taking the counting strategy seriously is a necessary requirement for Carroll and Sebens' reasoning to make sense.

Again, the plausibility of premises 4(a) and 4(c) strongly depends on the meaning of 'probability'. If probability is an estimator for the relative frequency with which a given type of event will be observed in a temporally ordered sequence of equal experiments (which at the very least it should be in the context of QM), why *should* two states whose swap leaves the total state of $\mathcal{S}$ unchanged (in the envariance-sense) be 'equiprobable' (premise 4(a))? The most reasonable answer seems to be that observes in both situations, the swapped and the unswapped one, will observe the results associated with the two states with unaltered frequencies. And why should they do so, given that the MWI is true? *That* still needs a physical *explanation* à la "because there will be equal fractions of observers who all observe these outcomes in both of these situations", which in turn requires a well defined branch count.

Premise 4(c) opens up a quite different option though. If we take it that this is the actual content of a state such as $|\psi_{\mathcal{S}*\mathcal{E}}\rangle = \sum_j \alpha_j |\sigma_j\rangle|\varepsilon_j\rangle$, that there is a peculiar kind of *correlation* between possibly observed states of systems and their 'environments', and *not* that the world is a 'global superposition' in which all outcomes occur, then we do not have to answer the questions raised before. QM is just that kind of a theory, a theory that describes correlations in the behavior of certain systems which otherwise behave stochastically. It is clear that we have left the context of the MWI with this interpretation of 4(c), and we are thus back to confronting (B). It is hence not at all clear that the result *does* support the MWI.

In fact, Barnum [6] has offered a version of Zurek's proof, based on a *no-signaling* constraint. If the environment $\mathcal{E}$ is a spatially remote system and one could manipulate the probabilities for outcomes of measurements on $\mathcal{S}$ by merely acting on $\mathcal{E}$ (with $\hat{U}_{\mathcal{E}} = \mathbb{1}_{\mathcal{S}} \otimes \hat{u}_{\mathcal{E}}$), then this would open up the possibility for faster-than-light signaling, known to be associated with casual paradoxes and the like [52]. Moreover, the existence of strong or rather close-to-perfect correlations in states $|\psi_{\mathcal{S}*\mathcal{E}}\rangle$ where $\mathcal{S}$ and $\mathcal{E}$ are spacelike separated systems is of course well known from the violations of Bell inequalities [5, 10]. We hence take it that 4(c) is the most plausible out of the three, because it has the clearest empirical support. But for reasons given before, this in turn means that Zurek's derivation of the Born rule does *not* (strongly) support the MWI after all.

In addition, we have diagnosed Zurek's proof to crucially rely on *unitary* transformations. But why does QM implement state transformations by unitary operators? The preservation of superposition could be one suitable reason, since the superposition principle is a quite unique feature of QM (with suitable meaning attached to kets), contributing massively to its success. But preservation of superposition could also be achieved by other linear transformations, not just unitary ones. The clearest reason why unitary operators pervade QM is that they are *norm preserving*. And why should they be norm-preserving? So that *probabilities* are conserved and never exceed unity.[30] Note the analogy to accusations of circularity in decoherence-based proofs: we cannot rely on decoherence because it relies on partial tracing and partial tracing is inevitably motivated probabilistically. But then we should not rely on unitarity either, because unitarity is invoked to ensure norm preservation and norm preservation is inevitably motivated probabilistically [41].

Based on these arguments, Zurek's proof may be interpreted as one way of explicating the implicit probabilistic content of a – rather special – probabilistic theory, and as thereby deriving only the precise form of the rule according to which the probabilities contained in the theory should be assigned (Born's rule). This view does not fit well with the MWI though, according to which "*every* outcome with nonvanishing support in the wave function will occur (in some branch) with probability one." (The emphasis is that of Carroll and Sebens.) [16]

[30] An analogous argument is presented by Mohrhoff [57], but the present author was first made aware of it by his main supervisor, Dr. Klesse. Moreover, in [41] it is shown how statistical considerations crucially figure as reasons for the unitarity of state transformations in QM.

5 Conclusions

We have analyzed, in this thesis, a central problem of the MWI, the problem of what 'probability' means in a branching multiverse and how it enters the QM formalism. As we have demonstrated, the two parts of the problem are closely related, because any purported proof of the Born rule must appeal to premises that convey meaning to the word 'probability'. It would be a mistake to believe that, in a context where *interpretation* is at stake, a formal proof, devoid of meaning can at all explain how probability enters into the MWI.[31] In any case it must be elucidated *why*, in a fully deterministic universe which is 'all wavefunction' and in which the appearance of multiple classical worlds arises in virtue of the joint dynamics of systems and their environments, certain observations are made by conscious observers with a certain frequency. That is, it is of crucial importance to *recapture* the statistical predictions of QM within the interpretational framework set by the MWI, since they constitute its "empirical heart" [49].

We have here considered two approaches to including probabilistic content into the MWI. But we have argued, in the first case, that the assumptions which the proof takes as premises are unreasonable, given what its result postulates, and in the second case, that the result is best interpreted *not* as supporting the inclusion of probabilities into the MWI, but as one way to demonstrate that QM is an inherently probabilistic theory.

In the first case, we have followed Maudlin's objection that in all existing version of a proof of the Born rule from merely decision-theoretic premises, these must be set up such that a decisively Everettian form of rationality is forestalled from the outset, namely, caring about what happens to *all* future versions of oneself. It hence appears to be rational for Everettians to exactly *not* conform to the Born rule at times.

In the second case, we have demonstrated that questionable counting strategies are crucially involved in deriving the Born rule. And we have argued that, in virtue of the crucial role of *unitarity* and correlations in Zurek's proof, the best way to make sense of the proof and simultaneously avoid the aforementioned problems is a way that does not support the MWI at all.

But maybe all hope is not lost, and one can too make sense of counting strategies, and hence rescue one version of Zurek's proof as supporting the MWI. The suggestion would be to find a suitable natural grain, giving rise to a suitable branch count, in a way that is compatible with decoherence. One could, for instance, argue that there is such a thing as a natural grain, exactly at the level where probabilities become relevant. Namely, probabilities (whatever they 'really are') are used by conscious agents to make predictions about their future experience, and any physical system that can support consciousness requires (we may take it) a minimum of complexity. That minimum of complexity might suffice to 'naturally coarse grain' the 'wave function of the universe', i.e., might impose a superstructure onto it, allowing for a counting strategy as employed in the Zurek/Carrol and Sebens-approach. Zeh [83], in his original paper on decoherence, has suggested that the MWI corresponds to "a 'localization of consciousness' not only in space and time, but also in certain Hilbert-space components [...]." So maybe a natural grain can be provided by appeal to the minimal complexity (and hence 'macroscopicity',

[31] A similar opinion is expressed by Kent [47], who writes: "Wherever one thinks of the scientific status of many worlds quantum theory, one cannot reasonably [...] think it is so obvious how to translate equations into statements about a many-worlds reality that arguments and explanations are redundant."

implying a certain grain) that is required of a physical system to carry conscious experience. Suppose, for instance, that a few neurons were sufficient to support consciousness. Then that would still mean a tremendous number of atoms, jointly in decohered states at any instance of a conscious observer's stream of perception.

Alternatively, one attempt to could come from the natural grain *within* conscious experience. In analyzing Zurek's strategy of fine-graining the system $\mathcal{C}$ such as to keep track of the frequency of a certain outcome, we have argued that the use of real numbers must be treated as an idealization anyway, since irrational frequencies cannot be observed in any conceivable experiment. So maybe the natural grain is provided by the minimum of *coarse-graining* that is required to support a conscious perception (an actual 'outcome'). Suppose, for example, that the system is in some non-trivial superposition state $|\psi\rangle = \int \mathrm{d}\boldsymbol{x}\, \psi(\boldsymbol{x})\, |\boldsymbol{x}\rangle$, and that it is measured for its position. Then the maximal accuracy to which its position *can* be determined by a conscious observer is a kind of 'box state' $\chi_\Gamma(\boldsymbol{x})$ (with $\Gamma \subset \mathbb{R}^3$) like that used by Zurek in the treatment of probability densities, because there is a natural lower limit for conscious *distinguishability* of positions (of a visually perceivable pointer, indicative of the system's position, say). Of course, one could object here that $\mathbb{R}^3$ can be sliced up into boxes in an arbitrary way; but a non-arbitrary slicing will always be provided by the diameters of the actual experimental setup, at least once a coordinate frame is chosen.

These two perspectives on the relation between conscious experience and physical complexity or 'coarseness' constitute tentative hints in a direction that may raise hopes for invoking counting strategies after all. But it is not clear that they *are* compatible with decoherence and the 'branching structure' it implies. Rae e.g. believes that

> if the likelihood of observing a particular result is proportional to the number of associated branches, the complexity introduced by decoherence should actually result in the outcome of a measurement being completely unpredictable. The situation is similar to chaos in classical mechanics or to turbulence in hydrodynamics, whose onset certainly does not lead to increased predictability. [61]

Moreover, the suggestions above do not at all address the problem that the 'branching dynamics' defy the rules of ordinary probability calculus (cf. the example of the law of total probability in section 2.2). So in summary, it is not clear that any such strategy can be applied after all. Additionally, conscious observation would here play a fundamental role in understanding QM again, so that a central virtue of the MWI (as it is taken by many) clearly vanishes: that it removes the need to fundamentally appeal to 'consciousness' or 'observation' in the interpretation of the theory.

In the light of the forgoing arguments, we are, at present, lead to judge that the status of probabilities in the MWI remains at best unclear. But what about the other virtues that we have laid out in section 2? What about, first of all, Wallace's claim that the MWI is just QM interpreted properly? Maudlin has here argued that this suggestion is in error, for the entire success of QM rests on its application in the standard formulation (i.e., with definite outcomes and probabilities for those). But the standard formulation, Maudlin has argued,

> is a set of (somewhat vague) rules, a recipe more than a perfectly sharp algorithm, that allows one to make predictions. [...] But since the recipe does not have the

> form of a description of the world [...], there is just nothing that counts as 'taking it literally'. If someone gives me a recipe—a set of directions—for making a fluffy omelet [...], and the recipe turns out to be highly reliable and always to get good results, I might well wonder why. But it makes no sense to suggest that I can answer that question by just 'taking it literally'. [53]

Indeed, it is doubtworthy that the MWI should count as 'just QM taken literally'. The extension of unitary dynamics and Hilbert space kinematics to measuring devices and observes clearly is a step beyond the original application of the formalism, beyond the 'recipe'. And we have argued that QM may be construed as a *fundamentally* probabilistic theory, exactly in virtue of the unitarity of the dynamics.

What, however, about the application of QM to cosmological scales and problems? We mentioned, in section 2, that to some physicists the MWI is the 'natural' framework for doing quantum cosmology. But this opinion is not ubiquitously shared across physics either. For instance, Fuchs and Peres have argued that

> We have experimental evidence that quantum theory is successful in the range from 10^{-10} to 10^{15} atomic radii; we have no evidence that it is universally valid. [...] Indeed, a common question is whether the universe has a wavefunction. There are two ways to understand this. If this 'wavefunction of the universe' has to give a complete description of everything, including ourselves, we [...] get [...] meaningless paradoxes. On the other hand, if we consider just a few collective degrees of freedom, such as the radius of the universe, its mean density, total baryon number, and so on, we can apply quantum theory only to these degrees of freedom, which do not include ourselves and other insignificant details. This is not essentially different from quantizing the magnetic flux and the electric current in a SQUID while ignoring the atomic details. [29]

The point is that QM can of course always be applied to the cosmos in an *instrumental* fashion, i.e., as a kind of recipe, as Maudlin has called it, for handling certain *aspects* of the universe. To be sure, this raises many daunting questions as to why we have to resort to a formalism so strange as that of QM to describe these aspects of our universe, and in particular, how to interpret decoherence. But the philosophical implications of this should be no 'stranger' than is the MWI itself, and there are no serious 'non-strange' alternatives available.

What, then, *are* the alternatives to the MWI? We have argued above that 'the Copenhagen interpretation', beloved by physicists though it may be, is an ambiguous term, something that mostly offers a recipe as to how to use the formalism in certain well-defined situations. There are, of course, Bohmian mechanics [26] and objective collapse interpretations [31, 59], both of which also interpret QM 'realistically', i.e., as describing the world as it actually is. But these alternatives come with 'unnecessary' modifications of the formalism (so far without testable consequences) and they are just as much associated with individual difficulties as is the MWI. Other trends (such as *Qbism* [55]) attempt to make sense of the 'Copenhagen-spirit' while adopting less of a mere 'shut up and calculate'-attitude, but it remains to be seen whether this is a fruitful project either. In any case, the present investigation suggests that the MWI cannot clearly be favored among its rivals as long as the problem of probabilities persists. And, as the author hopes to have shown, at present it clearly does.

References

[1] David Albert. Probability in the Everett Picture. In Simon Saunders, Jonathan Barrett, Adrian Kent, and David Wallace, editors, *Many worlds?: Everett, quantum theory, & reality*. OUP Oxford, 2010.

[2] David Albert and Barry Loewer. Interpreting the many worlds interpretation. *Synthese*, 77(2):195–213, 1988.

[3] M. Arndt, O. Nairz, J. Petschinka, and A. Zeilinger. High contrast interference with C_{60} and C_{70}. *Comptes Rendus de l'Académie des Sciences-Series IV-Physics*, 2(4):581–585, 2001.

[4] Alain Aspect. Introduction: John Bell and the second quantum revolution. In J. S. Bell, editor, *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press, second edition, 2004.

[5] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental Test of Bell's Inequalities Using Time-Varying Analyzers. *Physical Review Letters*, 49(25):1804–1807, 1982.

[6] Howard Barnum. No-signalling-based version of zurek's derivation of quantum probabilities: A note on "environment-assisted invariance, entanglement, and probabilities in quantum physics". *arXiv preprint quant-ph/0312150*, 2003.

[7] J.A. Barrett. *The Quantum Mechanics of Minds and Worlds*. OUP Oxford, 1999.

[8] Jeffrey A. Barrett. On the Nature of Measurement Records in Relativistic Quantum Field Theory. In Meinard Kuhlmann, Holger Lyre, and Andrew Wayne, editors, *Ontological Aspects of Quantum Field Theory*, pages 165–180. New Jersey, London: World Scientific Publishing, 2002.

[9] Jean-Louis Basdevant and Jean Dalibard. *Quantum Mechanics*. Berlin, Heidelberg: Springer, second edition, 2005.

[10] J.S. Bell. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*. Cambridge University Press, second edition, 2004.

[11] J. Binney and D. Skinner. *The Physics of Quantum Mechanics*. Oxford, New York: Oxford University Press, 2014.

[12] F. Boge. On Modern Approaches to the Einsteinian View of Quantum States. *philsci preprint 12013*, 2016.

[13] Michel Brune, E Hagley, J Dreyer, X Maitre, A Maali, C Wunderlich, JM Raimond, and S Haroche. Observing the progressive decoherence of the "meter" in a quantum measurement. *Physical Review Letters*, 77(24):4887, 1996.

[14] J. Bub. *Interpreting the Quantum World*. Cambridge University Press, 1999.

[15] Paul Busch, Pekka J. Lahti, and Peter Mittelstaedt. *The Quantum Theory of Measurement*. Lecture Notes in Physics. Berlin, Heidelberg: Springer, second revised edition, 1996.

[16] Sean M Carroll and Charles T Sebens. Many worlds, the born rule, and self-locating uncertainty. In *Quantum Theory: A Two-Time Success Story*, pages 157–169. Springer, 2014.

[17] C. M. Caves. Notes on Zurek's derivation of the quantum probability rule. Online: info.phys.unm.edu/ caves/reports/ZurekBornderivation.ps. Last modified 2005 July 29.

[18] Richard Dawid and Karim PY Thébault. Many worlds: decoherent or incoherent? *Synthese*, 192(5):1559–1580, 2015.

[19] Rafael de la Madrid. The role of the rigged Hilbert space in Quantum Mechanics. *arXiv preprint quant-ph/0502053*, 2005.

[20] Bernard d'Espagnat. *Conceptual Foundations of Quantum Mechanics*. Reading, Massachusttes: Perseus Books, second edition, 1999.

[21] David Deutsch. Quantum theory of probability and decisions. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 455, pages 3129–3137. The Royal Society, 1999.

[22] Bryce S DeWitt. Quantum theory of gravity. i. the canonical theory. *Physical Review*, 160(5):1113, 1967.

[23] Bryce S. DeWitt. Quantum Mechanics and Reality. In B.S. Dewitt and N. Graham, editors, *The Many Worlds Interpretation of Quantum Mechanics*, Princeton Series in Physics, pages 155–165. Princeton University Press, 1973.

[24] B.S. Dewitt and N. Graham, editors. *The Many Worlds Interpretation of Quantum Mechanics*. Princeton Series in Physics. Princeton University Press, 1973.

[25] P. A. M. Dirac. *The Principles of Quantum Mechanics*. Oxford: Clarendon Press, fourth (revised) edition, 1958.

[26] D. Dürr, S. Goldstein, and N. Zanghì. *Quantum Physics Without Quantum Philosophy*. Springer Berlin Heidelberg, 2012.

[27] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, 47:777–780, 1935.

[28] Hugh Everett III. "Relative state" formulation of quantum mechanics. *Reviews of modern physics*, 29(3):454, 1957.

[29] C. A. Fuchs and A. Peres. Quantum theory needs no 'interpretation'. *Physics Today*, 53(3):70–71, 2000.

[30] Murray Gell-Mann and James B Hartle. Quantum mechanics in the light of quantum cosmology. *Complexity, entropy and the physics of information*, 8, 1990.

[31] Gian Carlo Ghirardi, Alberto Rimini, and Tullio Weber. Unified dynamics for microscopic and macroscopic systems. *Physical Review D*, 34(2):470, 1986.

[32] Domenico Giulini. Superselection Rules. In Daniel Greenberger, Klaus Hentschel, and Friedel Weinert, editors, *Compendium of Quantum Physics. Concepts, Experiments, History and Philosophy*, pages 771–779. Berlin, Heidelberg: Springer, 2009.

[33] H Greaves and D Wallace. Empirical consequences of symmetries. *The British Journal for the Philosophy of Science*, page axt005, 2013.

[34] Hilary Greaves. Understanding Deutsch's probability in a deterministic multiverse. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 35(3):423–456, 2004.

[35] Hilary Greaves. Probability in the everett interpretation. *Philosophy Compass*, 2(1):109–128, 2007.

[36] T. Heinosaari and M. Ziman. *The Mathematical Language of Quantum Theory. From Uncertainty to Entanglement*. Cambridge, New York: Cambridge University Press, 2012.

[37] Werner Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3-4):172–198, 1927.

[38] Werner Heisenberg. *Physics and Philosophy. The Revolution in Modern Science*. London: Allen & Unwin Ltd., third edition, 1971 [1958].

[39] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of modern physics*, 81(2):865, 2009.

[40] Don Howard. Who Invented the 'Copenhagen Interpretation'? A Study in Mythology. *Philosophy of Science*, 71(5):669–682, 2002.

[41] R. I. G. Hughes. *The Structure and Interpretation of Quantum Mechanics*. Cambridge: Harvard University Press, 1989.

[42] Gregg Jaeger. *Quantum Information. An Overview*. New York: Springer Science+Business Media, LLC, 2007.

[43] E. T. Jaynes. *Probability Theory. The Logic of Science*. Cambridge, New York: Cambridge University Press, 2003.

[44] David Jennings and Matthew Leifer. No return to classical reality. *Contemporary Physics*, 57(1):60–82, 2015.

[45] William Johnston. The Weighted Hermite Polynomials Form a Basis for $L^2(\mathbb{R})$. *The American Mathematical Monthly*, 121(3):249–253, 2014.

[46] E. Joos, H.D. Zeh, C. Kiefer, D. Giulini, J. Kupsch, and I.-O. Stamatescu. *Decoherence and the Appearance of a Classical World in Quantum Theory*. Berlin, Heidelberg: Springer, second edition, 2003.

[47] Adrian Kent. Does it Make Sense to Speak of Self-Locating Uncertainty in the Universal Wave Function? Remarks on Sebens and Carroll. *Foundations of Physics*, 45(2):211–217, 2015.

[48] Daniel Kleppner and Roman Jackiw. One Hundred Years of Quantum Physics. *Science*, 289(5481):893–898, 2000.

[49] Peter J Lewis. Quantum sleeping beauty. *Analysis*, 67(293):59–65, 2007.

[50] Michael Lockwood. 'Many Minds' Interpretations of Quantum Mechanics. *The British journal for the philosophy of science*, 47(2):159–188, 1996.

[51] Gerhart Lüders. Über die Zustandsänderung durch den Messprozess. *Annalen der Physik*, 443(5-8):322–328, 1951.

[52] Tim Maudlin. *Quantum Non-Locality and Relativity. Metaphysical Intimations of Modern Physics*. Mladen, Oxford: Wiley-Blackwell, third edition, 2011.

[53] Tim Maudlin. Critical Study: David Wallace, The Emergent Multiverse: Quantum Theory According to the Everett Interpretation. *Noûs*, 48(4):794–808, 2014.

[54] N. David Mermin. What's Wrong With This Pillow? *Physics Today*, 42(4):9, 1989.

[55] N. David Mermin. Why QBism is not the Copenhagen interpretation and what John Bell might have thought of it. *arXiv preprint 1409.2454v1*, 2014.

[56] Peter Mittelstaedt. *The Interpretation of Quantum Mechanics and the Measurement Process*. Cambridge, New York: Cambridge University Press, 1998.

[57] Ulrich Mohrhoff. Probabilities from envariance? *International Journal of Quantum Information*, 2(02):221–229, 2004.

[58] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge, New York: Cambridge University Press, 10th anniversary edition, 2010.

[59] Roger Penrose. On gravity's role in quantum state reduction. *General relativity and gravitation*, 28(5):581–600, 1996.

[60] Asher Peres. *Quantum Theory: Concepts and Methods*, volume 72 of *Fundamental Theories of Physics*. New York, Boston: Kluwer, 2002.

[61] Alastair IM Rae. Everett and the Born rule. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 40(3):243–250, 2009.

[62] George G. Roussas. *An Introduction to Measure-Theoretic Probability*. Amsterdam, Boston: Elsevier, second edition, 2014.

[63] G.G. Roussas. *Introduction to Probability*. Elsevier Academic Press, 2007.

[64] Simon Saunders. Time, quantum mechanics, and probability. *Synthese*, 114(3):373–404, 1998.

[65] L.J. Savage. *The Foundations of Statistics*. Dover Publications, second edition, 1972.

[66] M. Schlosshauer. *Decoherence and the Quantum to Classical Transition*. Berlin, Heidelberg: Springer, second edition, 2007.

[67] Maximilian Schlosshauer and Arthur Fine. On Zurek's derivation of the Born rule. *Foundations of Physics*, 35(2):197–213, 2005.

[68] Maximilian Schlosshauer, Johannes Kofler, and Anton Zeilinger. A snapshot of foundational attitudes toward quantum mechanics. *Studies in History and Philosophy of Modern Physics*, 44(3):222–230, 2013.

[69] Erwin Schrödinger. The Present Situation in Quantum Mechanics. A Translation of Schrödinger's "Cat Paradox" Paper. In John Archibald Wheeler and Wojciech Hubert Zurek, editors, *Quantum Theory and Measurement*. Princeton, New Jersey: Princeton University Press, 1983 [1935]. Translated by John D. Trimmer.

[70] Gerhard Schurz. *Philosophy of Science. A Unified Approach*. New York, London: Routledge, 2014.

[71] Franz Schwabl. *Quantum Mechanics*. Berlin, Heidelberg: Springer, fourth edition, 2007.

[72] Max Tegmark. The Interpretation of Quantum Mechanics: Many Worlds or Many Words? *Fortschritte der Physik*, 46(6-8):855–862, 1998.

[73] Paul Teller. *An Interpretive Introduction to Quantum Field Theory*. Princeton, New Jersey: Princeton University Press, 1995.

[74] Lev Vaidman. Probability in the many-worlds interpretation of quantum mechanics. In Yemima Ben-Menahem and Meir Hemmo, editors, *Probability in physics*, pages 299–311. Springer, 2012.

[75] R. von Mises. *Mathematical Theory of Probability and Statistics*. New York: Dover Publications Inc., second revised edition, 1981 [1957].

[76] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, third edition, 1953.

[77] John von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press, 1955 [1932]. Translated by Robert T. Beyer.

[78] D. Wallace. Quantum probability and decision theory, revisited. *arXiv preprint quant-ph/0211104*, 2002.

[79] D. Wallace. Everettian rationality: defending deutsch's approach to probability in the everett interpretation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 34(3):415–439, 2003.

[80] D. Wallace. Quantum probability from subjective likelihood: improving on deutsch's proof of the probability rule. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 38(2):311–332, 2007.

[81] D. Wallace. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. OUP Oxford, 2012.

[82] E. Wigner. Remarks on the mind-body question. In John Heil, editor, *Philosophy of Mind: A Guide and Anthology*, pages 866–877. Oxford, New York: Oxford University Press, 2004 [1967].

[83] H Dieter Zeh. On the interpretation of measurement in quantum theory. *Foundations of Physics*, 1(1):69–76, 1970.

[84] Wojciech H Zurek. Environment-induced superselection rules. *Physical Review D*, 26(8):1862, 1982.

[85] Wojciech H Zurek. Decoherence and the transition from quantum to classical-revisited. *Los Alamos Science*, 27:86–109, 2002.

[86] Wojciech Hubert Zurek. Environment-assisted invariance, entanglement, and probabilities in quantum physics. *Physical review letters*, 90(12):120404, 2003.

[87] Wojciech Hubert Zurek. Probabilities from entanglement, Born's rule $p_k = |\psi_k|^2$ from envariance. *Physical Review A*, 71(5):052105, 2005.