The Importance of Quantum Decoherence in Brain Processes

Max Tegmark
Institute for Advanced Study, Olden Lane, Princeton, NJ 08540; max@ias.edu
Dept. of Physics, Univ. of Pennsylvania, Philadelphia, PA 19104
(Submitted to Phys. Rev. E July 2 1999, accepted October 25)

Based on a calculation of neural decoherence rates, we argue that that the degrees of freedom of the human brain that relate to cognitive processes should be thought of as a classical rather than quantum system, *i.e.*, that there is nothing fundamentally wrong with the current classical approach to neural network simulations. We find that the decoherence timescales ($\sim 10^{-13} - 10^{-20}$ seconds) are typically much shorter than the relevant dynamical timescales ($\sim 10^{-3} - 10^{-1}$ seconds), both for regular neuron firing and for kink-like polarization excitations in microtubules. This conclusion disagrees with suggestions by Penrose and others that the brain acts as a quantum computer, and that quantum coherence is related to consciousness in a fundamental way.

I. INTRODUCTION

In most current mainstream biophysics research on cognitive processes, the brain is modeled as a neural network obeying classical physics. In contrast, Penrose [1,2], and others have argued that quantum mechanics may play an essential role, and that successful brain simulations can only be performed with a quantum computer. The main purpose of this paper is to address this issue with quantitative decoherence calculations.

The field of artificial neural networks (for an introduction, see, *e.g.*, [4–6]) is currently booming, driven by a broad range of applications and improved computing resources. Although the popular neurological models come in various levels of abstraction, none involve effects of quantum coherence in any fundamental way. Encouraged by successes in modeling memory, learning, visual processing, *etc.* [7,8], many workers in the field have boldly conjectured that a sufficiently complex neural network could in principle perform all cognitive processes that we associate with consciousness.

On the other hand, many authors have argued that consciousness can only be understood as a quantum effect. For instance, Wigner [9] suggested that consciousness was linked to the quantum measurement problem[1], and this idea has been greatly elaborated by Stapp [3]. There have been numerous suggestions that consciousness is a macroquantum effect, involving superconductivity [12], superfluidity [13], electromagnetic fields [14], Bose condensation [15,16], superflourescence [17] or some other mechanism [18,19]. Perhaps the most concrete one is that of Penrose [2], proposing that this takes place in microtubules, the ubiquitous hollow cylinders that among other things help cells maintain their shapes. It has been argued that microtubules can process information like a cellular automaton [20], and Penrose suggests that they operate as a quantum computer. This idea has been further elaborated employing string theory methods [21–27].

The make-or-break issue for all these quantum models is whether the relevant degrees of freedom of the brain can be sufficiently isolated to retain their quantum coherence, and opinions are divided. For instance, Stapp has argued that interaction with the environment is probably small enough to be unimportant for certain neural processes [28], whereas Zeh [29], Zurek [30], Scott [31], Hawking [32] and Hepp [33] have conjectured that environment-induced coherence will rapidly destroy macrosuperpositions in the brain. It is therefore timely to try to settle the issue with detailed calculations of the relevant decoherence rates. This is the purpose of the present work.

The rest of this paper is organized as follows. In Section II, we briefly review the open system quantum mechanics necessary for our calculations, and introduce a decomposition into three subsystems to place the problem in its proper context. In Section III, we evaluate decoherence rates both for neuron firing and for the microtubule processes proposed by Penrose *et al.*, relegating some technical details to the Appendix. We conclude in Section IV by discussing the implications of our results, both for modeling cognitive brain processes and for incorporating them into a quantum-mechanical treatment of the rest of the world.

II. SYSTEMS AND SUBSYSTEMS

In this section, we review those aspects of quantum mechanics for open systems that are needed for our calculations, and introduce a classification scheme and a subsystem decomposition to place the problem at hand in its appropriate context.

[1] Interestingly, Wigner changed his mind and gave up this idea [10] after he became aware in of the first paper on decoherence in 1970 [11].

A. Notation

Let us first briefly review the quantum mechanics of subsystems. The state of an arbitrary quantum system is described by its density matrix ρ, which left in isolation will evolve in time according to the Schrödinger equation

$$\dot{\rho} = -i[H, \rho]/\hbar. \tag{1}$$

It is often useful to view a system as composed of two subsystems, so that some of the degrees of freedom correspond to the 1^{st} and the rest to the 2^{nd}. The state of subsystem i is described by the reduced density matrix ρ_i obtained by tracing (marginalizing) over the degrees of freedom of the other: $\rho_1 \equiv \mathrm{tr}\,_2\rho$, $\rho_2 \equiv \mathrm{tr}\,_1\rho$. Let us decompose the Hamiltonian as

$$H = H_1 + H_2 + H_{int}, \tag{2}$$

where the operator H_1 affects only the 1^{st} subsystem and H_2 affects only the 2^{nd} subsystem. The interaction Hamiltonian H_{int} is the remaining nonseparable part, defined as $H_{int} \equiv H - H_1 - H_2$, so such a decomposition is always possible, although it is generally only useful if H_{int} is in some sense small.

If $H_{\mathrm{int}} = 0$, *i.e.*, if there is no interaction between the two subsystems, then it is easy to show that $\dot{\rho}_i = -i[H_i, \rho_i]/\hbar$, $i = 1, 2$, that is, we can treat each subsystem as if the rest of the Universe did not exist, ignoring any correlations with the other subsystem that may have been present in the full non-separable density matrix ρ. It is of course this property that makes density matrices so useful in the first place, and that led von Neumann to invent them [34]: the full system is assumed to obey equation (1) simply because its interactions with the rest of the Universe are negligible.

B. Fluctuation, dissipation, communication and decoherence

In practice, the interaction H_{int} between subsystems is usually not zero. This has a number of qualitatively different effects:

1. Fluctuation
2. Dissipation
3. Communication
4. Decoherence

The first two involve transfer of energy between the subsystems, whereas the last two involve exchange of information. The first three occur in classical physics as well - only the last one is a purely quantum-mechanical phenomenon.

For example, consider a tiny colloid grain (subsystem 1) in a jar of water (subsystem 2). Collisions with water molecules will cause **fluctuations** in the center-of-mass position of the colloid (brownian motion). If its initial velocity is high, **dissipation** (friction) will slow it down to a mean speed corresponding to thermal equilibrium with the water. The dissipation timescale τ_{diss}, defined as the time it would take to lose half of the initial excess energy, will in this case be of order $\tau_{\mathrm{coll}} \times (M/m)$, where τ_{coll} is the mean-free time between collisions, M the colloid mass M and m is the mass of a water molecule. We will define **communication** as exchange of information. The information that the two subsystems have about each other, measured in bits, is

$$I_{12} \equiv S_1 + S_2 - S, \tag{3}$$

where $S_i \equiv -\mathrm{tr}\,_i\rho_i \log\rho_i$ is the entropy of the i^{th} subsystem, $S \equiv -\mathrm{tr}\,\rho\log\rho$ is the entropy of the total system, and the logarithms are base 2. If this mutual information is zero, then the states of the two systems are uncorrelated and independent, with the density matrix of the separable form $\rho = \rho_1 \otimes \rho_2$. If the subsystems start out independent, any interaction will at least initially increase the subsystem entropies S_i, thereby increasing the mutual information, since the entropy S of the total system always remains constant.

This apparent entropy increase of subsystems, which is related to the arrow of time and the 2^{nd} law of of thermodynamics [35], occurs also in classical physics. However, quantum mechanics produces a qualitatively new effect as well, known as **decoherence** [11,36,37], suppressing off-diagonal elements in the reduced density matrices ρ_i. This effect destroys the ability to observe long-range quantum superpositions within the subsystems, and is now rather well-understood and uncontroversial [30,38–42] – the interested reader is referred to [43] and a recent book on decoherence [44] for details. For instance, if our colloid was initially in a superposition of two locations separated by a centimeter, this macrosuperposition would for all practical purposes be destroyed by the first collision with a water molecule, *i.e.*, on a timescale τ_{dec} of order τ_{coll}, with the quantum superposition surviving only on scales below the de de Broigle wavelength of the water molecules [45,46].[2] This means

[2]Decoherence picks out a preferred basis in the quantum-mechanical Hilbert space, termed the "pointer basis" by Zurek [36], in which superpositions are rapidly destroyed and classical behavior is approached. This normally includes the position basis, which is why we never experience superpositions of objects in macroscopically different positions. Decoherence is quite generic. Although it has been claimed that this preferred basis consists of the maximal set of commuting observables that also commute with H_{int} (the "microstable basis" of Omnes [43]), this is in fact merely a sufficient condition, not a necessary one. If $[H_{\mathrm{int}}, x] = 0$ for some observable x but $[H_{\mathrm{int}}, p] \neq 0$ for its conjugate p, then the interaction

that $\tau_{\rm diss}/\tau_{\rm dec} \sim M/m$ in our example, *i.e.*, that decoherence is much faster than dissipation for macroscopic objects, and this qualitative result has been shown to hold quite generally as well (see [43] and references therein). Loosely speaking, this is because each microscopic particle that scatters off of the subsystem carries away only a tiny fraction m/M of the total momentum, but essentially all of the necessary information.

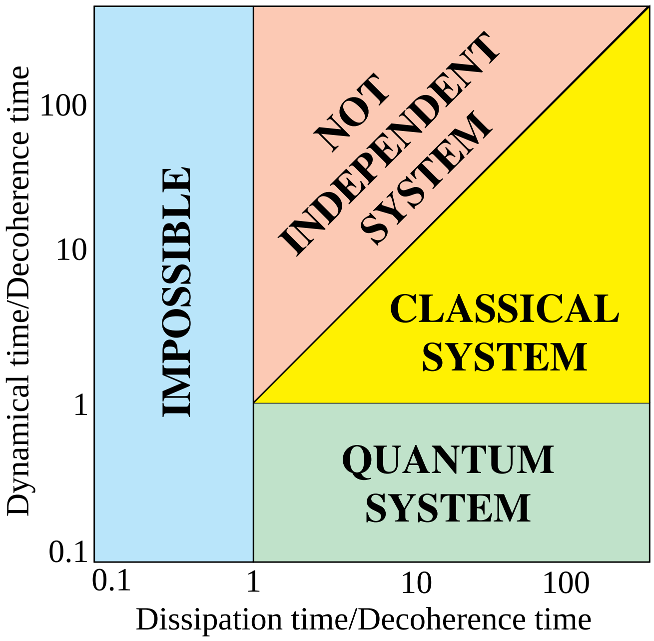

FIG. 1. The qualitative behavior of a subsystem depends on the timescales for dynamics, dissipation and decoherence. This classification is by necessity quite crude, so the boundaries should not be thought of as sharp.

C. Classification of systems

Let us define the *dynamical timescale* $\tau_{\rm dyn}$ of a subsystem as that which is characteristic of its internal dynamics. For a planetary system or an atom, $\tau_{\rm dyn}$ would be the orbital frequency.

The qualitative behavior of a system depends on the ratio of these timescales, as illustrated in Figure 1. If $\tau_{\rm dyn} \ll \tau_{\rm dec}$, we are are dealing with a true quantum system, since its superpositions can persist long enough to be dynamically important. If $\tau_{\rm dyn} \gg \tau_{\rm diss}$, it is hardly meaningful to view it as an independent system at all, since its internal forces are so week that they are dwarfed by the effects of the surroundings. In the intermediate case where $\tau_{\rm dec} \ll \tau_{\rm dyn} \lesssim \tau_{\rm diss}$, we have a familiar classical system.

The relation between $\tau_{\rm dec}$ and $\tau_{\rm diss}$ depends only on the *form* of $H_{\rm int}$, whereas the question of whether $\tau_{\rm dyn}$ falls between these values depends on the *normalization* of $H_{\rm int}$ in equation (2). Since $\tau_{\rm dec} \sim \tau_{\rm diss}$ for microscopic (atom-sized) systems and $\tau_{\rm dec} \ll \tau_{\rm diss}$ for macroscopic ones, Figure 1 shows that whereas macroscopic systems can behave quantum-mechanically, microscopic ones can never behave classically.

D. Three systems: subject, object and environment

Most discussions of quantum statistical mechanics split the Universe into two subsystems [47]: the object under consideration and everything else (referred to as the *environment*). Since our purpose is to model the observer, we need to include a third subsystem as well, the *subject*. As illustrated in Figure 2, we therefore decompose the total system into three subsystems:

- The **subject** consists of the degrees of freedom associated with the subjective perceptions of the observer. This does not include any other degrees of freedom associated with the brain or other parts of the body.
- The **object** consists of the degrees of freedom that the observer is interested in studying, *e.g.*, the pointer position on a measurement apparatus.
- The **environment** consists of everything else, *i.e.*, all the degrees of freedom that the observer is not paying attention to. By definition, these are the degrees of freedom that we always perform a partial trace over.

---

will indeed cause decoherence for x as advertised. But this will happen even if $[H_{\rm int}, x] \neq 0$ — all that matters is that $[H_{\rm int}, p] \neq 0$, *i.e.*, that the interaction Hamiltonian contains ("measures") x.

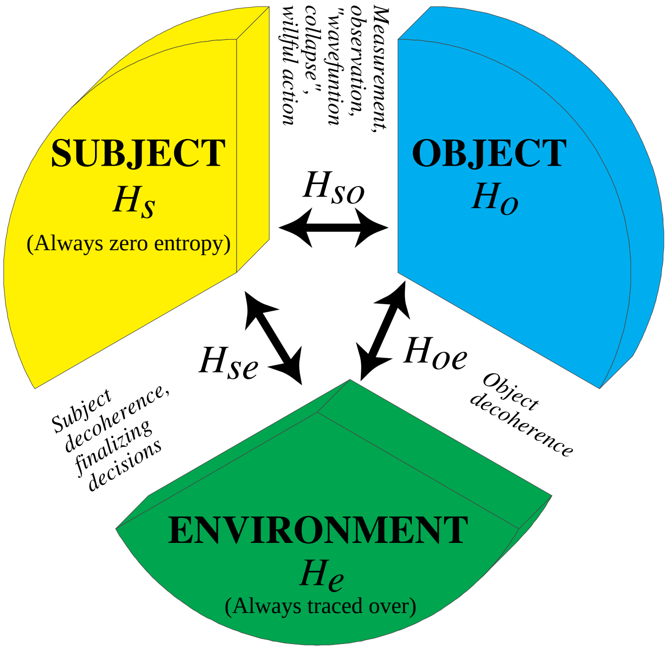

FIG. 2. An observer can always decompose the world into three subsystems: the degrees of freedom corresponding to her subjective perceptions (the subject), the degrees of freedom being studied (the object), and everything else (the environment). As indicated, the subsystem Hamiltonians H_s, H_o, H_e and the interaction Hamiltonians H_{so}, H_{oe}, H_{se} can cause qualitatively very different effects, which is why it is often useful to study them separately. This paper focuses on the interaction H_{se}.

Note that the first two definitions are very restrictive. Whereas the subject would include the entire body of the observer in the common way of speaking, only very few degrees of freedom qualify as our subject or object. For instance, if a physicist is observing a Stern-Gerlach apparatus, the vast majority of the $\sim 10^{28}$ degrees of freedom in the the observer and apparatus are counted as environment, not as subject or object.

The term "perception" is used in a broad sense in item 1, including thoughts, emotions and any other attributes of the subjectively perceived state of the observer.

The practical usefulness in this decomposition lies in that one can often neglect everything except the object and its internal dynamics (given by H_o) to first order, using simple prescriptions to correct for the interactions with the subject and the environment. The effects of both H_{so} and H_{oe} have been extensively studied in the literature. H_{so} involves quantum measurement, and gives rise to the usual interpretation of the diagonal elements of the object density matrix as probabilities. H_{oe} produces decoherence, selecting a preferred basis and making the object act classically if the conditions in Figure 1 are met.

In contrast, H_{se}, which causes decoherence directly in the subject system, has received relatively little attention. It is the focus of the present paper, and the next section is devoted to quantitative calculations of decoherence in brain processes, aimed at determining whether the subject system should be classified as classical or quantum in the sense of Figure 1. We will return to Figure 2 and a more detailed discussion of its various subsystem interactions in Section IV.

III. DECOHERENCE RATES

In this section, we will make quantitative estimates of decoherence rates for neurological processes. We first analyze the process of neuron firing, widely assumed to be central to cognitive processes. We also analyze electrical excitations in microtubules, which Penrose and others have suggested may be relevant to conscious thought.

A. Neuron firing

Neurons (see Figure 3) are one of the key building blocks of the brain's information processing system. It is widely believed that the complex network of $\sim 10^{11}$ neurons with their nonlinear synaptic couplings is in some way linked to our subjective perceptions, *i.e.*, to the subject degrees of freedom. If this picture is correct, then if H_s or H_{so} puts the subject into a superposition of two distinct mental states, some neurons will be in a superposition of firing and not firing. How fast does such a superposition of a firing and non-firing neuron decohere?

Let us consider this process in more detail. For introductory reviews of neuron dynamics, the reader is referred to, *e.g.*, [48–50]. Like virtually all animal cells, neurons have ATP driven pumps in their membranes which push sodium ions out of the cell into the surrounding fluids and potassium ions the other way. The former process is slightly more efficient, so the neuron contains a slight excess of negative charge in its "resting" state, corresponding to a potential difference $U_0 \approx -0.07\,\mathrm{V}$ across the axon membrane ("axolemma"). There is an inherent instability in the system, however. If the potential becomes substantially less negative, then voltage-gated sodium channels in the axon membrane open up, allowing Na^+ ions to come gushing in. This makes the potential still less negative, causes still more opening, *etc.* This chain reaction, "firing", propagates down the axon at a speed of up to 100 m/s, changing the potential difference to a value U_1 that is typically of order $+0.03\,\mathrm{V}$ [49].

The axon quickly recovers. After less than ~ 1 ms, the sodium channels close regardless of the voltage, and large potassium channels (also voltage gated, but with a time delay) open up allowing K^+ ions to flow out and restore the resting potential U_0. The ATP driven pumps quickly restore the Na^+ and K^+ concentrations to their initial values, making the neuron ready to fire again if triggered. Fast neurons can fire over 10^3 times per second.

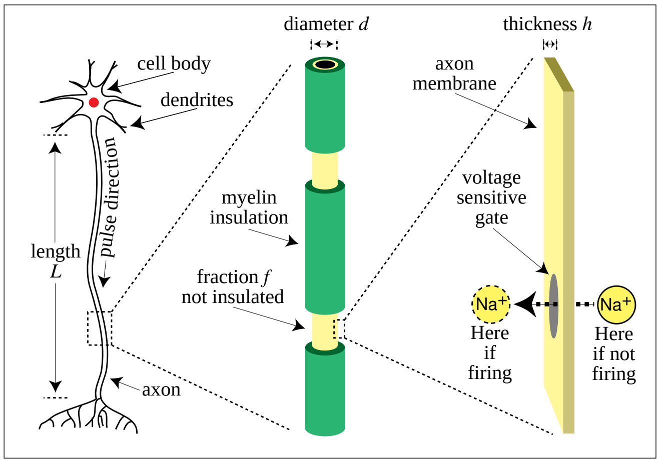

FIG. 3. Schematic illustration of a neuron (left), a section of the myelinated axon (center) and and a piece of its axon membrane (right). The axon is typically insulated (myelinated) with small bare patches every 0.5 mm or so (so-called Nodes of Ranvier) where the voltage-sensitive sodium and potassium gates are concentrated [51,52]. If the neuron is in a superposition of firing and not firing, then $N \sim 10^6$ Na^+ ions are in a superposition of being inside and outside the cell (right).

Consider a small patch of the membrane, assumed to be roughly flat with uniform thickness h as in Figure 3. If there is an excess surface density $\pm\sigma$ of charge near the inside/outside membrane surfaces, giving a voltage differential U across the membrane, then application of Gauss' law tells us that $\sigma = \epsilon_0 E$, where the electric field strength in the membrane is $E = U/h$ and ϵ_0 is the vacuum permittivity. Consider an axon of length L and diameter d, with a fraction f of its surface area bare (not insulated with myelin). The total active surface area is thus $A = \pi dLf$, so the total number of Na^+ ions that migrate in during firing is

$$N = \frac{A\sigma}{q} = \frac{\pi dLf\epsilon_0(U_1 - U_0)}{qh}, \tag{4}$$

where q is the ionic charge ($q = q_e$, the absolute value of the electron charge). Taking values typical for central nervous system axons [52,53], $h = 8\,nm$, $d = 10\,\mu\text{m}$, $L = 10\,\text{cm}$, $f = 10^{-3}$, $U_0 = -0.07\,\text{V}$ and $U_1 = +0.03\,\text{V}$ gives $N \approx 10^6$ ions, and reasonable variations in our parameters can change this number by a few orders of magnitude.

B. Neuron decoherence mechanisms

Above we saw that a quantum superposition of the neuron states "resting" and "firing" involves of order a million ions being in a spatial superposition of inside and outside the axon membrane, separated by a distance of order $h \sim 10\,\text{nm}$. In this subsection, we will compute the timescale on which decoherence destroys such a superposition.

In this analysis, the object is the neuron, and the superposition will be destroyed by any interaction with other (environment) degrees of freedom that is sensitive to where the ions are located. We will consider the following three sources of decoherence for the ions:

1. Collisions with other ions
2. Collisions with water molecules
3. Coloumb interactions with more distant ions

There are many more decoherence mechanisms [44–46]. Exotic candidates such as quantum gravity [54] and modified quantum mechanics [55,56] are generally much weaker [46]. A number of decoherence effects may be even stronger than those listed, *e.g.*, interactions as the ions penetrate the membrane — the listed effects will turn out to be so strong that we can make our argument by simply using them as lower limits on the actual decoherence rate.

Let ρ denote the density matrix for the position $\mathbf{r}$ of a single Na^+ ion. As reviewed in the Appendix, all three of the listed processes cause ρ to evolve as

$$\rho(\mathbf{x}, \mathbf{x}', t_0 + t) = \rho(\mathbf{x}, \mathbf{x}', t_0) f(\mathbf{x}, \mathbf{x}', t) \tag{5}$$

for some function f that is independent of the ion state ρ and depends only on the interaction Hamiltonian H_{int}. This assumes that we can neglect the motion of the ion itself on the decoherence timescale — we will see that this condition is met with a broad margin.

1. Ion–ion collisions

For scattering of environment particles (processes 1 and 2) that have a typical de Broigle wavelength λ, we have [46]

$$f(\mathbf{x}, \mathbf{x}', t) = e^{-\Lambda t\left(1 - e^{-|\mathbf{x}' - \mathbf{x}|^2/2\lambda^2}\right)} \approx \begin{cases} e^{-|\mathbf{x}' - \mathbf{x}|^2 \Lambda t/2\lambda^2} & \text{for } |\mathbf{x}' - \mathbf{x}| \ll \lambda, \\ e^{-\Lambda t} & \text{for } |\mathbf{x}' - \mathbf{x}| \gg \lambda. \end{cases} \tag{6}$$

Here Λ is the scattering rate, given by $\Lambda \equiv n\langle\sigma v\rangle$, where n is the density of scatterers, σ is the scattering cross section and v is the velocity. The product σv is averaged over a the velocity distribution, which we take to be a thermal (Boltzmann) distribution for corresponding to $T = 37°\text{C} \approx 310\,\text{K}$. The gist of equation (6) is that a single collision decoheres the ion down to the de Broigle wavelength of the scattering particle. The information I_{12} communicated during the scattering is $I_{12} \sim \log_2(\Delta x/\lambda)$ bits, where Δx is the initial spread in the position of our particle.

Since the typical de Broigle wavelength of a Na^+ ion (mass $m \approx 23 m_p$) or H_2O molecule ($m \approx 18 m_p$) is

$$\lambda = \frac{2\pi\hbar}{\sqrt{3mkT}} \approx 0.03\,\mathrm{nm} \tag{7}$$

at 310K, way smaller than the the membrane thickness $h \sim 10\,\mathrm{nm}$ over which we need to maintain quantum coherence, we are clearly in the $|\mathbf{x}' - \mathbf{x}| \gg \lambda$ limit of equation (6). This means that the spatial superposition of an ion decays exponentially Λ^{-1}, of order its mean free time between collisions. Since the superposition of the neuron states "resting" and "firing" involves N such superposed ions, it thus gets destroyed on a timescale $\tau \equiv (N\Lambda)^{-1}$.

Let us now evaluate τ. Coulomb scattering between two ions of unit charge gives substantial deflection angles ($\theta \sim 1$) with a cross section or order[3]

$$\sigma \sim \left(\frac{gq^2}{mv^2}\right)^2, \tag{9}$$

where v is the relative velocity and $g \equiv 1/4\pi\epsilon_0$ is the Coulomb constant. In thermal equilibrium, the kinetic energy $mv^2/2$ is of order kT, so $v \sim \sqrt{kT/m}$. For the ion density, let us write $n = \eta n_{H_2O}$, where the density of water molecules n_{H_2O} is about $(1\,\mathrm{g/cm^3})/(18m_p) \sim 10^{23}/\mathrm{cm}^3$ and η is the relative concentration of ions (positive and negative combined). Typical ion concentrations during the resting state are $[\mathrm{Na}^+]$ =9.2 (120) mmol/l and $[\mathrm{K}^+]$ =140 (2.5) mmol/l inside (outside) the axon membrane [48], corresponding to total $\mathrm{Na}^+ + \mathrm{K}^+$ concentrations of $\eta \approx 0.00027$ (0.00022) inside (outside). To be conservative, we will simply use $\eta \approx 0.0002$ throughout. Ion–ion collisions therefore destroy the superposition on a timescale

$$\tau \sim \frac{1}{Nn\sigma v} \sim \frac{\sqrt{m(kT)^3}}{Ng^2q_e^4 n} \sim 10^{-20}\,\mathrm{s}. \tag{10}$$

2. Ion–water collisions

Since H_2O molecules are electrically neutral, the cross-section is dominated by their electric dipole moment $p \approx 1.85\,\mathrm{Debye} \approx (0.0385\,\mathrm{nm}) \times q_e$. We can model this dipole as two opposing unit charges separated by a distance $y \equiv p/q_e \ll b$, so summing the two corresponding contributions from equation (8) gives a deflection angle

$$\theta \approx \frac{2gq_e p}{mv^2b^2}. \tag{11}$$

This gives a cross section

$$\sigma = \pi b^2 \sim \frac{gq_e p}{mv^2}. \tag{12}$$

for scattering with large ($\theta \sim 1$) deflections. Although σ is smaller than for the case of ion–ion collisions, n is larger because the concentration factor η drops out, giving a final result

$$\tau \sim \frac{1}{Nn\sigma v} \sim \frac{\sqrt{mkT}}{Ngq_e pn} \sim 10^{-20}\,\mathrm{s} \tag{13}$$

3. Interactions with distant ions

As shown in the Appendix, long-range interaction with a distant (environment) particle gives

$$f(\mathbf{r}, \mathbf{r}', t) = \widehat{p}_2\left[\mathbf{M}(\mathbf{r}' - \mathbf{r})t/\hbar\right], \tag{14}$$

up to a phase factor that is irrelevant for decoherence. Here $\widehat{p}_2$ is the Fourier transform of $p_2(\mathbf{r}) \equiv \rho_2(\mathbf{r}, \mathbf{r})$, the probability distribution for the location of the environment particle. M is the 3×3 Hessian matrix of second derivatives of the interaction potential of the two particles at their mean separation. A slightly less general formula was derived in the seminal paper [45]. For roughly thermal states, ρ_2 (and thus p) is likely to be well approximated by a Gaussian [58,59]. This gives

$$f(\mathbf{r}, \mathbf{r}', t) = e^{-\frac{1}{2}(\mathbf{r}'-\mathbf{r})^t\mathbf{M}^t\mathbf{\Sigma}\mathbf{M}(\mathbf{r}'-\mathbf{r})t^2/\hbar^2}, \tag{15}$$

where $\mathbf{\Sigma} = \langle\mathbf{r}_2\mathbf{r}_2^t\rangle - \langle\mathbf{r}_2\rangle\langle\mathbf{r}_2^t\rangle$ is the covariance matrix of the location of the environment particle. Decoherence is destroyed when the exponent becomes of order unity, *i.e.*, on a timescale

$$\tau \equiv \left[(\mathbf{r}' - \mathbf{r})^t\mathbf{M}^t\mathbf{\Sigma}\mathbf{M}(\mathbf{r}' - \mathbf{r})\right]^{-1/2}\hbar. \tag{16}$$

Assuming a Coulomb potential $V = gq^2/|\mathbf{r}_2 - \mathbf{r}_1|$ gives $\mathbf{M} = (3\widehat{\mathbf{a}}\widehat{\mathbf{a}}^t - \mathbf{I})gq^2/a^3$ where $\mathbf{a} \equiv \mathbf{r}_2 - \mathbf{r}_1 = a\widehat{\mathbf{a}}$, $|\widehat{\mathbf{a}}| = 1$. For thermal states, we have the isotropic case $\mathbf{\Sigma} = (\Delta x)^2\mathbf{I}$, so equation (16) reduces to

$$\tau = \frac{\hbar a^3}{gq^2|\mathbf{r}' - \mathbf{r}|\Delta x}\left(1 + 3\cos^2\theta\right)^{-1/2}, \tag{17}$$

where $\cos\theta \equiv \widehat{\mathbf{a}} \cdot (\mathbf{r}' - \mathbf{r})/|\mathbf{r}' - \mathbf{r}|$. To be conservative, we take Δx to be as small as the uncertainty principle allows. With the thermal constraint $(\Delta p)^2/m \lesssim kT$ on the momentum uncertainty, this gives

---

[3] If the first ion starts at rest at $\mathbf{r}_1 = (0,0,0)$ and the second is incident with $\mathbf{r}_2 = (vt, b, 0)$, then a very weak scattering with deflection angle $\theta \ll 1$ will leave these trajectories roughly unchanged, the radial force $F = gq^2/|\mathbf{r}_1 - \mathbf{r}_2|^2$ merely causing a net transverse acceleration [57]

$$\Delta v_y = \int_{-\infty}^{\infty} \frac{\widehat{\mathbf{y}} \cdot \mathbf{F}}{m} dt = \int_{-\infty}^{\infty} \frac{gq^2 b\,dt}{[b^2 + (vt)^2]^{3/2}} = \frac{2gq^2}{mvb}. \tag{8}$$

The approximation breaks down as the deflection angle $\theta \approx \Delta v_y/v$ approaches unity. This occurs for $b \sim gq^2/mv^2$, giving $\sigma = \pi b^2$ as in equation (9).

$$\Delta x = \frac{\hbar}{2\Delta p} \sim \frac{\hbar}{\sqrt{mkT}}. \tag{18}$$

Substituting this into equation (17) and dividing by the number of ions N, we obtain the decoherence timescale

$$\tau \sim \frac{a^3\sqrt{mkT}}{Ngq^2|\mathbf{r}'-\mathbf{r}|}. \tag{19}$$

caused by a single environment ion a distance a away. Each such ion will produce its own suppression factor f, so we need to sum the exponent in equation (15) over all ions. Since the tidal force $\mathbf{M} \propto a^{-3}$ causes the exponent to drop as a^{-6}, this sum will generally be dominated by the very closest ion, which will typically be a distance $a \sim n^{-1/3}$ away. We are interested in decoherence for separations $|\mathbf{r}'-\mathbf{r}| = h$, the membrane thickness, which gives

$$\tau \sim \frac{\sqrt{mkT}}{Ngq_e^2nh} \sim 10^{-19}\,\mathrm{s}. \tag{20}$$

The relation between these different estimates is discussed in more detail in the Appendix.

C. Microtubules

Microtubules are a major component of the cytoskeleton, the "scaffolding" that helps cells maintain their shapes. They are hollow cylinders of diameter $D = 24\,nm$ made up of 13 filaments that are strung together out of proteins known as tubulin dimers. These dimers can make transitions between two states known as α and β, corresponding to different electric dipole moments along the axis of the tube. It has been argued that microtubules may have additional functions as well, serving as a means of energy and information transfer [20]. A model has been presented whereby the dipole-dipole interactions between nearby dimers can lead to long-range polarization and kink-like excitations that may travel down the microtubules at speeds exceeding 1 m/s [60].

Penrose has gone further and suggested that the dynamics of such excitations can make a microtubule act like a quantum computer, and that microtubules are the site of of human consciousness [2]. This idea has been further elaborated [21–24] employing methods from string theory, with the conclusion that quantum superpositions of coherent excitations can persist for as long as a second before being destroyed by decoherence. See also [61,62]. This was hailed as a success for the model, the interpretation being that the quantum gravity effect on microtubules was identified with the human though process on this same timescale.

This decoherence rate $\tau \sim 1\,\mathrm{s}$ was computed assuming that quantum gravity is the main decoherence source. Since this quantum gravity model is somewhat controversial [32] and its effect has been found to be more than 20 orders of magnitude weaker than other decoherence sources in some cases [46], it seems prudent to evaluate other decoherence sources for the microtubule case as well, to see whether they are in fact dominant. We will now do so.

Using coordinates where the x-axis is along the tube axis, the above-mentioned models all focus on the time-evolution of $p(x)$, the average x-component of the electric dipole moment of the tubulin dimers at each x. In terms of this polarization function $p(x)$, the net charge per unit length of tube is $-p'(x)$. The propagating kink-like excitations [60] are of the form

$$p(x) = \begin{cases} +p_0 & \text{for } x \ll x_0, \\ -p_0 & \text{for } x \gg x_0, \end{cases} \tag{21}$$

where the kink location x_0 propagates with constant speed and has a width of order a few tubulin dimers. The polarization strength p_0 is such that the total charge around the kink is $Q = -\int p'(x)dx = 2p_0 \sim 940q_e$, due to the presence of 18 Ca^{2+} ions on each of the 13 filaments contributing to p_0 [60].

Suppose that such a kink is in two different places in superposition, separated by some distance $|\mathbf{r}'-\mathbf{r}|$. How rapidly will the superposition be destroyed by decoherence? To be conservative, we will ignore collisions between polarized tubulin dimers and nearby water molecules, since it has been argued that these may be in some sense ordered and part of the quantum system [24] – although this argument is difficult to maintain for the water outside the microtubule, which permeates the entire cell volume. Let us instead apply equation (19), with $N = Q/q_e \sim 10^3$. The distance to the nearest ion will generally be less than $a = R + n^{-1/3} \sim 26\,\mathrm{nm}$, where the tubulin diameter $D = 24\,\mathrm{nm}$ dominates over the inter-ion separation $n^{-1/3} \sim 2\,\mathrm{nm}$ in the fluid surrounding the microtubule. Superpositions spanning many tubuline dimers ($|\mathbf{r}'-\mathbf{r}| \gg D$) therefore decohere on a timescale

$$\tau \sim \frac{D^2\sqrt{mkT}}{Ngq_e^2} \sim 10^{-13}\,\mathrm{s}. \tag{22}$$

due to the nearest ion alone. This is quite a conservative estimate, since the other $nD^3 \sim 10^3$ ions that are merely a small fraction further away will also contribute to the decoherence rate, but it is nonetheless 6-7 orders of magnitude shorter than the estimates of Mavromatos & Nanopoulos [25–27]. We will comment on screening effects below.

1. Decoherence summary

Our decoherence rates are summarized in Table 1. How accurate are they likely to be?

In the calculations above, we generally tried to be conservative, erring on the side of underestimating the decoherence rate. For instance, we neglected that N potassium ions also end up in superposition once the neuron firing is quenched, we neglected the contribution of other abundant ions such as Cl^- to η, and and we ignored collisions with water molecules in the microtubule case.

Since we were only interested in order-of-magnitude estimates, we made a number of crude approximations, *e.g.*, for the cross sections. We neglected screening effects because the decoherence rates were dominated by the particles closest to the system, *i.e.*, the very same particles that are responsible for screening the charge from more distant ones.

Table 1. Decoherence timescales.

Object	Environment	$\tau_{\rm dec}$
Neuron	Colliding ion	10^{-20}s
Neuron	Colliding H_2O	10^{-20}s
Neuron	Nearby ion	10^{-19}s
Microtubule	Distant ion	10^{-13}s

IV. DISCUSSION

A. The classical nature of brain processes

The calculations above enable us to address the question of whether cognitive processes in the brain constitute a classical or quantum system in the sense of Figure 1. If we take the characteristic dynamical timescale for such processes to be $\tau_{\rm dyn} \sim 10^{-2}\,{\rm s} - 10^{0}\,{\rm s}$ (the apparent timescale of *e.g.*, speech, thought and motor response), then a comparison of $\tau_{\rm dyn}$ with $\tau_{\rm dec}$ from Table 1 shows that processes associated with either conventional neuron firing or with polarization excitations in microtubules fall squarely in the classical category, by a margin exceeding ten orders of magnitude. Neuron firing itself is also highly classical, since it occurs on a timescale $\tau_{\rm dyn} \sim 10^{-3} - 10^{-4}\,{\rm s}$ [53]. Even a kink-like microtubule excitation is classical by many orders of magnitude, since it traverses a short tubule on a timescale $\tau_{\rm dyn} \sim 5\times 10^{-7}\,{\rm s}$ [60].

What about other mechanisms? It is worth noting that if (as is commonly believed) different neuron firing patterns correspond in some way to different conscious perceptions, then consciousness itself cannot be of a quantum nature even if there is a yet undiscovered physical process in the brain with a very long decoherence time. As mentioned above, suggestions for such candidates have involved, *e.g.*, superconductivity [12], superfluidity [13], electromagnetic fields [14], Bose condensation [15,16], superflourescence [17] and other mechanisms [18,19]. The reason is that as soon as such a quantum subsystem communicates with the constantly decohering neurons to create conscious experience, everything decoheres.

How extreme variations in the decoherence rates can we obtain by changing our model assumptions? Although the rates can be altered by a few of orders of magnitudes by pushing parameters such as the neuron dimensions, the myelination fraction or the microtubule kink charge to the limits of plausibility, it is clearly impossible to change the basic conclusion that $\tau_{\rm dec} \ll 10^{-3}\,{\rm s}$, *i.e.*, that we are dealing with a classical system in the sense of Figure 1. Even the tiniest neuron imaginable, with only a single ion ($N = 1$) traversing the cell wall during firing, would have $\tau_{\rm dec} \sim 10^{-14}\,{\rm s}$. Likewise, reducing the effective microtubule kink charge to a small fraction of q_e would not help.

How are we to understand the above-mentioned claims that brain subsystems can be sufficiently isolated to exhibit macroquantum behavior? It appears that the subtle distinction between dissipation and decoherence timescales has not always been appreciated.

B. Implications for the subject-object-environment decomposition

Let us now discuss the subsystem decomposition of Figure 2 in more detail in light of our results. As the figure indicates, the virtue of this decomposition into subject, object and environment is that the subsystem Hamiltonians $H_{\rm s}$, $H_{\rm o}$, $H_{\rm e}$ and the interaction Hamiltonians $H_{\rm so}$, $H_{\rm oe}$, $H_{\rm se}$ can cause qualitatively very different effects. Let us now briefly discuss each of them in turn.

Most of these processes are schematically illustrated in Figure 4 and Figure 5, where for purposes of illustration, we have shown the extremely simple case where both the subject and object have only a single degree of freedom that can take on only a few distinct values (3 for the subject, 2 for the object). For definiteness, we denote the three subject states $|\ddot{-}\rangle$, $|\ddot{\smile}\rangle$ and $|\ddot{\frown}\rangle$, and interpret them as the observer feeling neutral, happy and sad, respectively. We denote the two object states $|{\uparrow}\rangle$ and $|{\downarrow}\rangle$, and interpret them as the spin component ("up" or "down") in the z-direction of a spin-1/2 system, say a silver atom. The joint system consisting of subject and object therefore has only $2\times 3 = 6$ basis states: $|\ddot{-}{\uparrow}\rangle$, $|\ddot{-}{\downarrow}\rangle$, $|\ddot{\smile}{\uparrow}\rangle$, $|\ddot{\smile}{\downarrow}\rangle$, $|\ddot{\frown}{\uparrow}\rangle$, $|\ddot{\frown}{\downarrow}\rangle$. In Figures 4 and 5, we have therefore plotted ρ as a 6×6 matrix consisting of nine two-by-two blocks.

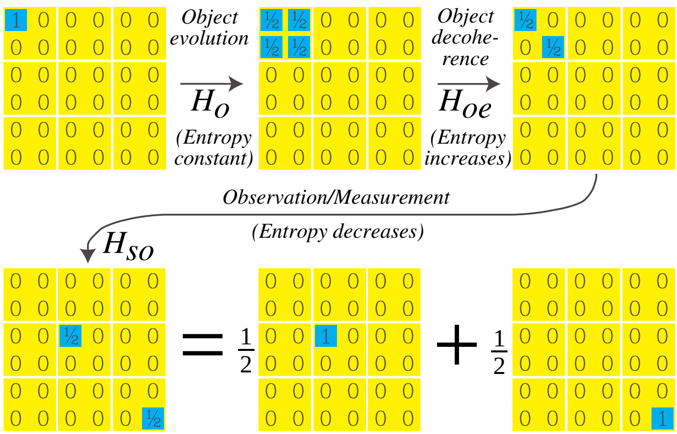

FIG. 4. Time evolution of the 6×6 density matrix for the basis states $|\ddot{-}\uparrow\rangle$, $|\ddot{-}\downarrow\rangle$, $|\ddot{\smile}\uparrow\rangle$, $|\ddot{\smile}\downarrow\rangle$, $|\ddot{\frown}\uparrow\rangle$, $|\ddot{\frown}\downarrow\rangle$ as the object evolves in isolation, then decoheres, then gets observed by the subject. The final result is a statistical mixture of the states $|\ddot{\smile}\uparrow\rangle$ and $|\ddot{\frown}\downarrow\rangle$, simple zero-entropy states like the one we started with.

1. Effect of H_o: constant entropy

If the object were to evolve during a time interval t without interacting with the subject or the environment ($H_{so} = H_{oe} = 0$), then according to equation (1) its reduced density matrix ρ_o would evolve into $U\rho_o U^\dagger$ with the same entropy, since the time-evolution operator $U \equiv e^{-iH_o t}$ is unitary.

Suppose the subject stays in the state $|\ddot{-}\rangle$ and the object starts out in the pure state $|\uparrow\rangle$. Let the object Hamiltonian H_o correspond to a magnetic field in the y-direction causing the spin to precess to the x-direction, *i.e.*, to the state $(|\uparrow\rangle+|\downarrow\rangle)/\sqrt{2}$. The object density matrix ρ_o then evolves into

$$\begin{aligned}\rho_o = U|\uparrow\rangle\langle\uparrow|U^\dagger &= \frac{1}{2}(|\uparrow\rangle+|\downarrow\rangle)(\langle\uparrow|+\langle\downarrow|)\\ &= \frac{1}{2}(|\uparrow\rangle\langle\uparrow|+|\uparrow\rangle\langle\downarrow|+|\downarrow\rangle\langle\uparrow|+|\downarrow\rangle\langle\downarrow|),\end{aligned} \quad (23)$$

corresponding to the four entries of 1/2 in the second matrix of Figure 4.

This is quite typical of pure quantum time evolution: a basis state eventually evolves into a superposition of basis states, and the quantum nature of this superposition is manifested by off-diagonal elements in ρ_o. Another familiar example of this is the familiar spreading out of the wave packet of a free particle.

2. Effect of H_{oe}: increasing entropy

This was the effect of H_o alone. In contrast, H_{oe} will generally cause decoherence and increase the entropy of the object. As discussed in detail in Section III and the Appendix, it entangles it with the environment, which suppresses the off-diagonal elements of the reduced density matrix of the object as illustrated in Figure 4. If H_{oe} couples to the z-component of the spin, this destroys the terms $|\uparrow\rangle\langle\downarrow|$ and $|\downarrow\rangle\langle\uparrow|$. Complete decoherence therefore converts the final state of equation (23) into

$$\rho_o = \frac{1}{2}(|\uparrow\rangle\langle\uparrow|+|\downarrow\rangle\langle\downarrow|), \quad (24)$$

corresponding to the two entries of 1/2 in the third matrix of Figure 4.

3. Effect of H_{so}: decreasing entropy

Whereas H_{oe} typically causes the apparent entropy of the object to increase, H_{so} typically causes it to decrease. Figure 4 illustrates the case of an ideal measurement, where the subject starts out in the state $|\ddot{-}\rangle$ and H_{so} is of such a form that gets perfectly correlated with the object. In the language of Section II, an ideal measurement is a type of communication where the mutual information I_{12} between the subject and object systems is increased to its maximum possible value. Suppose that the measurement is caused by H_{so} becoming large during a time interval so brief that we can neglect the effects of H_s and H_o. The joint subject+object density matrix ρ_{so} then evolves as $\rho_{so} \mapsto U\rho_{so}U^\dagger$, where $U \equiv \exp\left[-i\int H_{so}dt\right]$. If observing $|\uparrow\rangle$ makes the subject happy and $|\downarrow\rangle$ makes the subject sad, then we have $U|\ddot{-}\uparrow\rangle = |\ddot{\smile}\uparrow\rangle$ and $U|\ddot{-}\downarrow\rangle = |\ddot{\frown}\downarrow\rangle$. The state given by equation (24) would therefore evolve into

$$\rho_o = \frac{1}{2}U(|\ddot{-}\rangle\langle\ddot{-}|)\otimes(|\uparrow\rangle\langle\uparrow|+|\downarrow\rangle\langle\downarrow|)U^\dagger \quad (25)$$

$$= \frac{1}{2}(U|\ddot{-}\uparrow\rangle\langle\ddot{-}\uparrow|U^\dagger + U|\ddot{-}\downarrow\rangle\langle\ddot{-}\downarrow|U^\dagger \quad (26)$$

$$= \frac{1}{2}(|\ddot{\smile}\uparrow\rangle\langle\ddot{\smile}\uparrow| + |\ddot{\frown}\downarrow\rangle\langle\ddot{\frown}\downarrow|), \quad (27)$$

as illustrated in Figure 4. This final state contains a mixture of two subjects, corresponding to definite but opposite knowledge of the object state. According to both of them, the entropy of the object has decreased from one bit to zero bits.

In general, we see that the object decreases its entropy when it exchanges information with the subject and increases when it exchanges information with the environment.[4] Loosely speaking, the entropy of an object decreases while you look at it and increases while you don't[5].

---

[4] If n bits of information are exchanged with the environment, then equation (3) shows that the object entropy will increase by this same amount if the environment is in thermal equilibrium (with maximal entropy) throughout. If we were to know the state of the environment initially (by our definition of environment, we do not), then both the object and environment entropy will typically increase by $n/2$ bits.

[5] Here and throughout, we are assuming that the total system, which is by definition isolated, evolves according to the Schrödinger equation (1). Although modifications of the Schrödinger equation have been suggested by some authors, either in a mathematically explicit form as in [55,56] or verbally as a so-called reduction postulate, there is so far no experimental evidence suggesting that modifications are necessary. The original motivations for such modifications were

1. to be able to interpret the diagonal elements of the density matrix as probabilities and
2. to suppress off-diagonal elements of the density matrix.

The subsequent discovery by Everett [64] that the probability interpretation automatically appears to hold for almost all observers in the final superposition solved problem 1, and is discussed in more detail in, *e.g.*, [29,66–74]. The still more

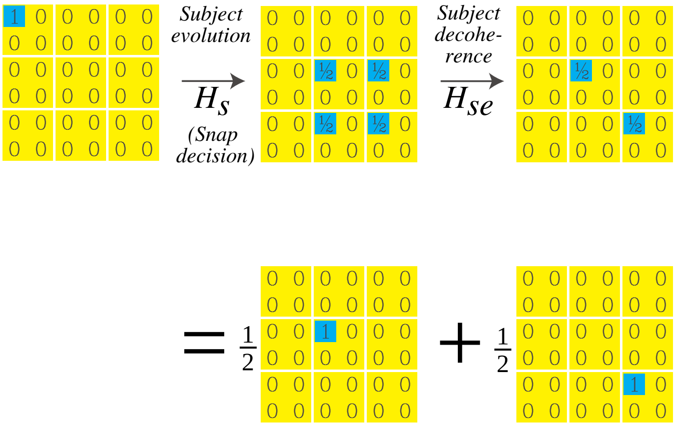

FIG. 5. Time evolution of the same 6×6 density matrix as in Figure 4 when the subject evolves in isolation, then decoheres. The object remains in the state $|\uparrow\rangle$ the whole time. The final result is a statistical mixture of the two states $|\smile\uparrow\rangle$ and $|\frown\uparrow\rangle$.

4. Effect if H_s: the thought process

So far, we have focused on the object and discussed effects of its internal dynamics (H_o) and its interactions with the environment (H_{oe}) and subject (H_{so}). Let us now turn to the subject and consider the role played by its internal dynamics (H_s) and interactions with the environment (H_{se}). In his seminal 1993 book, Stapp [3] presents an argument about brain dynamics that can be summarized as follows.

1. Since the brain contains $\sim 10^{11}$ synapses connected together by neurons in a highly nonlinear fashion, there must be a huge number of metastable reverberating patters of pulses into which the brain can evolve.

2. Neural network simulations have indicated that the metastable state into which a brain does in fact evolves depends sensitively on the initial conditions in small numbers of synapses.

3. The latter depends on the locations of a small number of calcium atoms, which might be expected to be in quantum superpositions.

4. Therefore, one would expect the brain to evolve into a quantum superposition of many such metastable configurations.

5. Moreover, the fatigue characteristics of the synaptic junctions will cause any given metastable state to become, after a short time, unstable: the subject will then be forced to search for a new metastable configuration, and will therefore continue to evolve into a superposition of increasingly disparate states.

If different states (perceptions) of the subject correspond to different metastable states of neuron firing patterns, a definite perception would eventually evolve into a superposition of several subjectively distinguishable perceptions.

We will follow Stapp in making this assumption about H_s. For illustrative purposes, let us assume that this can happen even at the level of a single thought or snap decision where the outcome feels unpredictable to us. Consider the following experiment: the subject starts out with a blank face and counts silently to three, then makes a snap decision on whether to smile or frown. The time-evolution operator $U \equiv \exp\left[-i\int H_s dt\right]$ will then have the property that $U|\ddot{-}\rangle = (|\smile\rangle + |\frown\rangle)/\sqrt{2}$, so the subject density matrix ρ_s will evolve into

$$\begin{aligned}\rho_s = U|\ddot{-}\rangle\langle\ddot{-}|U^\dagger &= \frac{1}{2}(|\smile\rangle + |\frown\rangle)(\langle\smile| + \langle\frown|)\\ &= \frac{1}{2}(|\smile\rangle\langle\smile| + |\smile\rangle\langle\frown| + |\frown\rangle\langle\smile| + |\frown\rangle\langle\frown|),\end{aligned} \tag{28}$$

corresponding to the four entries of 1/2 in the second matrix in Figure 5.

5. Effect of H_{se}: subject decoherence

Just as H_{oe} can decohere the object, H_{se} can decohere the subject. The difference is that whereas the object can be either a quantum system (with small H_{oe}) or a classical system (with large H_{oe}), a human subject *always* has a large interaction with the environment. As we showed in Section III, $\tau_{dec} \ll \tau_{dyn}$ for the subject, *i.e.*, the effect of H_{se} is faster than that of H_s by many orders of magnitude. This means that we should strictly speaking not think of macrosuperpositions such as equation (28) as first forming and then decohering as in Figure 5 — rather, subject decoherence is so fast that such superpositions decohere already during their process of formation. Therefore we are never even close to being able to perceive superpositions of different perceptions. Reducing object decoherence (from H_{oe}) during measurement would make no difference, since decoherence would take place in the brain long before the transmission of the appropriate sensory input through sensory nerves had been completed.

C. H_e and H_{soe}

The environment is of course the most complicated system, since it contains the vast majority of the degrees of

---

recent discovery of decoherence [11,36,37] solved problem 2, as well as explaining so-called superselection rules for the first time (why for instance the position basis has a special status) [44].

freedom in the total system. It is therefore very fortunate that we can so often ignore it, considering only those limited aspects of it that affect the subject and object.

For the most general H, there can also be an ugly irreducible residual term $H_{soe} \equiv H - H_s - H_o - H_e - H_{so} - H_{oe} - H_{se}$.

D. Implications for modeling cognitive processes

For the neural network community, the implication of our result is "business as usual", *i.e.*, there is no need to worry about the fact that current simulations do not incorporate effects of quantum coherence. The only remnant from quantum mechanics is the apparent randomness that we subjectively perceive every time the subject system evolves into a superposition as in equation (28), but this can be simply modeled by including a random number generator in the simulation. In other words, the recipe used to prescribe when a given neuron should fire and how synaptic coupling strengths should be updated may have to involve some classical randomness to correctly mimic the behavior of the brain.

1. Hyper-classicality

If a subject system is to be a good model of us, H_{so} and H_{se} need to meet certain criteria: decoherence and communication are necessary, but fluctuation and dissipation must be kept low enough that the subject does not lose its autonomy completely.

In our study of neural processes, we concluded that the subject is not a quantum system, since $\tau_{\rm dec} \ll \tau_{\rm dyn}$. However, since the dissipation time $\tau_{\rm diss}$ for neuron firing is of the same order as its dynamical timescale, we see that in the sense of Figure 1, the subject is not a simple classical system either. It is therefore somewhat misleading to think of it as simply some classical degrees of freedom evolving fairly undisturbed (only interacting enough to stay decohered and occasionally communicate with the outside world). Rather, the semi-autonomous degrees of freedom that constitute the subject are to be found at a higher level of complexity, perhaps as metastable global patters of neuron firing.

These degrees of freedom might be termed "hyper-classical": although there is nothing quantum-mechanical about their equations of motion (except that they can be stochastic), they may bear little resemblance with the underlying classical equations from which they were derived. Energy conservation and other familiar concepts from Hamiltonian dynamics will be irrelevant for these more abstract equations, since neurons are energy pumped and highly dissipative. Other examples of such hyper-classical systems include the time-evolution of the memory contents of a regular (highly dissipative) digital computer as well as the motion on the screen of objects in a computer game.

2. Nature of the subject system

In this paper, we have tacitly assumed that consciousness is synonymous with certain brain processes. This is what Lockwood terms the "identity theory" [66]. It dates back to Hobbes (~1660) and has been espoused by, *e.g.*, Russell, Feigl, Smart, Armstrong, Churchland and Lockwood himself. Let us briefly explore the more specific assumption that the subject degrees of freedom *are* our perceptions. In this picture, some of the subject degrees of freedom would have to constitute a "world model", with the interaction H_{so} such that the resulting communication keeps these degrees of freedom highly correlated with selected properties of the outside world (object + environment). Some such properties, *i.e.*,

- the intensity of the electromagnetic on the retina, averaged through three narrow-band filters (color vision) and one broad-band filter (black-and-white vision),
- the spectrum of air pressure fluctuations in the ears (sound),
- the chemical composition of gas in the nose (smell) and solutions in the mouth (taste),
- heat and pressure at a variety of skin locations,
- locations of body parts,

are tracked rather continuously, with the corresponding mutual information I_{12} between subject and surroundings remaining fairly constant. Persisting correlations with properties of the past state of the surroundings (memories) further contribute to the mutual information I_{12}. Much of I_{12} is due to correlations with quite subtle aspects of the surroundings, *e.g.*, the contents of books. The total mutual information I_{12} between a person and the external world is fairly low at birth, gradually grows through learning, and falls when we forget. In contrast, most innate objects have a very small mutual information with the rest of the world, books and diskettes being notable exceptions.

The extremely limited selection of properties that the subject correlates with has presumably been determined by evolutionary utility, since it is known to differ between species: birds perceive four primary colors but cats only one, bees perceive light polarization, *etc.* In this picture, we should therefore not consider these particular ("classical") aspects of our surroundings to be more fundamental than the vast majority that the subject system is uncorrelated with. Morover, our perception of *e.g.* space is as subjective as our perception of color, just as suggested by *e.g.* [50].

3. The binding problem

One of the motivations for models with quantum coherence in the brain was the so-called binding problem. In the words of James [75,76], "the only realities are the separate molecules, or at most cells. Their aggregation into a 'brain' is a fiction of popular speech". James' concern, shared by many after him, was that consciousness did not seem to be spatially localized to any one small part of the brain, yet subjectively feels like a coherent entity. Because of this, Stapp [3] and many others have appealed to quantum coherence, arguing that this could make consciousness a holistic effect involving the brain as a whole.

However, non-local degrees of freedom can be important even in classical physics, For instance, oscillations in a guitar string are local in Fourier space, not in real space, so in this case the "binding problem" can be solved by a simple change of variables. As Eddington remarked [77], when observing the ocean we perceive the moving waves as objects in their own right because they display a certain permanence, even though the water itself is only bobbing up and down. Similarly, thoughts are presumably highly non-local excitation patterns in the neural network of our brain, except of a non-linear and much more complex nature. In short, this author feels that there is no binding problem.

4. Outlook

In summary, our decoherence calculations have indicated that there is nothing fundamentally quantum-mechanical about cognitive processes in the brain, supporting the Hepp's conjecture [33]. Specifically, the *computations* in the brain appear to be of a classical rather than quantum nature, and the argument by Lisewski [78] that quantum corrections may be needed for accurate modeling of some details, *e.g.*, non-Markovian noise in neurons, does of course not change this conclusion. This means that although the current state-of-the-art in neural network hardware is clearly still very far from being able to model and understand cognitive processes as complex as those in the brain, there are no quantum mechanical reasons to doubt that this research is on the right track.

Acknowledgements: The author wishes to thank the organizers of the Spaatind-98 and Gausdal-99 winter schools, where much of this work was done, and Mark Alford, Philippe Blanchard, Carlton Caves, Angelica de Oliveira-Costa, Matthew Donald, Andrei Gruzinov, Piet Hut, Nick Mavromatos, Henry Stapp, Hans-Dieter Zeh and Woitek Zurek for stimulating discussions and helpful comments. Support for this work was provided by the Sloan Foundation and by NASA though grant NAG5-6034 and Hubble Fellowship HF-01084.01-96A from STScI, operated by AURA, Inc. under NASA contract NAS5-26555.

APPENDIX: DECOHERENCE FORMULAS

The quantitative effect of decoherence from both short range interactions (scattering) and long-range interactions was first derived in a seminal paper by Joos & Zeh [45]. Since our application involved scattering between particles of comparable mass, we used a generalized version of these results that included the effect of recoil [46]. In this Appendix, we derive a slightly generalized formula for long-range interactions, and briefly comment on the relation between these short-range and long-range limiting cases.

1. Decoherence due to tidal forces

Even if the dissipation and fluctuation caused by $H_{\rm int}$ is dynamically unimportant, H_1 and H_2 can be neglected in equation (2) when calculating the decoherence effect in the many cases where the interaction Hamiltonian decoheres the object on a timescale far below the dynamical time. In this approximation, we consider two particles with an interaction $H = H_{\rm int} = V(\mathbf{r}_2 - \mathbf{r}_1)$ for some potential V. According to equation (1), the two-particle density matrix ρ therefore evolves as

$$\begin{aligned}&\rho(\mathbf{r}_1, \mathbf{r}_1', \mathbf{r}_2, \mathbf{r}_2', t_0 + t)\\ &= \rho(\mathbf{r}_1, \mathbf{r}_1', \mathbf{r}_2, \mathbf{r}_2', t) e^{-i[V(\mathbf{r}_2-\mathbf{r}_1)-V(\mathbf{r}_2'-\mathbf{r}_1')]/\hbar}. \end{aligned} \tag{A1}$$

Following [45], we assume that the two particles are fairly localized near their initial average positions

$$\mathbf{r}_i^0 \equiv \langle \mathbf{r}_i \rangle_0 = \mathrm{tr}\,[\mathbf{r}_i \rho_i(t_0)], \tag{A2}$$

$i = 1, 2$, and approximate the potential by its second order Taylor expansion

$$\begin{aligned}V(\mathbf{r}_2 - \mathbf{r}_1) \approx{}& V(\mathbf{a}) - \mathbf{F} \cdot (\mathbf{x}_2 - \mathbf{x}_1)\\ &+ \frac{1}{2}(\mathbf{x}_2 - \mathbf{x}_1)^t \mathbf{M} (\mathbf{x}_2 - \mathbf{x}_1). \end{aligned} \tag{A3}$$

Here $\mathbf{F} \equiv = -\nabla V(\mathbf{a})$ is the average force, $\mathbf{M}$ is the Hessian matrix $\mathbf{M}_{ij} \equiv \partial_i \partial_j V(\mathbf{a})$ and $\mathbf{a} \equiv \mathbf{r}_2^0 - \mathbf{r}_1^0$. We have introduced relative coordinates $\mathbf{x}_i \equiv \mathbf{r}_i - \mathbf{r}_i^0$. Assuming that the two particles are independent initially as in [45], *i.e.*, that $\rho(t_0)$ takes the separable form $\rho(\mathbf{x}_1, \mathbf{x}_1', \mathbf{x}_2, \mathbf{x}_2', t_0) = \rho_1(\mathbf{x}_1, \mathbf{x}_1', t_0)\rho_2(\mathbf{x}_2, \mathbf{x}_2', t_0)$, this gives

$$\begin{aligned}&\rho_1(\mathbf{x}_1, \mathbf{x}_1', t_0 + t) = \mathrm{tr}\,_2 \rho(t_0 + t) = \\ &\int \rho(\mathbf{x}_1, \mathbf{x}_1', \mathbf{x}, \mathbf{x}, t_0 + t) d^3x = \rho_1(\mathbf{x}_1, \mathbf{x}_1', t_0) f(\mathbf{x}_1, \mathbf{x}_1', t), \end{aligned} \tag{A4}$$

where

$$f(\mathbf{x}_1,\mathbf{x}_1',t)\approx$$
$$e^{i\phi(\mathbf{x}_1,\mathbf{x}_1',t)}\int\rho_2(\mathbf{x}_2,\mathbf{x}_2',t_0)e^{-it(\mathbf{x}_1'-\mathbf{x}_1)^t\mathbf{M}\mathbf{x}_2/\hbar}d^3x_2=$$
$$e^{i\phi(\mathbf{x}_1,\mathbf{x}_1',t)}\widehat{p}_2[\mathbf{M}(\mathbf{x}_1'-\mathbf{x}_1)t/\hbar]. \tag{A5}$$

Here the phase factor

$$e^{i\phi(\mathbf{x},\mathbf{x}',t)}\equiv e^{\frac{i}{\hbar}\left[\mathbf{F}\cdot(\mathbf{x}'-\mathbf{x})+\frac{1}{2}\mathbf{x}'^t\mathbf{M}\mathbf{x}'-\frac{1}{2}\mathbf{x}^t\mathbf{M}\mathbf{x}\right]} \tag{A6}$$

is of no importance for decoherence, since it does not suppress the magnitude $|\rho_1(\mathbf{x}_1,\mathbf{x}_1',t)|$ of the off-diagonal elements – it merely causes momentum transfer related to fluctuation and dissipation. It is the other term that causes decoherence. $\widehat{p}_2$ is the Fourier transform of $p_2(\mathbf{x})\equiv\rho_2(\mathbf{x},\mathbf{x},t_0)$, the probability distribution for the location of the environment particle.

2. Properties of the effect

Let us briefly discuss some qualitative features of equation (A5). Since $\widehat{p}_2(\mathbf{0})=\int p_2(\mathbf{x}_2)d^3x_2=\mathrm{tr}\,\rho_2=1$, $\rho_1(\mathbf{x},\mathbf{x}')$ remains unchanged on the diagonal $\mathbf{x}=\mathbf{x}'$. This is because $H_{\rm int}$ is not changing the position of our our object particle, merely its momentum. Since the mean position $\langle\mathbf{x}_2\rangle=\int p_2\mathbf{x}_2d^3x_2=\mathrm{tr}\,[\mathbf{x}_2\rho_2]=0$ vanishes (using equation (A2)), we have $\nabla\widehat{p}_2(\mathbf{0})=\mathbf{0}$. In fact, $|f|$ takes a maximum on the diagonal, and the Riemann-Lebesgue Lemma shows that $|f|=|\widehat{p}_2|\leq 1$ whenever $\mathbf{x}\neq\mathbf{x}'$, with equality only for the unphysical case where p_2 is a delta function, *i.e.*, where the location of the environment particle is perfectly known. $\partial_i\partial_j|f(\mathbf{0})|=-\mathbf{M}^(\mathbf{x}_2\mathbf{x}_2^t\rangle\mathbf{M}t^2/2\hbar^2$, so so the larger $\langle\mathbf{x}_2\mathbf{x}_2^t\rangle$ is (*i.e.*, the more spread out the environment particle is), the closer to the diagonal decoherence will suppress our density matrix.

Since $\mathbf{M}$ is the shear matrix of the force field $-\nabla V$, we see that it is tidal forces that are causing the decoherence — the average force $\mathbf{F}$ simply contributes to the phase factor $e^{i\phi}$. Specifically, the rate at which our object degrees of freedom $\mathbf{r}_1$ decohere grows with the tidal force that it exerts on the environment: if the environment particle is spread out with $\langle\mathbf{x}_2\mathbf{x}_2^t\rangle$ large, experiencing a wide range of forces from the object, object decoherence is rapid. In the opposite situation, where the object is spread out and the environment is not, the object will experience strong classical tidal forces but no decoherence.

3. Relation between long-range and short-range decoherence

Above we derived the effect of decoherence from long-range tidal forces. Another interesting case that has been solved analytically [45] is that of short-range interactions that can be modeled as scattering events. If the scattering takes place during short enough a time interval that we can neglect the internal dynamics of the object, then its reduced density matrix changes as [46]

$$\rho_1(\mathbf{r},\mathbf{r}')\mapsto\rho_1(\mathbf{r},\mathbf{r}')\widehat{p}\left(\frac{\mathbf{r}'-\mathbf{r}}{\hbar}\right), \tag{A7}$$

where $p(\mathbf{q})$ is the probability distribution for the momentum transfer $\mathbf{q}$ in the collision. This equation generalizes the scattering result of [45] by including the effect of recoil. The larger the uncertainty in momentum transfer, the stronger the decoherence effect becomes, since widening p narrows its Fourier transform $\widehat{p}$. Changing the mean momentum transfer $\langle\mathbf{q}\rangle$ does not affect the decoherence, merely contributes a phase factor just as $\mathbf{F}$ did above. Typically, the last factor in equation (A7) destroys coherence down to scales of order the de Broigle wavelength of the scatterer, with directional modulations from the angular dependence of the scattering cross section. Generalization to a steady flux of scattering particles [46] gives equation (6).

Equation (A7) has striking similarities with the tidal force result of equation (A5): in both cases, the density matrix gets multiplied by the Fourier transform of a probability distribution. If fact, up to uninteresting phase factors, we can rewrite our equation (A5) in exactly the form of equation (A7) by redefining p to be the probability distribution for momentum transfer $\mathbf{q}=\mathbf{M}(\mathbf{x}_2-\mathbf{x}_1)t$ due to tidal forces for a fixed $\mathbf{x}_1$, *i.e.*,

$$p(\mathbf{q})\equiv p_2(\mathbf{x}_2)\frac{d^3x_2}{d^3q}=\frac{p_2(\mathbf{x}_1+\mathbf{M}^{-1}\mathbf{q}/t)}{t^3\det\mathbf{M}}. \tag{A8}$$

Fourier transforming this expression and substituting the result into equation (A7), we recover equation (A5) up to a phase factor.

Perhaps the simplest way to understand all these results is in terms of Wigner functions [79]. If $W(\mathbf{x}_1,\mathbf{p}_1)$ is the Wigner phase space distribution for the object particle, then any of the momentum-transferring interactions that we have considered will take the form

$$W(\mathbf{x}_1,\mathbf{p}_1)\mapsto\int W(\mathbf{x}_1,\mathbf{p}_1-\mathbf{q})p(\mathbf{q},\mathbf{x}_1)d^3q \tag{A9}$$

for some probability distribution p that may or may not depend on $\mathbf{x}_1$. Since the density matrix

$$\rho_1(\mathbf{x}_1,\mathbf{x}_1')=\int W\left(\frac{\mathbf{x}_1+\mathbf{x}_1'}{2},\mathbf{p}\right)e^{-i(\mathbf{x}-\mathbf{x}')\cdot\mathbf{p}}d^3p \tag{A10}$$

is just the Wigner function Fourier transformed in the momentum direction (and rotated by 45°), the convolution with p in equation (A9) reduces to a simple multiplication with $\widehat{p}$ in equation (A7).

[1] R. Penrose, *The Emperor's New Mind* (Oxford, Oxford Univ. Press, 1989).
[2] R. Penrose, in *The Large, the Small and the Human Mind*, ed. M. Longair (Cambridge, Cambridge Univ. Press, 1997).
[3] H. P. Stapp, *Mind, Matter and Quantum Mechanics* (Berlin, Springer, 1993).
[4] D. J. Amit, *Modeling Brain Functions* (Cambridge, Cambridge Univ. Press, 1989).
[5] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (Singapore, World Scientific, 1993).
[6] R. L. Harvey, *Neural Network Principles* (Englewood Cliffs, Prentice Hall, 1994).
[7] F. H. Eeckman and J. M. Bower, *Computation and Neural Systems* (Boston, Kluwer, 1993).
[8] D. R. McMillen, G. M. T D'Eleuterio, and J. R. P Halperin, Phys. Rev. E **59**, 6 (1999).
[9] E. P. Wigner, in *The Scientist Speculates: an Anthology of Partly-Baked Ideas, p284-302*, ed. I. J. Good (London, Heinemann, 1962).
[10] J. Mehra and A. S. Wightman, *The Collected Works of E. P. Wigner, Vol. VI, p271* (Berlin, Springer, 1995).
[11] H. D. Zeh, Found. Phys. **1**, 69 (1970).
[12] E. H. Walker, Mathematical Biosciences **7**, 131 (1970).
[13] L. H. Domash, in *Scientific Research on TM*, ed. D. W. Orme-Johnson and J. T. Farrow (Weggis, Switzerland, Maharishi Univ. Press, 1977).
[14] H. P. Stapp, Phys. Rev. D **28**, 1386 (1983).
[15] I. N. Marshall, New Ideas in Psychology **7**, 73 (1989).
[16] D. Zohar, *The Quantum Self* (New York, William Morrow, 1990).
[17] H. Rosu, Metaphysical Review **3**, 1, gr-qc/9409007 (1997).
[18] L. M. Ricciardi and H. Umezawa, Kibernetik **4**, 44 (1967).
[19] A. Vitiello, Int. J. Mod. Phys. **B9**, 973-89 (1996).
[20] S. R. Hameroff and R. C. Watt, Journal of Theoretical Biology **98**, 549 (1982).
S. R. Hameroff, *Ultimate Computing: Biomolecular Consciousness and Nanotechnology* (Amsterdam, North-Holland, 1987).
[21] D. V. Nanopoulos 1995, hep-ph/9505374
[22] N. Mavromatos and D. V. Nanopoulos 1995, hep-ph/9505401
[23] N. Mavromatos and D. V. Nanopoulos 1995, quant-ph/9510003
[24] N. Mavromatos and D. V. Nanopoulos 1995, quant-ph/9512021
[25] N. Mavromatos and D. V. Nanopoulos, Int. J. Mod. Phys B **12**, 517, quant-ph/9708003 (1998).
[26] N. Mavromatos and D. V. Nanopoulos 1998, quant-ph/9802063
[27] N. Mavromatos 1999, J. Bioelectrochemistry & Bioenergetics;48;273
[28] H. P. Stapp, Found. Phys. **21**, 1451 (1991).
[29] H. D. Zeh, quant-ph/9908084, *Epistemological Letters of the Ferdinand-Gonseth Association* **63:0** (Biel, Switzerland, 1981)
[30] W. H. Zurek, Phys. Today **44 (10)**, 36 (1991).
[31] A. Scott, J. Consciousness Studies **6**, 484 (1996).
[32] S. Hawking, in *The Large, the Small and the Human Mind*, ed. M. Longair (Cambridge, Cambridge Univ. Press, 1997).
[33] K. Hepp, in *Quantum Future*, ed. P. Blanchard and A. Jadczyk (Berlin, Springer, 1999).
[34] J. von Neumann, *Matematische Grundlagen der Quanten-Mechanik* (Springer, Berlin, 1932).
[35] H. D. Zeh, *The Arrow of Time, 3rd ed.* (Springer, Berlin, 1999).
[36] W. H. Zurek, Phys. Rev. D **24**, 1516 (1981).
[37] W. H. Zurek, Phys. Rev. D **26**, 1862 (1982).
[38] W. H. Zurek, reprint LAUR 84-2750, in *Non-Equilibrium Statistical Physics*, ed. G. Moore and M. O. Sculy (New York, Plenum, 1984).
[39] E. Peres, Am. J. Phys. **54**, 688 (1986).
[40] P. Pearle, Phys. Rev. A **39**, 2277 (1989).
[41] M. R. Gallis and G. N. Fleming, Phys. Rev. A **42**, 38 (1989).
[42] W. H. Unruh and W. H. Zurek, Phys. Rev. D **40**, 1071 (1989).
[43] R. Omnès, Phys. Rev. A **56**, 3383 (1997).
[44] D. Giulini, E. Joos, C. Kiefer, J. Kupsch, I. O. Stamatescu, and H. D. Zeh, *Decoherence and the Appearance of a Classical World in Quantum Theory* (Springer, Berlin, 1996).
[45] E. Joos and H. D. Zeh, Z. Phys. B **59**, 223 (1985).
[46] M. Tegmark, Found. Phys. Lett. **6**, 571 (1993).
[47] R. P. Feynman, *Statistical Mechanics* (Reading, Benjamin, 1972).
[48] B. Katz, *Nerve, Muscle, and Synapse* (New York, McGraw-Hill, 1966).
[49] J. P. Schadé and D. H. Ford, *Basic Neurology, 2nd ed.* (Amsterdam, Elsevier, 1973).
[50] A. G. Cairns-Smith, *Evolving the Mind* (Cambridge, Cambridge Univ. Press, 1996).
[51] P. Morell and W. T. Norton, Sci. Am. **242**, 74 (1980).
[52] A. Hirano and J. A. Llena, in *The Axon*, ed. S. G. Waxman, J. D. Kocsis, and P. K. Stys (New York, Oxford Univ. Press, 1995).
[53] J. M. Ritchie, in *The Axon*, ed. S. G. Waxman, J. D. Kocsis, and P. K. Stys (New York, Oxford Univ. Press, 1995).
[54] J. Ellis, S. Mohanty, and Nanopoulos D V, Phys. Lett. B **221**, 113 (1989).
[55] P. Pearle, Phys. Rev. D **13**, 857 (1976).
[56] G. C. Ghirardi, A. Rimini, and T. Weber, Phys. Rev. D **34**, 470 (1986).
[57] J. D. Jackson, *Classical Electrodynamics* (New York, Wiley, 1975).
[58] W. H. Zurek, S. Habib, and J. P. Paz, Phys. Rev. Lett. **70**, 1187 (1993).
[59] M. Tegmark and H. S. Shapiro, Phys. Rev. E **50**, 2538 (1994).
[60] M. V. Satarić, J. A. Tuszyński, and R. B. Žakula, Phys. Rev. E **48**, 589 (1993).
[61] H. C. Rosu, Phys. Rev. E **55**, 2038 (1997).
[62] H. C. Rosu, Nuovo Cimento D **20**, 369 (1998).
[63] H. S. Stapp 1999,
Attention, Intention, and Mind in Quantum Physics and *Quantum Ontology and Mind-Matter Synthesis*, available

at *www-physics.lbl.gov/ stapp/stappfiles.html.*
[64] H. Everett III, Rev. Mod. Phys. **29**, 454 (1957).
H. Everett III, *The Many-Worlds Interpretation of Quantum Mechanics, B. S. DeWitt and N. Graham* (Princeton, Princeton Univ. Press, 1986).
[65] J. A. Wheeler, Rev. Mod. Phys. **29**, 463;1957 (1957).
L. M. Cooper and D. van Vechten, Am. J. Phys **37**, 1212 (1969).
B. S. DeWitt, Phys. Today **23**, 30 (1971).
[66] M. Lockwood, *Mind, Brain and the Quantum* (Cambridge, Blackwell, 1989).
[67] D. Deutsch *The Fabric of Reality* (Allen Lane, New York, 1997).
[68] D. N. Page A 1995, gr-qc/9507025
[69] M. J. Donald 1997, quant-ph/9703008
[70] M. J. Donald 1999, quant-ph/9904001
[71] L. Vaidman 1996, quant-ph/9609006, *Int. Stud. Phil. Sci.*, in press
[72] T. Sakaguchi 1997, quant-ph/9704039
[73] M. Tegmark, quant-ph/9709032, Fortschr. Phys. **46**, 855 (1997).
[74] M. Tegmark, gr-qc/9704009, Annals of Physics **270**, 1 (1998).
[75] W. James, *The Principles of Psychology* (New York, Holt, 1890).
[76] W. James 1904, in *The Writings of William James, pp169-183*, ed. J. J. McDermott (Chicago, Univ. Chicago Press, 1977).
[77] A. Eddington, *Space, Time & Gravitation* (Cambridge, Cambridge Univ. Press, 1920).
[78] A. M. Lisewski 1999, quant-ph/9907052
[79] E. P. Wigner, Phys. Rev. **40**, 749 (1932).
M. Hillery, R. H. O'Connell, M. O. Scully & Wigner E P, Phys. Rep. **106**, 121 (1984).
Y. S. Kim and M. E. Noz, *Phase Space Picture of Quantum Mechanics: Group Theoretical Approach* (Singapore, World Scientific, 1991).