Towards a Theory of Mind

ABRAHAM BOYARSKY

Department of Mathematics and Statistics, Concordia University, Montreal, Canada H4B 1R6

The language of nonlinear dynamical systems and ergodic theory is used to present a theoretical framework for the study of mind. The basic space X consists of the collection of all brain images (clusters of activated neurons) that are relevant to consciousness. The dynamics of the brain is modelled by means of a discrete time transformation T which takes a cluster of activated brain cells into another cluster of activated brain cells. The space X is partitioned into subcollections of brain images, namely those generated by the five senses and by other processes that produce brain images relevant to consciousness. It is argued that T is a Markov transformation with respect to this partition of X. This leads to the existence of an object μ referred to as an SRB measure which possesses properties that make it a candidate for mind: μ is 'aware' of the brain images in its support; μ is time-invariant and acts as an attractor into which all orbits of (conscious) brain images settle. Furthermore, the dynamical systems model for mind allows the estimation of brain information rates and provides a framework in which a number of mind related issues can be discussed.

Keywords: Consciousness, Theory of mind, Dynamical system, Invariant measure

1. INTRODUCTION

This note is written from the perspective that it is better to have some theory of mind – even one fraught with simplifications – than to have no theory at all, which is the current state. There are scientific efforts underway – most notably from the perspective of quantum mechanics [4,5,9,10,16] – to explain consciousness, but there is no general science of mind. There is no theoretical starting point, no basic model from which more sophisticated ones can develop. Einstein's Relativity Theory emerged from the rudimentary attempts of Copernicus, Galileo and Newton to describe the macroworld. These early scientists saw physical bodies and used telescopes to formulate a theory of celestial mechanics. The physical bodies of the mind are brain images and using instruments such as PET scans, mind scientists are now capable of formulating a dynamical theory of mind.

This note does not deal with explanations of mind, but rather attempts to establish a theoretical framework in which to study it. To do this we use the language of nonlinear dynamics and ergodic theory. We begin with a physical brain B that consists of a large number of neurons and define X to be the space of all images realized in B. By this we mean all possible collections of activated neurons associated

with conscious processes. A point $x \in X$ is a cluster of neurons, which does not have to consist of contiguous activated neurons, but can appear as a complicated, disconnected collection of such neurons. X is, in fact, a subset of a very large but finite dimensional space, R^n [3].

A living brain is a dynamical system and, as such, there must exist a process or processes by which brain images are transformed into brain images. These processes may operate in continuous or discrete time and may be described by differential or partial differential equations or by difference or integral equations or even by combinations of these. For simplicity and to facilitate the presentation of our model (and without loss of generality), we assume that the dynamics of brain images can be captured by a transformation T, which acts in discrete time. Although the brain is organized at many levels, we are concerned only with those levels which effect consciousness and, for those levels, we assume that T describes brain dynamics. We also assume that the brain structures and processes are constant over the time scale of interest. Hence T can be considered to be constant. If we allow for the evolution of brain structures and processes, B and T become time-dependent and the model of this note applies in a time-dependent manner.

Section 2 provides a brief overview of the relevant notions and notations from dynamical systems and ergodic theory. In Section 3 we argue that the space of brain images X can be partitioned into a finite number of subsets, namely those which originate from the five senses and from processes such as thought and speech that are relevant to consciousness. We let $\mathcal{P}$ denote the collection of these subsets of X and then argue that T is a Markov transformation, that is, that it maps any element P of $\mathcal{P}$ onto a collection of elements of $\mathcal{P}$. Under general conditions the dynamical system (X, T) possesses an object, μ, called an SRB measure, which describes the long term behaviour of brain image orbits and possesses properties that justify calling it mind. One of these properties is that μ acts as an attractor to most brain images whose orbits settle into it after a transitory trajectory of images. Another property is

that μ manifests 'awareness' of the brain and reveals the underlying brain dynamical system (X, T). In Section 4 we use the structure of the Markov model to estimate information rates for the conscious part of the brain. In Section 5 we discuss how the dynamical system model can be used to create a universal model in which brains are connected. In Section 6 we show that the dynamical system model has a connection to quantum theory in view of the fact that wave functions, when squared, give rise to probability density functions that are the densities of SRB measures for the spatial processes of quantum mechanics.

2. DYNAMICAL SYSTEM NOTATION

Let X be a space (i.e., a collection of points) and $T: X \rightarrow X$ a transformation. When we want to measure subsets of X, we first classify the subsets to be considered. These are usually the Borel subsets, a Borel subset being the smallest collection of subsets of X containing the open subsets (unions of balls) and is closed under taking complements and unions of countable many members of this collection. These subsets are said to be measurable. A function which associates to each Borel set a positive number in such a way that for each countable collection of disjoint sets A_i, one has $\mu(\bigcup A_i) = \sum \mu(A_i)$ is called a measure. It is a probability measure if $\mu(X) = 1$. We consider only probability measures. By the support of a measure we mean the smallest (closed) Borel subset S such that $\mu(S) = 1$. We say μ is supported on S. Furthermore, S is an invariant set which means that $T(S) \subseteq S$.

In dynamical systems theory we are interested in measures which are invariant under the transformation T. Such a measure is defined by the property that $\mu(T^{-1}A) = \mu(A)$ for all measurable subsets A, where $T^{-1}(A) = \{x: Tx \in X\}$, the set of points that are transformed into A after one application of T. We say μ is an ergodic measure if, for each Borel set A, for which $T^{-1}A = A$, we have $\mu(A) = 0$ or $\mu(A) = 1$. Ergodicity is an indecomposability property

which holds in many applications. One of the main consequences of ergodicity is the existence of dense orbits, that is, orbits that wander throughout the entire space X. If μ is ergodic, the Birkhoff Ergodic Theorem [12], implies: for each measurable subset A,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_A(T^i x) = \mu(A) \qquad (1)$$

for almost all x, where $\chi_A(x)$ is the indicator function of A, that is, $\chi_A(x)$ is 1 if $x \in A$ and 0 otherwise. It is possible that μ is a measure supported on a periodic orbit and then the statement of the Birkhoff Ergodic Theorem is not interesting as it applies only to those x's in the periodic orbit. We are, therefore, interested in invariant measures which are supported on large sets. One such class of measures **are** the SRB measures [14]. These are physically relevant measures in that they are the limits of experimentally found measures. More precisely, we say that a sequence of measures $\{\mu_n\}$ converges in the weak sense to the measure μ if for all continuous functions g, we have $\int g \, d\mu_n \to \int g \, d\mu$. The measure μ is called **an** SRB measure or, a physical measure, if

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x} = \mu \quad \text{weakly}, \qquad (2)$$

where δ_x is the point measure at x, often referred to as the Dirac measure. It has value 1 at x and is equal to 0 elsewhere. This point measure can be thought of as a function which identifies the point x. We can visualize it as a vertical line of height 1 located at the point x, much as a post in the ground identifies the spot into which it is rooted. Note that $\{\delta_{T^i x}\}$ traces the orbit of Dirac measures associated with the actual physical orbit $\{T^i x\}$, x being any initial point. (Note that Eq. (2) holds for more general weighting sequences than $\{1/n\}$.) Now, T transforms x to Tx. This induces a transformation T_* which transforms the Dirac measure δ_x to the Dirac measure δ_{Tx}, as depicted in Fig. 1.

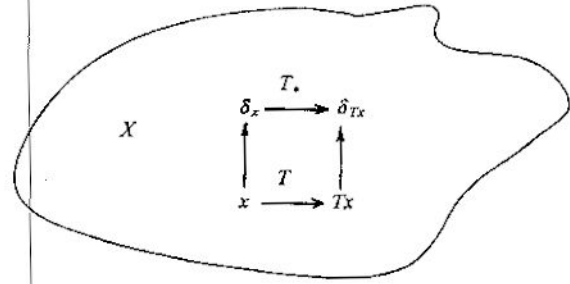The final notion we will need is that of a Markov transformation. Roughly, T is a Markov

FIGURE 1   Transformation of Dirac measures under T.

transformation on X if there exists a finite partition P of X, that is, a collection of subsets $\{P_1, P_2, \ldots, P_n\}$ of X such that T transforms each P_i onto some union of P_j's. For Markov transformations there exists a general ergodic theorem which guarantees the existence of an SRB measure [14]. Other sufficient conditions for the existence of SRB measures are expansion [7] and folding [3].

Among the properties of SRB measures are the following:

(i) μ acts on subsets of X, that is on space rather than in time and, as such, is always 'available' as a source of knowledge about the dynamical system. Furthermore, μ can be used to reveal T.

(ii) μ acts as a global attractor as evidenced by (2), i.e., for any starting point (in a large set of starting points), the time-averaged orbit converges (weakly) to μ.

As a simple example of the foregoing, consider the one dimensional transformation $T : X \to X$, where $X = [0, 1]$ and T is the triangle transformation defined by

$$T_x = \begin{cases} 2x, & \text{if } 0 \le x < \frac{1}{2}, \\ 2 - 2x, & \text{if } \frac{1}{2} < x \le 1, \end{cases}$$

and shown in Fig. 2.

Let the partition P of X consist of $P_1 = [0, \frac{1}{2})$ and $P_2 = [\frac{1}{2}, 1]$. Then T is Markov for this partition since $T(P_1) = P_1 \cup P_2$ and $T(P_2) = P_1 \cup P_2$. Obviously, T is onto as it maps each element of the partition
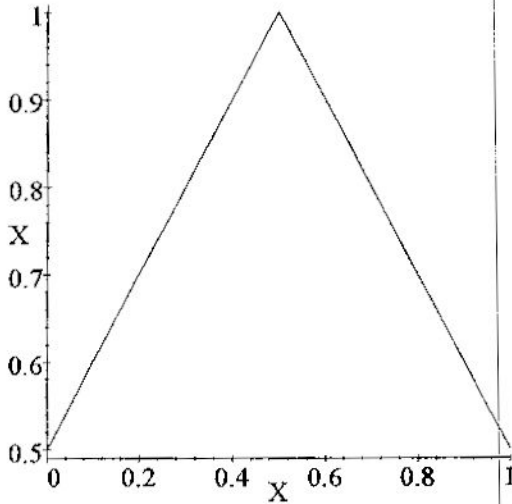
FIGURE 2   One dimensional triangle transformation T.

onto X. It is easy to show that the SRB measure for this system is $\mu = \lambda$, the Lebesgue measure which measures intervals (and more complicated sets) simply by their physical length. Hence,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x} = \lambda \qquad (3)$$

weakly for almost all x.

Let us now look more closely at Eq. (3) as this will help us in the sequel when we apply the foregoing ideas to brain dynamics. 'For almost all x' means 'for all x in a set of Lebesgue measure 1'. In this example the 'almost all' statement allows for the existence of many starting points x for which (3) fails – indeed this is the case if x is a periodic point of T. Then the orbit $\{\delta_{Tx}\}$ cycles through the finite number of points in the cycle and cannot converge to λ in the sense of (3). All the periodic points have Lebesgue measure 0, and hence if, for whatever reason, the orbit is nudged out of its periodic cycle, it is likely to become a point in the set of Lebesgue measure 1 and then the new orbit will quickly reflect, via Eq. (3), the SRB measure λ. Since T is an expanding transformation, it does not take much to perturb the periodic orbit from its cycle. In fact, this system is 'sensitive to initial conditions' [6]; the slightest

nudge is sufficient to expose the system's inherent chaotic behaviour and the new orbit will settle quickly into the system's SRB measure λ.

To summarize: for most starting points x, the SRB measure (which is known to exist) will come into actual being via the averaging process in (3). Due to the instability of periodic orbits (T is expanding), any cycle is eventually derailed to a point whose orbit is chaotic and the associated orbit of Dirac measures quickly settles into the SRB measure. The starting points that lead to chaotic orbits have Lebesgue measure 1 and hence are ubiquitously available to settle the dynamics into the SRB measure λ.

Once the SRB measure is known to exist, it is the flow of time and the time-averaging mechanism that displays the SRB measure. This averaging mechanism is nothing other than a kind of check-up of all the components of the brain system, that they are functioning properly and available for service.

We will use these rudimentay notions in the sequel to study brain dynamics. To prepare for the application in a higher dimensional setting, note that λ possesses the following properties.

Awareness   Let x^* be any point in X which is not a periodic point and let A^* be a very small set containing x^*. Since T is expanding, the orbit $\{T^i x^*\}_{i=0}^{n-1}$ moves quickly out of A^*, say in $m < n$ steps. Thus,

$$\lambda_{/A^*} \approx (m/n)\delta_{x^*} = c\delta_{x^*}, \qquad (4)$$

which is the (scaled) Dirac measure at x^*. Knowing this measure is the same as knowing x^*. Hence λ is 'aware' of x^*.

Inversion   Since X and the partition P are known, knowledge of λ determines (almost uniquely) the transformation T. This follows from a simple application of the matrix form of the Frobenius-Operator equation [2] for piecewise linear transformations.

In summary, the SRB measure λ has the following properties:

1. λ is 'aware' of almost all points in its support as evinced by (4).

2. Almost all orbits settle into λ and therefore it acts as a global attractor.
3. λ determines the dynamics T.

3. BRAIN FUNCTION AS A DYNAMICAL SYSTEM AND THE SRB MEASURE

We now introduce the special setting of [3] in which brain operation is viewed as a dynamical system on the space X of compact subsets of the physical space B, where B is that portion of the brain-matter in which processing of conscious brain images takes place. (We will have more to say about this in the sequel.) B is a subset of three-dimensional space, or a smooth manifold, which means that locally, around each point, the brain-matter resembles a segment of three-dimensional space.

A point $x \in X$ is a compact subset of B; it represents a collection of activated neurons. T is a transformation from X into X and operates on compact subsets, that is, on collections of activated neurons to produce a collection of activated neurons. The actual physical mechanism by which a brain image is transformed to another brain image is not important. What is important is the fact that, at the end of a unit of time (of the order of a second), a new brain image is created. This process is modelled by a discrete time transformation T.

We partition X into subsets according to where the brain images originate from, as shown in Fig. 3. Thus, P_1 is the collection of brain images induced by vision, P_2, by touch, and so on. Without loss of model generality, we postulate that there are 10 such subsets, namely the images associated with the five sensory inputs: vision, touch, sound, smell, taste, and the brain images associated with speech, thought, memory, motor activity and a category we refer to as 'miscellaneous' which collects the brain images associated with emotions, say, and other processes that we may have neglected. The actual number of these subsets of brain images is not relevant to the main idea of the present model. Let $\{P_1, P_2, \ldots, P_q\}$ denote the collection of the subsets forming the partition P of X. We claim that any P_i can be
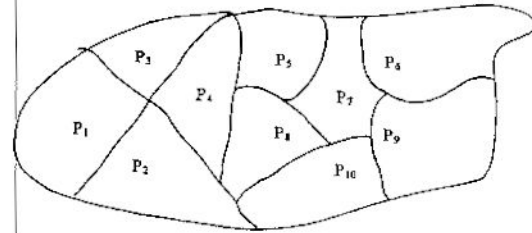

FIGURE 3 Partition P of X, the space of all brain images relevant to consciousness.

mapped by T onto all of X, that is, that any image in any P_i can be transformed by T to any other image in X. For example, a smell image can give rise to a thought image as occurs, for example, in the novel by Proust where the scent of a madelaine triggered an avalanche of thoughts. Analogously, an image associated with a motor action can stimulate an image associated with a sound. With the onto assumption, T becomes Markov.

In a more realistic model, elements of the partition of X map onto some – but not necessarily all – partition elements of X. The actual condition needed is irreducibility of the adjacency matrix [2], which is trivially satisfied if the transformation is piecewise onto. In general we need neither Markovness nor onto-ness. What is essential is an expanding property [7] or a folding property which is at the heart of chaotic behaviour [3]. As a starting point in modelling 'mind' and in order to expedite the presentation of the model, we are content with a simple Markov (piecewise onto) model of the brain image transformation T.

We now discuss how the SRB measure manifests 'awareness' of the brain. With the assumption of an 'interesting' transformation T (such as one that is Markov and piecewise onto or expanding), we have the existence of an SRB measure μ. Let us consider a brain orbit: for example, one that starts with a memory image which leads to a thought image which leads to a physical action image (such as a sigh image) which in turn leads to another thought image which leads to another memory image. This is a transitory brain image orbit, which must

ultimately – with loss of concentration, be nudged into a brain image (in the set of μ-measure 1) that exhibits the SRB measure by virtue of the averaging process of (3). Based on the foregoing, a possible brain experiment might go as follows: a brain is lured into forced concentration for some length of time and then the concentration is released; the brain image orbit during the transition from concentration to non-concentration is traced by PET scans. The images that occur after the release from concentration might shed light on the orbit of images that leads to μ. Methods for analyzing images are developed in [11,17].

Given a brain B, the space of images X on B, and a partition P of X together with the SRB measure μ, it is possible – as in the one-dimensional case – to derive the image transformation within a class of piecewise linear transformations [2]. This implies that μ is 'aware' of brain (Eq. (4)) and brain dynamics T.

4. INFORMATION RATES

The space of images X and the Markov transformation $T: X \to X$ constitute what is known as a Bernouilli shift [1], where the measure is the SRB measure μ has mass p_i on each element P_i of the partition P. Since $\mu(X) = 1, p_2 + \cdots + p_q = 1$, where $p_i > 0$. Hence the information rate of the brain dynamical system is given by

$$H(P, T) = -\sum_{i=1}^{q} p_i \log_2 p_i. \tag{5}$$

Equation (5) yields the information rate of the brain dynamical system treated as a single channel. The conscious brain, however, is able to handle many inputs at the same time – in effect acting as a parallel processor. Hence we assume it can handle input images from any of the $q = 10$ sources of conscious brain images simultaneously. Thus, we are dealing with the brain $B = B_1 \times B_2 \times \cdots \times B_q$ and the subsets given by $X = X_1 \times X_2 \times \cdots \times X_q$. The information rate for this q-parallel system is

given by [13, Theorem 2.50]

$$H_{\text{total}}(P, T) = -q \sum_{i=1}^{q} p_i \log_2 p_i. \tag{6}$$

Hence, the maximum information rate for this q-parallel system is $q \log_2 q$. For $q = 10$, the maximum information rate is approximately 33 bits/s which accords well with the estimate of between 15 and 50 bits/s given in [9, p. 60] by means of a completely different line of reasoning.

5. UNIVERSAL MODEL OF MIND

Let us now consider N brains $\{B_1, B_2, \ldots, B_N\}$ functioning at the same time. Since these brains are independent of each other we can display their dynamics on the space of images $X = X_1 \cup X_2 \cup \cdots \cup X_N$, where X_i denotes all the brain images on B_i. Let $T_i: X_i \to X_i$ be the image transformation on X_i Then $T_i(X_i) \subseteq X_i$, i.e., each space of images X_i is invariant under T_i. In Fig. 4 we show this multi-brain situation symbolically.

By changing T_i slightly we can obtain an interconnected brain system. Let us suppose that for each brain X_i, T_i is altered to $T_i^\sim$ as follows: on a very small set I_i in X_i, we let $T_i^\sim(I_i) = X$, that is, $T_i^\sim$ carries
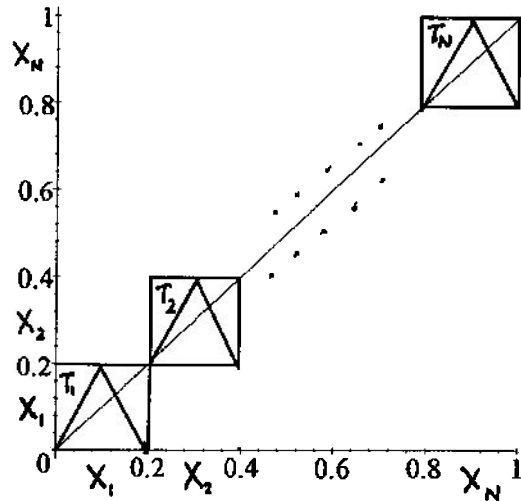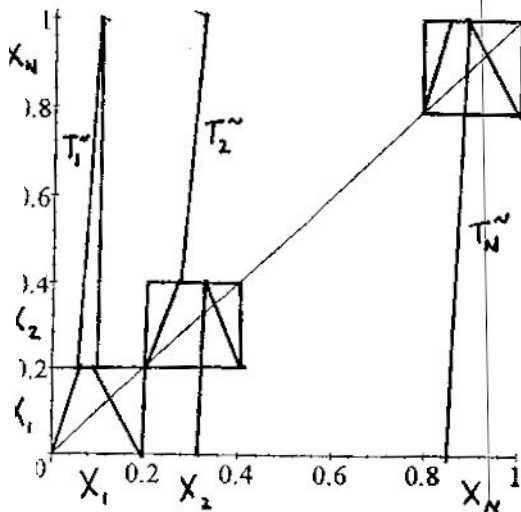

FIGURE 4 Independent brains.

FIGURE 5   Interacting brains.

a small set of special images to images in every other brain. This can be effected by a transformation such as is shown in Fig. 5.

For the transformations as shown in Fig. 5, there exists a global SRB measure μ on all of X which, when restricted to any B_i, is approximately equal to μ_i, the SRB measure for $(X_i, T_i^{\sim})$. Although an orbit in X_i will spend most of its time inside X_i it is nonetheless able to leave X_i – by hitting the right small collection of images – and then to wander through all of X, the universal brain. The global SRB measure μ ensures that there are dense orbits on all of X. These orbits connect all the brains. Finally if $T_i^{\sim}$ is very sharp in the region where it extend into other brains, communication with other brains can take as little as one iteration of $T_i^{\sim}$.

6. CONNECTION TO QUANTUM MECHANICS

In quantum mechanics, the wave function $\psi(x, t)$, when squared, is interpreted as a probability density function. Let us fix t and define $f_i(x) = |\psi(x, t)|^2$. Then $f(x)$ is the probability density function of a diffusion process [15]. This process, which is a generalized Brownian motion, can be modelled by

a point transformation or random map [8] whose SRB measure is defined by $f_i(x)$, i.e.,

$$\mu_t(A) = \int_A f_t(x)\, dx.$$

7. OBSERVATIONS

(i) Location of Mind

The existence of an SRB measure μ on X implies that the support S of μ ($\mu(S) = 1$) is a set that is invariant, i.e., that satisfies the condition $T(S) \subseteq S$. This suggests the existence of a location in the brain where a cluster of neurons can be directly connected to itself. Since the synaptic knobs on the axons of the neurons in S are likely to produce activity far removed from S, it is likely that the neurons in S interact via dendrodendritic synapses rather than axon synapses. Such a densely packed cluster of neurons with highly interacting dendrites would be a candidate for the seat of mind. Transforming images by means of dendrodendritic synapses is efficient (in that messages do not have to be relayed long distances) and is therefore less prone to error. In Crick's analysis of vision [5], the observed dense interconnections between layer 6 of the visual cortex and the thalamus suggests a location for the support of an SRB measure.

It is of interest to note that a brain image is a point in a very high dimensional space R^n [3] and, in this space, we can obtain a great diversity of images all contained in a small physical ball of R^n.

(ii) Mind, Body, Reality

In the mathematical formalism of this note, mind means the existence of the SRB measure μ on the space X of brain images, and body means the brain-matter together with the space of images X. The Birkhoff Ergodic Theorem for the system (X, T, μ) describes the intimate way in which mind and body are intermeshed. Since time and space are effectively the same in the Birkhoff theorem, we have the

following symbolic equation:

$$X(\text{body, brain}) + \textit{time}(\text{flow of time for an interesting transformation}) = \mu(\text{mind}).$$

The flow of time as enacted by iteration of an interesting tranformation T(Markov, expanding, folding) brings μ(mind) into actuality.

(iii) Relative Consciousness

In the human brain the thought process is sufficient to produce a folding transformation on the space of images since images produced by senses can be duplicated by thought. Hence, the human brain possesses interesting dynamics and an SRB measure. This is not the case for an inanimate object: a rock has no dynamics at all relative to itself and hence cannot give rise to any invariant measure that might have properties of mind. However, relative to a point outside the rock, it may have interesting dynamics. For example, consider the planets and stars moving relative to a (central) point C in the universe. It is conceivable that this dynamical system is governed by sufficiently complicated dynamics relative to C; then the collection of inanimate objectives would possess a 'relative mind' with respect to C. In this case, the role of the planets is analogous to that of the neurons in the brain, where each neuron does not exhibit mind on its own.

But it appears that only man has mind relative to himself by virtue of his ability to think, which is probably the minimal condition that gives rise to a folding transformation (and hence possibly an SRB measure) on the space of brain images.

(iv) Timelessness of Mind

Let the dynamical system (X, T) possess an SRB measure μ. Since μ is supported on S, which consists of pieces of the physical space B, S and μ are bound by the very existence of μ. Even if the underlying space B disintegrates, μ continues to exist, even if merely as a mathematical entity. Thus, mind – once created by an interesting transformation and the flow of time – exists forever after.

References

[1] P. Billingsley, *Ergodic Theory and Information*, R.E. Krieger Publishing Co., 1978.
[2] A. Boyarsky and P. Gora, *Laws of Chaos*, Birkhauser 1997.
[3] A. Boyarsky, P. Gora and V. Lyabumov, Dynamics on spaces of compact subsets with application to brain modeling, *Jour. Math. Anal. Appl.* **216**, 569–580, 1997.
[4] F. Crick and C. Koch, The problem of consciousness, Scientific American, Sept. 1992.
[5] F. Crick, *The Astonishing Hypothesis, The Scientific Search for the Soul*, Scriibners, N.Y., 1994.
[6] R. Devancy, *An Introduction to Chaotic Dynamical Systems*, 2nd edition, Addison Wesley, 1989.
[7] P. Gora and A. Boyarsky, Absolutely continuous invariant measures fo piecewise expanding C^2 transformations in R^n, *Israel Jour. Math.*, **67**(3), 272–286, 1989.
[8] P. Gora and A. Boyarsky, A lattice spacetime for the slit experiments of quantum mechanics, Physics Letters A 236, 263–269, 1997.
[9] N. Herbert, *Elemental Mind*, Dutton, 1993.
[10] J. Horgan, Can science explain consciousness? *Scientific American*, July 1994.
[11] D.G. Kendall and E.F. Harding, *Stochastic Analysis*, John Wiley, N.Y., 1974.
[12] R. Mane, *Ergodic Theory of Differentiable Dynamics*, Springer-Verlag, N.Y., 1987.
[13] N. Martin and J. England, *Mathematical Theory of Entropy*, Addison-Wesley, 1981.
[14] W. de Melo and S. van Stricn, *One-dimensional dynamics*, Springer-Verlag 1993.
[15] E. Nelson, *Dynamical Theory of Brownian Motion*, Princeton University Press, 1967.
[16] R. Penrose, *The Emporer's New Mind*, Oxford University Press, 1989.
[17] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, N.Y., 1982.