Rate Distortion Manifolds as Model Spaces for Cognitive Information

James F. Glazebrook
Department of Mathematics and Computer Science
Eastern Illinois University
600 Lincoln Avenue
Charleston IL 61920–3099, USA
E–mail: jfglazebrook@eiu.edu

Rodrick Wallace
Division of Epidemiology
The New York State Psychiatric Institute
Box 47, 1051 Riverside Drive
New York NY 10032, USA
E–mail: wallace@pi.cpmc.columbia.edu

Keywords: global workspace, entropy, cognition, rate distortion function, giant component, groupoid, stochastic resonance.

Received: December 18, 2007

The rate distortion manifold is considered as a carrier for elements of the theory of information proposed by C. E. Shannon combined with the semantic precepts of F. Dretske's theory of communication. This type of information space was suggested by R. Wallace as a possible geometric–topological descriptive model for incorporating a dynamic information based treatment of the Global Workspace theory of B. Baars. We outline a more formal mathematical description for this class of information space and further clarify its structural content and overall interpretation within prospectively a broad range of cognitive situations that apply to individuals, human institutions, distributed cognition and massively parallel intelligent machine design.

Povzetek: Predstavljena je formalna definicija prostora za opisovanje kognitivnih procesov.

1 Introduction

The concept of an *information space* seems to have various definitions and interpretations within the mathematical and life sciences literature. The quest for suitable models of cognitive processes in the large scale is likely to involve an eclectic array of techniques drawn from topology, geometry and category theory. For it appears sensible, if not absolutely necessary, to introduce structures that are comparatively weaker than the conventional 'smooth manifold' approach as it is often employed in mainstream dynamical systems, statistical inference, automata, and pattern/speech/facial recognition. Whereas in such situations we do not dispute the usefulness of Finsler or Riemannian structures (e.g. the well–known Fisher information metric about which much as been written), their comparative rigidity does not allow the flexibility of topological modeling which is necessary for the more organismic–like nature of cells of information as they function in the *local-to-global processing* of cognitive mechanisms. The quest then is to provide a descriptive framework of cognitive–interactive modules which are 'organisms' in their own right, possessing their own internal dynamics, semantic language, channels of communication and their own 'immune' systems. This affords them further richness of content compared to the schemata of classical neural networks, or to the over–simplified flow boxes of cybernetic processes and other stock–in–trade tools of traditional AI.

Wallace [125] has brought together the fundamental ideas of F. Dretske's semantic communication theory [39] with the Global Workspace theory of B. Baars [12] within a framework governed by the acclaimed theory of information of C. E. Shannon [32, 112] originating with necessary conditions gauging the reliability of a source entropy rate relative to a channel capacity. Subsequent ideas are blended in with the mathematical models for statistical mechanics/chemical thermodynamics as introduced by L. Onsager along with K. Wilson's theory of renormalization. In this way the overall treatment incorporates several notable examples of 20th century mathematical–physical creation. Motivated by the main results of Shannon's theory, Wallace [125] introduced the concept of a *rate distortion manifold* as a descriptive model for analyzing a schemata of information–based cognition in a range of contexts (such as e.g. psycho–social epidemics, sleep patterns, obesity, stress related illnesses, inattentional blindness and the language of the immune system [126, 127, 128, 129]). A prime ingredient used here is the concept of the *rate distortion function* $R(D)$ that characterizes the minimum rate of information needed to reproduce a source message within a specified fidelity D.

Although the basis for having a manifold atlas topology had been suggested in these works, there remains the task for a comprehensive description of a rate distortion manifold in more formal geometric and topological terms as a means of understanding and better representing the intrinsic dynamics underlying a wide range of cognitive processes. These may involve direct comparisons between individual/institutional cognition on the one hand, and distributed cognition/massively parallel computation on the other. In this way we may gain some insight into how failures in one category induce corresponding causes of failures in another.

Here we propose such a formal and conceptual treatment of the rate distortion manifold within a stretch of mathematical ideas, and at the same time provide a discussion of how these ideas are applicable to a variety of cognitive situations. The 'weaker' structures, incumbent to an extent on the abstract principles of the topology of manifolds and the theory of categories, are preferred over the less flexible structures which appear in the examples mentioned previously. More specifically, we consider certain topological spaces which are intrinsically path spaces over an alphabetical–coding system in the Dretske semantics of communication and adhering to the Shannon theory [125]. But viewed as particular (rate distortion) metric spaces with length structure, they are nevertheless conducive to admitting manifold/atlas topologies in a general sense, and moreover, may enjoy a sufficiently weaker notion of 'differentiability' beyond the framework of classical differential calculus. Putting it another way, we propose a rate distortion manifold to be a general 'cognitive information space' adapted to, or designed towards, analyzing a particular cognitive situation.. Such a space admits a manifold/atlas topology to encompass its framework, and which serves as *a descriptive 'cell' or 'organism' of cognitive information* gauged as such by certain principles of information/entropy and statistical physics.

By 'cognitive', we mean pertaining to cognition at large; that is, in relationship to the neurosciences, cross–cultural psycho–sociological and epidemiological/immunological phenomena (both *distributed* and *institutional*). In the context of the Global Neuronal Workspace [12], such an 'organism' is a constituent module of a 'broadcasting system' that cooperates or competes within a specified hierarchy of contexts. Symmetry-breaking within the underlying information network engenders a phase transition, and thus the complexity of information increases. The cultural models of [125, 135] apply not only to individual cognition, but also to the large–scale cognition of socio–political–environmental mechanisms which inevitably shape the vital drive of the former.

Underlying our proposed information spaces are cellular and combinatorial structures as, for instance, provided by the general topological notion of a *CW–complex* or of a *simplex* (obtained via a simplicial decomposition of a given topological space) which provides a route into the graph–theoretic concepts of mainstream network analysis. There is yet another categorical twist. We propose that the structures of our information spaces be richer and deeper than abstract categories, but it is even more meaningful if the actual objects of a given category are the information spaces themselves. The other slant is that these spaces may be reasonably viewed in terms of 'small' categories with invertible morphisms, that is, they can be realized via path equivalence as *groupoids* and therefore can be treated quite appropriately within the framework of the algebraic topology of groupoids, their actions and atlases of such (see e.g. [16, 21, 143] and Appendix I of this paper). Applicable to a broad range of cognitive models, this is a central idea already advanced in [125](cf [60]) and can be compared with recent approaches to biological and neural network systems as adopted by a number of other authors (see e.g. [15, 61, 66, 67, 118] and the many references therein). The cognitive 'cells' or 'modules' of information founded on rate distortion manifolds are also relevant in the modeling of *autopoietic* and related systems [93](cf [34, 60, 70]) which will be discussed in a later section. Future directions (which we will address later) are likely to involve the more ambitious tasks of dealing with informational processes and rate distortion theory within the context of such general categorical concepts as *groupoid atlases*, *stacks in groupoids* and the general concept of a *topos*.

2 Information spaces for general cognition

2.1 Characteristics of a cognitive information space

Let us mention that Chalmers [25](Chapt. 8), for instance, has suggested a possible framework for linking the processes of the physical and phenomenal worlds in terms of a conceptually–based information science cast within the mathematical methods of geometry and topology. This plan has a significant overlap with the development in [125] and the more formal treatment described herein. Such an approach seems imperative for advancing the Workspace structure within the Theater of Consciousness [12] by means of semantic information and dynamical processes. Moreover, the proximity to our formal description of a rate distortion manifold as an information space, will eventually become relevant. We rephrase the overall requirements within the scope of our proposed development:

1. The model assists in addressing the phenomenological aspect (states of context) in relationship to physical reality.

2. An information space could be viewed as in part rep-

resenting a 'causal pathway' embedded within some culture, but included are semantic, dynamic principles seeking to incorporate states of experience, properties often lacking in traditional cognitive theories.

3. We consider a model based upon the structure of contexts and language of thought within fluctuating paths of information, not entirely in the discrete sense, but possibly within a suitable notion of a 'continuum limit', a 'manifold–atlas topology', or a 'dynamic categorical process', as examples of possible working environments.

4. The manifold–atlas topology can be coupled with an information based, (weakly) stochastic structure where essential distinctions can be represented in terms of such properties as homotopy and diffeomorphism type, homology, curvature, etc. Thus we turn to state of the art techniques of differential geometry/topology, category theory, on the one hand, and on the other, to combine these techniques with those of information theory, stochastic processes and statistical mechanics. This spread of ideas reflects upon the eclectic framework proposed for extending the traditionally acceptable descriptive methods for studying cognitive processes and their emergence through orders of complexity.

5. By introducing simplicial methods to analyze the underlying combinatorial structure of the manifold, we may recover graph-theoretic models as suited to the navigation through various types of information highways [2], systems of coding, symbolic dynamics [89] and complexity [88].

6. The approach aims at constructing a geometrical/topological carrier for the Shannon information theorems about which the rate distortion manifold is formulated, thus leading to a framework for interacting cognitive modules upon which the prevailing cultural environment inevitably writes its image. Essential here is that the manifold accommodates a homology between the corresponding informational laws of asymptotic probability and certain thermodynamic limits of statistical mechanics. In this way, altered states of cognitive processes can be seen to be caused by phase transitions analogous to how the latter can induce sharp transformations between one thermodynamic state to another.

A wide range of descriptive possibilities are likely. Of these we could view the structure of awareness and experience as represented within the structure of an information space with phenomenal states. Conversely, such a predictive representation may feed its way back into the cognitive system with the enhanced prospects of obtaining an improved model which may lead to the eventual solution of a given problem.

A limitation of classical information theory is that it was not preoccupied with semantics. The theory was destined originally for the testing ground of noisy telephone exchanges – some time before 'The Brain is a Noisy Processor' became a standard assumption. Information in Shannon's theory evolved essentially within a combinatorial/probabilistic framework for representing how states are manifest within an information space. Its main constituents include:

a) Application of the asymptotic limit theorems of probability theory.

b) Mutual information.

c) The Shannon Coding Theorem (fixing signal and oscillator) assumes an optimal coding scheme involving noise so that the rate of error–free output of the signal will attain some positive value.

2.2 The Global Workspace

A principal aim is to apply rate distortion manifolds as descriptive features of the Global Workspace theory of [12, 14]. The general dictum goes as follows [14]:

(1) The brain can be viewed as a collection of distributed specialized networks (processors).

(2) Consciousness is associated with a Global Workspace in the brain – a fleeting memory capacity whose focal contents are widely distributed (broadcast) to many unconscious specialized networks.

(3) Conversely, a Global Workspace can serve to integrate many competing and cooperating input networks.

(4) Some unconscious networks called 'contexts', shape conscious contents. For example, unconscious parietal maps modulate visual feature cells that underlie the perception of color in the ventral stream.

(5) Such contents work together jointly to constrain conscious events.

(6) Motives and emotions can be viewed as goal contexts.

(7) Executive functions work as hierarchies of goal contexts.

Recent research [37, 38] has enhanced the validity of this model. As pointed out in [126], the special properties of representing embedding and interpenetrating contexts provide a framework for understanding the synergism of consciousness and mental disorders in humans within a socio–cultural context. This framework bears startling analogies with the institutional cognition of epidemics versus the public health sector as a phenomenon of disorder of information [129, 137, 140]. Indeed, institutions such as the latter may themselves function within their respective cultural environments as 'distributed' cognitive systems having their own sovereign mechanisms, implicitly different from that of humans but nevertheless influencing the degree of effectiveness of human involvement [70], a situation closely in tune with the organisms of environmental autopoietic systems [93] (see §9.4). But these large–scale cognitive mechanisms, although not constrained so much by biological evolution, are certainly prone to analogous cognitive disorders such as enviromental (psycho–social) stress, inattentional blindness, (social) network failure and many other ailments that plague human society.

As incorporating fundamental geometric techniques, rate distortion manifolds each possessing characteristic topological and geometric properties, along with their own internal dynamics, are thus proposed as descriptive cells within this blueprint for cognition. As abstract topological manifolds they are designed to model the shape and flow of information that can be adapted to analyze a broad range of cognitive situations. The next stage is to outline a more specific mathematical description of their structure.

3 Towards a rate distortion manifold

3.1 Manifold–atlas topology

The standard concept of a 'differentiable manifold' as to be found in e.g. [1, 87], is of long–standing importance in geometry and physics. However, for the sake of the more flexible structures as sought after, we need to have a handle on an even more general concept, namely that of an *atlas–manifold topology* (such as to be found in e.g. [21]).

The idea is to start by defining a weaker notion of 'function' valid for set-valued mappings. Let A and B be sets and consider a triple (A, B, F) where $F \subset A \times B$, with the property that if $(a, b), (a, b') \in F$, then $b = b'$. Such a triple is called a *partial function* between A and B, denoted $f : A \rightarrowtail B$, and written $f(a) = b$. The domain of f is the set of $a \in A$ such that $f(a)$ is defined. The concepts of composition, continuity, etc. apply in accordance their usual topological definitions. The domain of f can be any subset of A, and if B is a some scalar field such as $\mathbb{R}$ or $\mathbb{C}$, then the definition reduces to the standard one for that of a function.

Consider a set A and a family $\{A_\lambda\}_{\lambda \in \Lambda}$ of topological spaces, together with a partial function $f_\lambda : A \rightarrowtail A_\lambda$, for each $\lambda \in \Lambda$. A topology $\mathcal{T}$ on A is said to be *initial* with respect to $\{f_\lambda\}$ if for any topological space B, a partial function $k : B \rightarrowtail A_{\mathcal{T}}$ is continuous if and only if the composition $f_\lambda \circ k : B \rightarrowtail A_\lambda$ is continuous. Such a topology on A is the coarsest of topologies such that each $f_\lambda : A_{\mathcal{T}} \rightarrowtail A_\lambda$, is continuous.

Let E be a topological space and let M be a set. An *E–chart* on M is an injective partial function $\varphi : M \rightarrowtail E$ whose image is open in E. For some indexing set $\mathcal{I}$, an *E–atlas* for M consists of a family $\mathcal{A} = \{\varphi_\alpha\}_{\alpha \in \mathcal{I}}$ of E–charts for M such that if $\varphi_\alpha, \varphi_\beta : M \rightarrowtail E$ are charts in $\mathcal{A}$, then the composition $\varphi_\beta \circ \varphi_\alpha^{-1} : E \longrightarrow E$, is continuous.

Suppose then we are given such an E–atlas $\mathcal{A}$, and let M have the initial topology with respect to all E–charts in $\mathcal{A}$. Then $\varphi_\alpha^{-1} : E \rightarrowtail M$ is continuous, since $\varphi_\beta \circ \varphi_\alpha^{-1} : E \longrightarrow E$, is continuous for all $\beta \in \mathcal{A}$. Therefore, φ_α maps its domain homeomorphically to its image. We may call E the *model space* of the atlas in accordance with the terminology in the case where M is a manifold in the more concrete sense, and when E is some suitable vector space (which could be infinite dimensional).

In order to realize a suitable cognitive information space, we probe beyond some of the typical manifold structures of information geometry and so the above abstraction has potential value. Information geometry can involve using parametric and non–parametric probability densities in order to construct appropriate statistical manifolds for inference. In the parametric case, befitting a Fisher metric structure say, the manifold can be treated from the point of view of a Euclidean topology, whereas in the non–parametric case (useful for e.g. perception/recognition as in [122]), other topologies have to be considered leading to infinite dimensional manifolds modeled, say, on spaces of operators (such as projections) and which include special Banach spaces with differing topologies (e.g. exponential convergence as introduced in [101] and applied in [145]).

In this respect the model space E might be taken as a carrier space for operator–valued probability densities, suitable say, for dealing with 'sharp' or 'fuzzy' stochastic processes (as in e.g. [64]). Manifolds for statistical inference, stochastic processes and those serving as a descriptive mechanism for modeling the various information highways are seemingly too rigid in structure for effectively describing cognitive cells. In the latter case, we consider these as significantly influenced by linguistic and cultural factors. Additionally, there are potentially useful structures weaker than the standard manifold/atlas, such as that of an *orbifold atlas* which would accommodate certain types of singularities [96] (see Appendix I §11.2). More generally, there is the notion of a *groupoid atlas* [16] (see §7.4) which incorporates groupoid actions and thus may be viewed as an abstract dynamical system in its own way (see §7.4).

3.2 The information space (X, s_X)

More specifically, suppose $E = E^\Gamma$ is a high dimensional state space modeled on some 'alphabetical/coding/syntatical' structure denoted Γ. This is instrumental for a semantic base–model following the dictum of F. Dretske [39, 125]. In mathematical terms, we grant the possibility that E as a kind of *state space* may be formally structured as a vector space which may possess certain properties such as local convexity, etc. For instance, E could be taken to be the underlying vector space of a general *events algebra* in the sense of [64]. We also leave open the possibility that E is endowed with some norm denoted $\| \cdot \|$, although we may not always insist on this property.

We proceed to consider a set $X \subset E$, where points $x \in X$ correspond to paths of convoluted signals; typically, $x = (a_0, a_1, \ldots, a_n, \ldots)$ where a_k represents some functional composition of internal and external signals. In this respect

X could be deemed to be a 'weak path space' over E. It is quite possible that X could be considered as having a filtered or cellular structure (as will be described below). The path space (X, s_X) with model (or atlas) space E, is considered as a metric space with metric $s_X \equiv d(x, \hat{x})$ induced by a distortion measure d (see §3.4). This measure then leads to defining a rate distortion function complying with Shannon's theorem (see §3.4 and §4.3 later).

Following [125], suppose we take a decision oscillator generating an output as given by a set valued (partial) function $h : X \longrightarrow B$. For instance, we set $B = B_0 \cup B_1$ where

$$h(x) \in B_0 \equiv b_0, \dots, b_k, \tag{3.1}$$

if the pattern is not recognized, and

$$h(x) \in B_1 \equiv b_k, \dots, b_m, \tag{3.2}$$

if the pattern is recognized.

The set B is prospectively one that is highly extensible and could be viewed as the underlying set of a suitably constructed algebra of responses or events. Also, the fact that higher order cognitive decisions and several options of response along a given path are likely to be necessary, suggests further intrinsic properties needed for sets of the type B_0 and B_1.

Remark 3.1. Note that patterns may well undergo a filtering in stages of recognition. Thus a generalization is to suppose that B_0 and B_1 admit countable filtrations of the sort:

$$B_0 = B_0^0 \subseteq B_0^1 \subseteq B_0^2 \subseteq \cdots$$
$$B_1 = B_1^0 \subseteq B_1^1 \subseteq B_1^2 \subseteq \cdots$$

where at level j we have set $B_0^j \equiv b_0^j, \dots, b_k^j$, and $B_1^j \equiv b_{k+1}^j, \dots, b_m^j$.

3.3 The Shannon entropy

Shannon conceived of *entropy* as a measure H of the capacity of a communications system to transmit information. The idea was to directly tie a given response rate $r(t)$ to a function of the probability of achieving $r(t)$. In a more specific way we will recall below some of the basic results of the theory in terms of *meaningful paths*.

For each $n \in \mathbb{N}$, let $N(n)$ denote the number of paths of length n beginning with a particular a_0 with $h(a_0) \in B_0$, and leading to the condition that $h(x) \in B_1$. We call such paths *meaningful* and, for cognitive reasons, regard $N(n)$ to be much less than the number of all paths of length n . Further, we assume that the limit

$$H \equiv \lim_{n \longrightarrow \infty} \frac{\log[N(n)]}{n}, \tag{3.3}$$

exists, is finite, and is independent of the path x. Such a cognitive process is then said to be *ergodic*. The non–ergodic case (more pertinent to cognition) will be discussed later.

Relative to the path space (X, s_X), we define a corresponding ergodic information source $\mathbf{X}$ with stochastic variables $\mathbf{X}_j$ having joint and conditional probabilities $P(a_0, \dots, a_n)$ and $P(a_n | a_0, \dots, a_{n-1})$ respectively, so that the joint and conditional Shannon probabilities may be defined and satisfy the relations [32]:

$$\begin{aligned} H[\mathbf{X}] &= \lim_{n \longrightarrow \infty} \frac{\log[N(n)]}{n} \\ &= \lim_{n \longrightarrow \infty} H(\mathbf{X}_n | \mathbf{X}_0, \dots, \mathbf{X}_{n-1}) \\ &= \lim_{n \longrightarrow \infty} \frac{H(\mathbf{X}_0, \dots, \mathbf{X}_n)}{n+1}. \end{aligned} \tag{3.4}$$

Such an information source is considered to be *dual* to the ergodic process.

Remark 3.2. Technically, $\mathbf{X}$ is taken to be an *adiabatically, piecewise stationary, ergodic* (APSE) information source (see e.g. [6, 32] and explanations relative to cognitive modules in [125, 135]).

The Shannon–McMillan theorem provides 'a law of large numbers' and permits the definition of uncertainties in terms of cross sectional sums of the form

$$H = -\sum_k P_k \log[P_k], \tag{3.5}$$

where the P_k are derived from a probability distribution and satisfy $\sum_k P_k = 1$. Different languages can be defined by different divisions of the total universe of possible responses into various pairs of sets B_0, B_1 above, or by insisting upon more one than response in B_1 along a path. Allocating the full set of possible responses into B_0, B_1 may necessitate engaging higher order cognitive decisions.

3.4 The Rate Distortion Theorem

Following [125], suppose we have an (ergodic) information source $\mathbf{Y}$ with output from a particular alphabet generating sequences of the form

$$y^n = y_1, \dots, y_n \tag{3.6}$$

'digitalized' in some sense, and inducing a chain of 'digitalized' values

$$b^n = b_1, \dots, b_n \tag{3.7}$$

where the b–alphabet is considered more restricted than the y–alphabet. In this way, b^n is *deterministically retranslated* into a reproduction of the signal y^n. That is, each b^n is mapped onto a unique n–length, y–sequence in the alphabet of $\mathbf{Y}$:

$$b^m \longrightarrow \hat{y}^n = \hat{y}_1, \dots, \hat{y}_n. \tag{3.8}$$

We remark that many y^n sequences may be mapped onto the same retranslation sequence $\hat{y}^n$, the set of which is denoted $\widehat{\mathbf{Y}}$; this may be interpreted as a loss of information.

A *distortion measure* $d : \mathbf{Y} \times \widehat{\mathbf{Y}} \longrightarrow \mathbb{R}^+$, between paths y^n and $\hat{y}^n$ is defined as

$$d(y^n, \hat{y}^n) = \frac{1}{n} \sum_{i=1}^{n} \underline{d}(y_j, \hat{y}_j), \tag{3.9}$$

for some suitable distance function $\underline{d}$ (such as the Hamming distance). Suppose that with each path $y^n \in \mathbf{Y}$ and each b^n–path retranslation $\hat{y}^n \in \widehat{\mathbf{Y}}$ into the y–language, we consider the associated individual, joint, and conditional probability distributions

$$p(y^n)\ ,\ p(\hat{y}^n)\ ,\ p(y^n|\hat{y}^n). \tag{3.10}$$

The *average distortion* is then defined to be

$$D = \sum_{y^n} p(y^n)\ d(y^n, \hat{y}^n). \tag{3.11}$$

For the corresponding strings $\mathbf{Y}$ (incoming), $\widehat{\mathbf{Y}}$ (outgoing), the Shannon uncertainty rule is

$$\begin{aligned} I(\mathbf{Y}, \widehat{\mathbf{Y}}) &\equiv H(\mathbf{Y}) - H(\mathbf{Y}|\widehat{\mathbf{Y}}) \\ &= H(\mathbf{Y}) + H(\widehat{\mathbf{Y}}) - H(\mathbf{Y}, \widehat{\mathbf{Y}}). \end{aligned} \tag{3.12}$$

Definition 3.1. The *information rate distortion function* $R(D)$ for a source sequence $\mathbf{Y}$, a retranslated sequence $\widehat{\mathbf{Y}}$, along with a distortion measure $d : \mathbf{Y} \times \widehat{\mathbf{Y}} \longrightarrow \mathbb{R}^+$, is defined as follows.

Let $\Upsilon = \sum_{(y,\hat{y})}\ p(y)\ p(y|\hat{y})\ d(y, \hat{y})$. Then

$$R(D) = \sum_{p(y,\hat{y})\ :\ \Upsilon \leq D} I(\mathbf{Y}, \widehat{\mathbf{Y}}). \tag{3.13}$$

To explain this notation, the minimization is over all conditional distributions $p(y|\hat{y})$, for which the joint distribution $p(y, \hat{y}) = p(y)\ p(y|\hat{y})$ satisfies average distortion less than or equal to D.

The *Rate Distortion Theorem* (see e.g. [32, 36]) states that $R(D)$ *is the maximum achievable rate of information which does not exceed the distortion* D.

These are some of the basic ingredients for considering the optimal rate of precise information transfer in relationship to channel capacity and to which extent noise is hazardous to the system. The path–space modeled rate distortion manifolds in question, are assumed to comply with the above theorem implicitly. Moreover, as far as cognition is concerned, [125] postulates a fundamental homology with thermodynamic processes, quite similar to how distortion and fidelity in network information can be studied involving techniques of statistical physics such as the Ising lattice and spin–glass networks in conjunction with the usual industry of error correcting and coding (for related work in this direction see e.g. [98, 89, 116]).

3.5 Channel capacity

The *channel capacity* is defined to be

$$C \equiv \max_{P(\mathbf{X})} I(\mathbf{X}|\mathbf{Y}), \tag{3.14}$$

subject to the subsidiary condition that $\sum P(\mathbf{X}) = 1$. This is a measure of the maximum transmission rate of information across a channel with the likelihood of error tending to zero. Effectively, the critical trick of the Shannon Coding Theorem for sending a message with arbitrarily small error along the channel $\mathbf{Y}$ at any rate $R < C$, *is to encode it in longer and longer 'typical' sequences of the stochastic variable* $\mathbf{X}$*; that is, those sequences whose distribution of symbols approximates the probability distribution* $P(\mathbf{X})$ *above which maximizes* C.

Thus for an information source X, the Shannon entropy $H(\mathbf{X})$ as given in (3.3), can be seen to satisfy for a given channel capacity C, the inequality

$$H(\mathbf{X}) \leq C. \tag{3.15}$$

If $S(n)$ is the number of such 'typical' sequences of length n, then $\log[S(n)] \approx nH(\mathbf{X})$. Some consideration shows that $S(n)$ is much less than the total number of possible messages of length n. Thus, as $n \to \infty$, only a vanishingly small fraction of all possible messages is meaningful in this sense. This observation, after some considerable development, is a principle that allows the Shannon Coding Theorem to work so well. In sum, the prescription is to encode messages in typical sequences, which are sent at very nearly the capacity of the channel. As the encoded messages become longer and longer, their maximum possible rate of transmission without error approaches channel capacity as a limit (for details see [32, 79, 112]).

Rate distortion manifolds may be characterized by a type of inversion of this procedure. Examples of noisy channels are telephone lines, optical wave guides and interplanetary plasmas around which a message is to be structured so as to attain an optimal error-free transmission rate. These examples are, relatively speaking, fixed on the timescale of most messages, as are most sociogeographic networks. Indeed, the capacity of a channel, is defined by varying the probability distribution of the 'message' process $\mathbf{X}$ so as to maximize $I(\mathbf{X}|\mathbf{Y})$. For instance, suppose there is some message $\mathbf{X}$ so critical that its probability distribution must remain fixed. The trick is to fix the distribution $P(\mathbf{X})$ but to *modify the channel*; that is, to tune it so as to maximize $I(\mathbf{X}|\mathbf{Y})$. The *dual* channel capacity C^* is then defined as

$$C^* \equiv \max_{P(\mathbf{Y}), P(\mathbf{Y}|\mathbf{X})} I(\mathbf{X}|\mathbf{Y}). \tag{3.16}$$

But

$$C^* = \max_{P(\mathbf{Y}), P(\mathbf{Y}|\mathbf{X})} I(\mathbf{Y}|\mathbf{X}), \tag{3.17}$$

since we have

$$\begin{aligned} I(\mathbf{X}|\mathbf{Y}) &= H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{X}, \mathbf{Y}) \\ &= I(\mathbf{Y}|\mathbf{X}). \end{aligned} \tag{3.18}$$

Thus, in a purely formal mathematical sense, *the message transmits the channel*, and there will indeed be, according to the Shannon Coding Theorem, a channel distribution $P(\mathbf{Y})$ which maximizes C^*. Variations on this theme are realized in [134](see also [135]).

3.6 Noise in the system

In many sensory and cognitive systems not all noise corrupts the processing of information. Indeed, adding noise under the right circumstances may actually amplify and enhance the transmission, and may even reduce randomness in the system by the presence of *stochastic resonance*. In biological circumstances, the effect is often detected in large ion channels in the presence of stochastic processes, such as in the Hodgkin–Huxley model for instance [57]. Noise itself is not without its own peculiar 'linguistics' and semantic coding. Standard martingale analysis coupled with stochastic resonance reveals in [129] that the noise of a socio–economic structure (in the form, say, of misguided or regressive social policies), is most likely than not a major catalyst for the spreading of endemic illnesses, psycho–social disorders, therapeutic failure, inadequate public health services, and the deterioration of urban residential districts, as much as these factors influence each other [137, 138, 139].

3.7 A canonical model (M, s_M)

In some cases it will be necessary to project to a manifold model for the path space (X, s_X) based on this local description, but to one that is inherently less complicated and more conducive to standard topological/geometrical techniques. So let us proceed to define a topological space M which can be associated to X via a suitable map. Initially, we can grant M the structure of a metric space with a distortion measure s_M as induced by s_X in a sense to be made precise. We may assume that M admits an E–atlas manifold topology with a system of E–charts $\{(V, \varphi_V)\}$ while thinking of E as a suitable state space as above. Several possible ways to proceed are discussed below. To an extent (M, s_M) could be viewed as a more structured, simplified information space serving as a *canonical model* for the path space (X, s_X) (but as pointed out earlier, we do not insist on the speciality of Finsler or Riemannian spaces). Thus we consider a procedure similar to a dimensional reduction. We wish then to specify a projection map

$$\Pi : X \longrightarrow M, \tag{3.19}$$

with suitable properties, such as surjective, Lipschitz, etc., which will be outlined below. To an extent this will reflect the nature of information sources in (X, s_X), be they ergodic, or non–ergodic.

Remark 3.3. The use of the the term 'canonical' is similar in spirit to how intricate and chaotic systems of neural networks can be transformed into certain blueprints (canonical models) representing the dynamics of reduced systems of differential equations, and which can thus be studied with the standard techniques of differentiable dynamical systems theory (as in e.g. [72]).

3.8 Length space structures

Hypothesizing a class of 'admissible paths' in (M, s_M) leads to giving the latter a *length space* in the sense of [24]. In our case, the admissible paths are to be considered as 'meaningful' in the sense introduced below (and for which M is considered as a locally path connected space).

To see how a length structure induced by the metric s_M arises, let $\gamma : [a, b] \longrightarrow M$ be a (continuous) path in M and choose a partition $\mathcal{J}$ of the interval $[a, b]$, that is, a finite collection of points $\mathcal{J} = \{y_0, \ldots, y_N\}$ such that

$$a = y_0 \leq y_1 \leq y_2 \leq \cdots \leq y_N = b. \tag{3.20}$$

We can define the length of γ with respect to the metric s_M as the supremum of the sums over all partitions $\mathcal{J}$:

$$L(\gamma) = L_{s_M}(\gamma) := \sup_{\mathcal{J}} \sum_{i=1}^{N} s_M(\gamma(y_{i-1}), \gamma(y_i)). \tag{3.21}$$

The length structure induced by s_M can then be specified in terms of: a) all continuous paths parametrized by closed intervals are admissible, and b) the length is given by the function L in (3.21).

Consequently, we can draw upon generalizations of some traditional (but elementary) differential–geometric concepts in terms of a length structure. For instance, a curve

$$\gamma : [a, b] \longrightarrow M, \tag{3.22}$$

is said to be *rectifiable* if its length is finite, and a *shortest path* if its length is minimal among curves with the same endpoint, that is, $L(\gamma_1) \geq L(\gamma)$ for any curve γ_1 connecting $\gamma(a)$ and $\gamma(b)$. In particular, a curve $\gamma : I \longrightarrow M$ is said to be a *geodesic* if for every $t \in I$, there exists an interval J containing a neighborhood of t in I, such that $\gamma|J$ is a shortest path. We remark that the concept of 'geodesic' can also be formulated in the context of graphs and networks (to be discussed later).

The metric space (M, s_M) is said to be *complete* if there exists a shortest path between two languages A, A', and said to be (locally) *homogeneous* if for every A, A', there exists a (local) isometry $\mathfrak{I} : M \longrightarrow M$, such that $\mathfrak{I}(A) = A'$. Other possible length space structures could be considered thus allowing the flexibility of going beyond the traditional Finsler and Riemannian structures which are common frameworks for inference and stochastic processes.

Remark 3.4. Suppose in M we have an admissible class $\mathcal{A}$ of curves $\{\gamma(t)\}$. For $V \subset M$ an open set, suppose there is defined a nonnegative homogeneous function $\widetilde{F} = \widetilde{F}(x, v)$, where $x, v \in V$, that can be integrated over curves

$$\gamma = \gamma(t) : [a, b] \longrightarrow M, \tag{3.23}$$

in $\mathcal{A}$ invariant under reparametrization. The homogeneity condition implies the relationship $\widetilde{F}(x, kv) = |k|\widetilde{F}(x, v)$. For $\gamma \in \mathcal{A}$, the length between $A = \gamma(a)$ and $\widehat{A} = \gamma(b)$, may then be defined by

$$\ell_\gamma(A, \widehat{A}) = \ell(\gamma, a, b) = \int_a^b \widetilde{F}(\gamma(t), \gamma'(t))\, dt. \tag{3.24}$$

Observe that we have not required $\widetilde{F}(\ ,\ )$ to be a norm, or even that $\widetilde{F}$ is convex or symmetric, so the length space structure of M is weaker than that of a Finslerian structure (see e.g. [24]). A metric on M can be defined in the usual way by

$$s_M(A, \widehat{A}) = \inf\{\ell_\gamma(A, \widehat{A}) : \gamma(a) = A, \gamma(b) = \widehat{A}\}. \tag{3.25}$$

Given a local isometry $\Pi : X \longrightarrow M$, this metric may then be assumed to agree locally with s_X on pulling–back under Π .

Once M is endowed with an atlas–manifold topology we can then postulate M as a CW–complex, that is, a space constructed from a collection of points via the successive attachment of cells. The topology then is *weak*, meaning that a set in M is closed if and only if its intersection with every cell is closed. Viewed as a CW–complex, M then has the same homotopy type of a simplicial complex which affords further considerations particularly when reducing matters to a skeletal–like, graph–theoretic analysis.

Remark 3.5. The (path) space X may indeed be very complex in its structure of informational data, whereas the canonical model M (possibly via dimensional reduction of the former) by dint of its manifold structure, is expected to be more conducive to geometric analysis as may be the case for nonlinear optimization. Such principles apply in work concerning imaging–recognition data as in [121] where multi–imaging 'noisy' data in a carrier akin to X is analyzed (such as with regard to pixel intensity) and is projected to a local metric structure via maps such as Π. The convergence of subsequent 'data manifolds' in reduced dimensions may then yield the exact model for a solution space. In [121] examples include intricate 'Swiss Roll' data spaces which are dimensionally reduced to some convex region in $\mathbb{R}^n$. Here geodesic distances between data points are calculated, and typically one wants to minimize a cost function based on an operational norm of data differences. Such examples (re. pixel intensity) do not generally fit into the context of rate distortion theory and semantic communication, and for our purposes further considerations are clearly necessary. These we will proceed to discuss in the following.

3.9 Stationary ergodic information sources

For a rate distortion manifold, an alternative procedure is to consider stationary ergodic information sources, although not all cognitive processes are expected to be of this type. Here the Khinchin's E–property [79](p. 74) is evoked. Under the ergodicity assumption, the path space (X, s_X) can be partitioned into high and low probability subsets, $X = X_h \cup X_\ell$.

The projection map $\Pi : X \longrightarrow M$ can be specified as follows. Each equivalence class of paths in the appropriate space is identified with its associated language–of–thought characterized by a stationary ergodic information source having a source uncertainty $H(A)$, where A is taken as the language having a set of paths Ax . Thus for $x \in X$, we define the projection Π via $x \mapsto \Pi(x) = A$ (the language having a set of paths Ax). There are other possible variations on this theme.

Note also that we have restrictions

$$\begin{aligned} \Pi|X_h &: X_h \longrightarrow M \\ \Pi|X_\ell &: X_\ell \longrightarrow M. \end{aligned} \tag{3.26}$$

For each $A \in M$, let $U \subset M$ be an open set consisting of approximately similar languages near A. For all pairs of languages $A, \widehat{A} \in U$, let us suppose that we have available a suitable metric $s_M(A, \widehat{A})$ induced by the path space metric s_X. As an alternative to the metric in (3.25), we may choose a metric of the form [134]

$$s_M(A, \widehat{A}) \equiv \left| \int_{A,\widehat{A}} d(Ax, \widehat{A}x) - \int_{A,A} d(Ax, A\hat{x}) \right|, \tag{3.27}$$

where Ax and $\widehat{A}x$ are paths in the languages $A, \widehat{A}$ respectively, d is the distortion measure, and the second term is a 'self–distance' for the language A, such that $s_M(A, A) = 0,\ s_M(A, \widehat{A}) > 0, A \neq \widehat{A}$.

Since choosing stationary ergodic sources presents a different scenario to the non–ergodic sources, we may rethink the appropriate properties assumed by the projection Π. Possibilities might include: the projection

$$\Pi : (X, s_X) \longrightarrow (M, s_M), \tag{3.28}$$

is a local isometry, or, Π is *Lipschitz*, meaning there exists a constant $C > 0$, such that

$$s_M(\Pi(x), \Pi(\hat{x})) \leq C\, s_X(x, \hat{x}). \tag{3.29}$$

Next, for each $A \in U$, we consider a source uncertainty $H(A)$, such that *the information source derivative* $\nabla_s H(A)$ is defined to be

$$\nabla_s H(A) \equiv \lim_{s \longrightarrow 0} \frac{H(\widehat{A}) - H(A)}{s(\widehat{A}, A)}, \tag{3.30}$$

when this limit exists and is finite. A number of concepts follow from the basic principles of calculus, such as the *logarithmic derivative*

$$\nabla_s(\log H(A)) = \frac{\nabla_s H(A)}{H(A)}, \qquad (H(A) \neq 0) \tag{3.31}$$

a measure of the relative rate of change of the source uncertainty through language of thought.

Prior to admitting an atlas–manifold topology on M, we remark that the metric s_M defined in (3.25) appears to be *intrinsic* in the sense that it arises from a supposed length structure on M induced by the above integration over languages, and not by the restriction of a metric on some ambient space.

3.10 Non–ergodic information sources

Non–ergodic information sources are likely to be favorable options for the purpose of understanding more complex cognitive processes. Suppose here the path space X consists of length n high probability paths $x_n \longrightarrow x$ (as $n \longrightarrow \infty$) that correspond to *non–ergodic* information sources. Let $N(n)$ be the number of high probability 'grammatical' and 'syntactical' paths of length n having $h(a_0) \in B_0$, and leading to $h(x) \in B_1$. Such paths are called *meaningful* where once more the limit

$$H \equiv \lim_{n \longrightarrow \infty} \frac{\log\ N(n)}{n}, \tag{3.32}$$

exists, but is now generally taken to be path *dependent*.

We have a partial function $h : X \longrightarrow B$, where B denotes a set of pattern responses for which given $x_n \longrightarrow x$, we have

$$\lim_{n \longrightarrow \infty} h(x_n) = h(x). \tag{3.33}$$

For all $x \in X$, we take an open set $U \subset X$ such that for all such $x \in U$, the following conditions hold [125]:

(1) For all paths $\hat{x}_n \longrightarrow \hat{x} \in U$, a distortion measure $s_n \equiv d_U(x_n, \hat{x}_n)$ exists.

(2) For each path $x_n \longrightarrow x$ in U, there exists a pathwise invariant function $h(x_n) \longrightarrow h(x)$ [79].

(3) A function $F_U(s_n, n) \equiv f_n \longrightarrow f$ exists, such as for example

$$f_n = s_n, \frac{\log[s_n]}{n}, \text{ or } \frac{s_n}{n}. \tag{3.34}$$

(4) The limit

$$\lim_{n \longrightarrow \infty} \frac{h(x_n) - h(\hat{x}_n)}{f_n} \equiv \nabla_F h|_x, \tag{3.35}$$

exists and is finite.

In a similar way to stationary ergodic sources, we may consider introducing (M, s_M) as a more finely structured information space corresponding to a path space (X, s_X), using a dimensional reduction procedure via a projection map

$$\Pi : X \longrightarrow M, \tag{3.36}$$

with properties such as e.g. a surjective, local isometry, or Lipschitz, etc. that suitably projects the informational architecture on X to that on M.

Again, one might consider several options for the space M which would allow a weaker than Finsler structure. With regards to Remark 3.4, let us consider defined on each open set $V \subset M$, a nonnegative homogeneous function of two variables

$$\begin{aligned} \widetilde{F} &: V \times V \longrightarrow \mathbb{R}^+, \\ (u, v) &\mapsto \widetilde{F}(u, v). \end{aligned} \tag{3.37}$$

Let $A, \widehat{A}$ be points in $V \subset M$ corresponding respectively to paths of length n in X, denoted $x_n \longrightarrow x$ and $\hat{x}_n \longrightarrow \hat{x}$, under the map Π (so that we have $A = \Pi(x)$ and $\widehat{A} = \Pi(\hat{x})$).

For the first variable u, set $u = s_M$ where $s_M \equiv d_M(A, \widehat{A})$ denotes a choice of a suitable metric on M, and the second variable $v = v(n)$, so that $F(u, v) = F(s_M, v(n))$. Consider the pulled back function $\Pi^* \widetilde{F} = \widetilde{F} \circ \Pi$, with the property that on $U \subset M$, we have for some sufficiently small $\tilde{\epsilon} > 0$, the inequality

$$|\ F_U(s_n, n) - \Pi^* \widetilde{F}(s_M, v(n))\ | < \tilde{\epsilon}. \tag{3.38}$$

Of course, such considerations may be suitable for stationary ergodic sources as well. We leave open the possibility that a suitable metric space structure on M will be conducive to introducing methods from dynamical systems such as homoclinic points, hyperbolic sets, stable manifolds and related ideas (see e.g. [77]).

Remark 3.6. The constructions proposed above may be compared with that for an optimal manifold representation of (information) data in [28] for instance. Suppose $X \subset E$ is defined by a density function $\rho(x)$ and E is a vector space with norm $\|\cdot\|$. There is a stochastic map $\Pi : X \longrightarrow M$ which is seen as a projection to a lower dimensional manifold $M \subset E$. On M, a distortion measure $D(M, \Pi, \rho)$ is defined by

$$\int_{x \in E} \int_{m \in M} \rho(x)\ \Pi(x)\ \|x - m\|^2\ dx\ dm. \tag{3.39}$$

The map Π along with the density ρ determines a joint probability function $P(M, X)$ that allows calculation of the mutual information $I(X, M)$ between X (higher dimensional) and M its lower dimensional manifold representative M, as given by

$$\int_{x \in X} \int_{m \in M} P(x, m)\ \log\ [\frac{P(x, m)}{\rho(x)\ \Pi(x)}]\ dx\ dm. \tag{3.40}$$

3.11 Semimartingale processes and noise

Given the possible stochastic nature of bio–cognitive behavioral and response mechanisms, we expect the underlying processes to be driven to an extent by stochastic and noise–driven diffusion processes conducive to the creation of new information. This necessitates introducing a noise mechanism into the system as an agent towards self–organization and complexity (cf [9]), just as open systems far from equilibrium require some sort of internal 'amplification' in order to attain to a macroscopic dynamical structure (cf [108]).

We have already mentioned how martingale analysis is instrumental in describing stochastic resonance within cognitive and epidemiological systems. In order to build this feature into a rate distortion manifold we need to consider a more general approach. A *submartingale* on the real line $\mathbb{R}$ consists of those stochastic processes of the form $\Lambda + A$, where $\Lambda : C^0(\mathbb{R}) \times \mathbb{R}^+ \longrightarrow \mathbb{R}$ represents Brownian motion on $\mathbb{R}$ commencing at $0 \in \mathbb{R}$, possibly with a random time change, and A denotes a continuous increasing process on $\mathbb{R}$.

In order to translate these concepts so as to work on a rate distortion manifold, we will need a suitable measure such as that of (3.9). A suggested definition is provided in [45]: given some domain Ω, let us say that a stochastic process

$$\chi : \Omega \times \mathbb{R}^+ \longrightarrow (M, s_M), \tag{3.41}$$

is a *martingale on* (M, s_M), if for any convex function $g : U \longrightarrow \mathbb{R}$ defined on an open set $U \subset M$, the composition $g \circ \chi | U$ is a submartingale on $\mathbb{R}$. An example is Brownian motion

$$\Lambda^M : C_a^0(M) \times \mathbb{R}^+ \longrightarrow (M, s_M), \tag{3.42}$$

starting at $a \in M$.

A sum of a continuous local martingale and a process of finite variation gives rise to the notion of a *semimartingale* [49]. Such local martingales with respect to Brownian motion say, admit certain integral representations. If $\alpha \in \Omega^1(M, \mathbb{R})$ is a 1-form and ζ a semimartingale on M, then the real semimartingale $\int \langle d\alpha, \mathrm{d}\zeta \rangle$ (where $\langle\, , \,\rangle$ denotes the dual pairing, and $\mathbf{d}$ denotes derivation on tangent vectors) is called *the Stratonovich integral of* α *along* ζ. We refer to [49] §7 for further properties where this integral is denoted by $\int \langle \alpha, \delta\zeta \rangle$.

If M, N are (smooth) manifolds, one can define *the Stratonovich operator*

$$\mathfrak{E} = \{e(x, y)\}\ ,\ x \in M,\ y \in N, \tag{3.43}$$

to be a family of linear maps where the map $e(x, y) : T_xM \longrightarrow T_yN$ is a linear map that depends on (x, y) (to some degree of differentiability) thus defining a map $e : TM \times N \longrightarrow TN$. The latter also has a corresponding adjoint mapping $e^*(x, y) : T_y^*N \longrightarrow T_x^*M$. Given an M–valued semimartingale ζ, an N–valued semimartingale η is said to be a solution of the Stratonovich differential equation $\delta\eta = e(\zeta, \eta)\ \delta\zeta$, if for every 1–form α on N, there is the equality of Stratonovich integrals

$$\int \langle \alpha, \delta\eta \rangle = \int \langle e^*(\zeta, \eta)\alpha, \delta\zeta \rangle. \tag{3.44}$$

In terms of Markovian game theory the above concepts may be seen more concretely as follows. Let $\{\mathsf{X}_n\}$ be a sequence of stochastic variables defining a game (possessing noise), with conditional expectations given by

$$E(\mathsf{X}_{n+1} | \mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n) \equiv E(\mathsf{X}_{n+1}). \tag{3.45}$$

The definition of terms and interpretations are then:

(i) $E(\mathsf{X}_{n+1}|n) \geq \mathsf{X}_n$ –*submartingale* (favorable to player).

(ii) $E(\mathsf{X}_{n+1}|n) = \mathsf{X}_n$ –*martingale* (completely fair game).

(iii) $E(\mathsf{X}_{n+1}|n) \leq \mathsf{X}_n$ –*supermartingale* (favorable to the house).

As in [135](§5.1), this is exemplified for an epidemiological model whereby the 'player' is an infectious agent, X_n is the number of people infected at stage n (the player's fortune), and the 'house' is some socioeconomic system. A submartingale then represents a spreading infection, and a supermartingale represents a declining infection. The convolution of the community structure (the 'signal') with the opportunity structure (the 'noise') then leads to the simple epidemic model as a *generalized stochastic resonance*. Such situations have been studied by similar means in the case of childhood illnesses where, besides the internal transmission of an infection within a community, there is also an external effect due to individual migration between communities, and weak 'seasonality' together with low transmission levels are seen to induce stochastic effects with amplified noise that may generate resonance [3]. Our perspective is that models such as these could be treated in terms of *groupoids* and convoluted path space, as will be discussed in §7.1.

4 Equivalence relations and tuning

4.1 Equivalence relations

For the purpose of describing cognitive modules we can also append the structure(s) with equivalence relations $\mathcal{R}^X, \mathcal{R}^M, \mathcal{R}^E$ defined on X, M, E respectively, and a sequence of maps

$$(X, \mathcal{R}^X) \xrightarrow{\Pi} (M, \mathcal{R}^M) \xrightarrow{\varphi} (E, \mathcal{R}^E). \tag{4.1}$$

Suggestive of the orbit equivalence theorem (relative to a more abstract setting of e.g. [59, 100]), we will suppose the equivalence relations are tied by

$$\begin{aligned} \mathcal{R}^M &= \Pi \times \Pi(\mathcal{R}^X), \\ \mathcal{R}^E &= \varphi \times \varphi(\mathcal{R}^M). \end{aligned} \tag{4.2}$$

4.2 Tangent spaces

As we have suggested, a rate distortion manifold need not necessarily be a differentiable manifold in the conventional sense, but may admit an abstract differentiable space structure (such as that described in Appendix III). In some instances, however, we may have to address the question of tangency at a point $m \in M$ and thus assume that the tangent space T_mM is defined accordingly. We recall how T_mM can be defined in terms of an equivalence classes of curves. Consider the equivalence relation $c_1(R_m^M)c_2$ on curves c_1, c_2 as meaning: c_1 and c_2 are tangent at $m \in M$, if and only if $(\varphi \circ c_1)(R^E_{\varphi(m)})(\varphi \circ c_2)$ means they are tangent at the point $\varphi(m)$ in E. In which case, the equivalence class $[c]_m$ at $m \in M$, is defined to be the tangent space at m, and this is usually denoted by T_mM. This way of viewing tangency in terms of equivalence classes globalizes to the construction of the tangent bundle $TM \longrightarrow M$ as described in e.g. [1, 87].

From an information–theoretic point of view, this description of tangency is useful for characterizing 'tuning'. We give a more general interpretation. Suppose p_1, p_2 are paths (or 'sequences') in X that are projected by Π down to the manifold M; we keep p_1, p_2 to denote their projected images. For $m \in M$, let (U, φ) be a local chart with $m \in U$ and $\varphi(U) \subset E$. With regards to the equivalence relations $\mathcal{R}^M, \mathcal{R}^E$, we want to consider paths (sequences) p_1, p_2 as being *equivalently tuned* at $m \in M$, denoted $p_1(R^M_m)p_2$, if at $\varphi(m)$ in the atlas space E, the equivalence denoted $(\varphi \circ p_1)(R^E_{\varphi(m)})(\varphi \circ p_2)$, holds. The equivalence class of paths $[p]_m$ at $m \in M$ may then be thought of as *the tuning space at the point m in the manifold M*, in a sense corresponding to a focal point of attentive processing.

4.3 Higher and lower dimensional information sources

Let $\mathbf{X}$ and $\mathbf{Y}$ be information sources whereby $\dim\ \mathbf{X} \geq \dim\ \mathbf{Y}$. The 'higher' source $\mathbf{X}$ is one that may be considered 'fast' and the 'lower' source $\mathbf{Y}$ is considered 'slow'. Associated to $\mathbf{X}$ and $\mathbf{Y}$, are their respective path spaces with respective distortion metrics (X, s_X), and (Y, s_Y). We consider a projection map

$$\Phi : (X, s_X) \longrightarrow (Y, s_Y), \tag{4.3}$$

as complying with the following version of the Rate Distortion Theorem: *for any chosen maximum average distortion such that*

$$d(x, \Phi(x)) < \epsilon, \tag{4.4}$$

there is, in relationship to Φ, a maximum possible transmission rate δ, such that the average distortion will be less than ϵ [125].

The above condition can be represented in terms of a subset of the graph of Φ. Recalling that the graph Γ_Φ of Φ is given by

$$\Gamma_\Phi := \{(x, y) \in X \times Y : y = \Phi(x)\}, \tag{4.5}$$

we define a 'rate distortion' subset $\Gamma_{\mathrm{rd}} \subseteq \Gamma_\Phi$ by

$$\Gamma_{\mathrm{rd}} := \{\ (x, \Phi(x)) \in \Gamma_\Phi : d(x, \Phi(x)) < \epsilon\ \}. \tag{4.6}$$

Recall that each of (X, s_X) and (Y, s_Y) have associated lower dimensional and structured language of thought spaces M and N (with possible manifold/atlas topologies) and induced distortion metrics s_M and s_N, respectively. For each of these we have projections onto the lower dimensional source, Π_1 and Π_2 respectively (with whatever properties are assumed), and for which the diagram below commutes

$$\begin{array}{ccc} (X, s_X) & \xrightarrow{\Pi_1} & (M, s_M) \\ \Phi \big\downarrow & & \big\downarrow \Psi \\ (Y, s_Y) & \xrightarrow{\Pi_2} & (N, s_N) \end{array} \tag{4.7}$$

that is, $\Psi \circ \Pi_1 = \Pi_2 \circ \Phi$. In this way we see that induced on Ψ is the constraint of the Rate Distortion Theorem as it holds on Φ, such that the corresponding languages of thought adhere accordingly.

4.4 Tunable states

The genesis of a rate distortion manifold lies in the concept of a *generalized tunable retina model* as introduced in [125]. Specifically, let us suppose that threshold behavior for individual, distributed or institutional (cognitive) reaction requires some elaborate system of nonlinear relationships defining a set of renormalization parameters

$$\Omega_k \equiv \omega_1^k, ..., \omega_m^k. \tag{4.8}$$

The critical assumption is that there is a tunable zero order state, and any changes about that state are, in first order, relatively small, although their effects on a punctuated process may not be at all small. Thus, given an initial m-dimensional vector Ω_k, the parameter vector at time $k+1$, Ω_{k+1}, can, in first order, be written as

$$\Omega_{k+1} \approx \mathbf{R}_{k+1}\Omega_k, \tag{4.9}$$

where $\mathbf{R}_{k+1}$ is an $m \times m$ matrix, having m^2 components. If the initial parameter vector at time $k = 0$ is Ω_0, then at time k

$$\Omega_k = \mathbf{R}_k\mathbf{R}_{k-1} \ldots \mathbf{R}_1\Omega_0. \tag{4.10}$$

The interesting correlates of individual, institutional or machine consciousness are, in this development, *now represented by an information-theoretic path defined by the sequence of operators $\mathbf{R}_k$*, each member having m^2 components, for some m. The grammar/syntax of the path defined by these operators is associated with a dual information source, in the usual manner.

The effect of an information source of external signals $\mathbf{Y}$, is now seen in terms of more complex joint paths in Y and the $\mathbf{R}$-space (of operators) whose behavior is, again, governed by a mutual information splitting criterion according to the Joint Asymptotic Equipartition Theorem (a variant of the Shannon–McMillan Theorem). The complex sequence in m^2–dimensional $\mathbf{R}$–space has, by this construction, been projected down onto a parallel path, the smaller set of m-dimensional ω–parameter vectors $\Omega_0, \ldots, \Omega_k$.

If the punctuated tuning of institutional or machine attention is now characterized by a 'higher' dual information source – an embedding generalized language – so that the paths of the operators $\mathbf{R}_k$ are autocorrelated, then the autocorrelated paths in Ω_k represent output of a parallel information source which is, given rate distortion limitations, apparently a grossly simplified, and hence highly distorted, picture of the 'higher' conscious process represented by the $\mathbf{R}$-operators, having m as opposed to $m \times m$ components. High levels of distortion may not necessarily be the case for such a structure, *provided it is properly tuned to the incoming signal*. If it is inappropriately tuned, however, then distortion may be extraordinary.

Let us examine a single iteration in more detail, assuming now there is a (tunable) zero reference state, $\mathbf{R}_0$, for the sequence of operators $\mathbf{R}_k$, and that

$$\Omega_{k+1} = (\mathbf{R}_0 + \delta\mathbf{R}_{k+1})\Omega_k, \tag{4.11}$$

where $\delta\mathbf{R}_k$ is 'small' in some sense compared to $\mathbf{R}_0$. Note that in this analysis the operators $\mathbf{R}_k$ are implicitly, determined by linear regression. We thus can invoke a quasi-diagonalization in terms of $\mathbf{R}_0$. Let $\mathbf{Q}$ be the matrix of eigenvectors which Jordan–block–diagonalizes $\mathbf{R}_0$. Then

$$\mathbf{Q}\Omega_{k+1} = (\mathbf{Q}\mathbf{R}_0\mathbf{Q}^{-1} + \mathbf{Q}\delta\mathbf{R}_{k+1}\mathbf{Q}^{-1})\mathbf{Q}\Omega_k. \quad (4.12)$$

If $\mathbf{Q}\Omega_k$ is an eigenvector of $\mathbf{R}_0$, say Y_j with eigenvalue λ_j, it is possible to rewrite this equation as a generalized spectral expansion

$$\begin{aligned} Y_{k+1} &= (\mathbf{J} + \delta\mathbf{J}_{k+1})Y_j \equiv \lambda_j Y_j + \delta Y_{k+1} \\ &= \lambda_j Y_j + \sum_{i=1}^{n} a_i Y_i, \end{aligned} \quad (4.13)$$

where $\mathbf{J}$ is a block-diagonal matrix

$$\delta\mathbf{J}_{k+1} \equiv \mathbf{Q}\mathbf{R}_{k+1}\mathbf{Q}^{-1}, \quad (4.14)$$

and δY_{k+1} *has been expanded in terms of a spectrum of the eigenvectors of* $\mathbf{R}_0$, with

$$|a_i| \ll |\lambda_j|\,,\ |a_{i+1}| \ll |a_i|. \quad (4.15)$$

The point is that, provided $\mathbf{R}_0$ has been tuned so that this condition is true, the first few terms in the spectrum of this iteration of the eigenstate will contain most of the essential information about $\delta\mathbf{R}_{k+1}$. This appears quite similar to the detection of color in the retina, where three overlapping non–orthogonal eigenmodes of response are sufficient to characterize a huge plethora of color sensation. Here, if such a tuned spectral expansion is possible, a very small number of observed eigenmodes would suffice to permit identification of a vast range of changes, so that the rate distortion constraints become quite modest. That is, there will not be much distortion in the reduction from paths in $\mathbf{R}$–space to paths in Ω–space. Inappropriate tuning, however, can produce very marked distortion as in inattentional blindness (individual, institutional or machine–oriented), in spite of multitasking. We remark that higher order rate distortion manifolds are likely to give better approximations than lower ones, in the same sense that second order tangent structures give better, if more complicated, approximations in conventional differentiable manifolds. The formal mathematical background to this idea can be found in [102]; we will be more specific about this observation in §6.4.

Remark 4.1. A possible and more general geometric way of viewing such constructions is to consider a vector bundle $V \longrightarrow M$ and some operator $\mathbf{R} : \Gamma(V) \longrightarrow \Gamma(V)$ on sections of V. The above local description may serve to describe the (time) evolution of $\mathbf{R}$ and its spectral properties. In certain cases this reveals an associated 'spectral' set (or submanifold, which could be M itself). Alternatively, the tangent space TM (pointwise) may split into a particular 'eigenmode' decomposition, a feature often found in the field of differentiable dynamics (see e.g. [1, 77]).

4.5 Description of the rate distortion manifold as an information space

At this stage we can summarize some of the essential properties required for *a rate distortion manifold* (M, s_M) to serve as a cognitive information space within the Global Workspace setting.

(1) M with its distortion metric s_M and atlas–manifold topology serves as a canonical model of the path space (X, s_X), where $X = \mathcal{P}(E^\Gamma)$. The corresponding distortion measure leads to defining a rate distortion function on M complying with the Rate Distortion Theorem.

(2) In terms of the metric s_M, M admits a length space structure in the general sense as described (*a priori*, weaker than a Finsler structure say).

(3) M may possibly admit a 'weak' differentiable space structure in some suitable sense (for instance, in terms of the abstract calculus of manifolds such as [27, 55] described briefly in Appendix III).

(4) Both ergodic and non–ergodic processes may be considered. We may also require that M admits certain stochastic properties suited to representing e.g. stochastic resonance in the informational context. On the other hand, there may be situations where M carries a flow engendered by solutions to some wave equation or the time–evolution of an operator on sections of a vector bundle over M. Spectral eigenmode decompositions of TM may be expected as §4.4 suggests.

(5) For the purposes of a 'directed' theory of information, M may admit a partial ordering '$\leqslant$' and thus may admit an underlying 'partially ordered space' structure (see §6.6).

5 Thermodynamic limit and the Onsager relations

Feynman [53] (following in part the work of C. Bennett) considered the problem of extracting useful work from a transmitted message. The essential argument is that computing of any form, requires work. Consequently, on recalling from (3.3), the asymptotic limit

$$H \equiv \lim_{n \longrightarrow \infty} \frac{\log[N(n)]}{n}, \quad (5.1)$$

is postulated as formally homologous to the thermodynamic limit in the definition of the free energy density of a physical system as given by

$$F(K) = \lim_{V \longrightarrow \infty} \frac{\log[Z(K)]}{V}, \quad (5.2)$$

where F is the free energy density, K the inverse temperature, V the system volume and $Z(K)$ the partition function defined by the Hamiltonian of the system.

In [125] it is shown at some length how this homology permits the natural transfer of renormalization methods from statistical mechanics to information theory, producing phase transitions and analogs to evolutionary punctuation in systems characterized by piecewise, adiabatically stationary, ergodic information sources. Crosstalk, as a particular characteristic, may then serve as an 'inverse temperature parameter'.

This homology is essential for understanding the type of model spaces as described here. The point being, that the more intricate a cognitive process, measured by information source uncertainty, the greater its energy consumption. Biological phase changes appear to be ubiquitous in natural systems and can be expected to dominate information machine behaviors as well, particularly those which seek to emulate biological paradigms. In [133] these arguments are used to explore the differences and similarities between evolutionary punctuation in genetic and learning plateaus in neuronal systems.

As much as thermodynamic laws influence most kinds of vital phenomena, certain types of epidemiological and cognitive processes may be represented in terms of a thermodynamic limit on the processing capacity (as for instance in the case of inattentional blindness [128] or sleep patterns [127]).

In order to see how suitable models may be designed accordingly, consider the dual source uncertainty of a cognitive process as parametrized by a vector $\mathbf{K} \equiv (K_1, \ldots, K_n)$. In analogy with nonequilibrium thermodynamics we define the *disorder* as a function S given by

$$S \equiv H(K) - \sum_{j=1}^{m} K_j \frac{\partial H}{\partial K_j}. \tag{5.3}$$

Further expanding the homology, leads to defining the generalized Onsager relations of temporal dynamics

$$\frac{dK_j}{dt} = \sum_i L_{ji} \frac{\partial S}{\partial K_i}, \tag{5.4}$$

where $L = [L_{ij}]$ is a matrix of constants associated to the underlying cognitive phenomena. From the symmetric matrix

$$U = [U_{ij}] = [\frac{\partial^2 S}{\partial K_i \, \partial K_j}] = [U_{ji}], \tag{5.5}$$

one can define associated metric coefficients as follows:

$$g_{ij} = \frac{L^2}{2} \sum_k U_{ik} U_{kj}. \tag{5.6}$$

Next, consider the source uncertainty

$$\mathbf{K}(t) \equiv (K_1(t), \ldots, K_n(t)), \tag{5.7}$$

as time dependent and defining a (smooth) curve $\mathbf{K} : \mathbb{R}^+ \longrightarrow M \subset \mathbb{R}^n$, in a rate distortion manifold M. Use of standard procedures (see e.g. [1, 24]) leads to defining a suitable length space structure on M via a distance function s_M between languages $A, \widehat{A}$, suitably represented by points along some dynamic path in M. Here s_M is given by

$$s_M(A, \widehat{A}) = \int_A^{\widehat{A}} [\sum_{i,j} g_{ij} \frac{dK_i}{dt} \frac{dK_j}{dt}]^{\frac{1}{2}} \, dt. \tag{5.8}$$

Accordingly, the curve $\mathbf{K}(t)$ with respect to the above metric structure is a geodesic in M precisely when the second order equation

$$\frac{d^2 K_i}{dt^2} + \sum_{j,m} \Gamma^i_{jm} \frac{dK_j}{dt} \frac{dK_m}{dt} = 0, \tag{5.9}$$

is satisfied, where the Γ^i_{jm} denote the associated Christoffel symbols (see e.g [1, 24]).

Remark 5.1. One may hypothesize that under the right circumstances, geodesics sufficiently near to a reference state A_0 are bound by some estimate, and external physiological forcing must be imposed to effect a transition to a different condition. This, as is pointed out in [128], may be specified in terms of regions of fatal attraction and strong repulsion akin to Black or White hole phenomena which can either trap or deflect the path of consciousness.

5.1 The torus–sphere example: differing homologies on languages of cognition

Most textbooks on algebraic topology contain the relevant definitions and concepts pertaining to singular homology of the space M in terms of the constituent homology groups $H_k(M, \mathbb{Z})$ with integer coefficients (see e.g. [20, 87, 117]):

$$H_*(M) = \sum_{k=0}^{\dim M} H_k(M, \mathbb{Z}). \tag{5.10}$$

Loosely speaking, the $H_k(M, \mathbb{Z})$ are 'groups of cycles of differing dimensions' which contribute to an overall characteristic of the space, namely its *homological structure*. In particular, if $M = U \cup V$ is the union of two open sets, then a finer analysis can be made in terms of the Mayer–Vietoris homology sequence

$$\begin{aligned} \ldots \longrightarrow H_q(U \cap V) \xrightarrow{f} H_q(U) \oplus H_q(V) \xrightarrow{g} H_q(M) \\ \longrightarrow H_{q-1}(U \cap V) \longrightarrow \cdots \end{aligned} \tag{5.11}$$

where f denotes the map induced by a signed inclusion $a \mapsto (-a, a)$ and g is that of the sum $(a, b) \mapsto a + b$.

Rate distortion manifolds pertaining to distinct homotopy types are then expected to represent distinct homologies in relationship to information, intrinsic languages of cognition and culture, and to which such homological techniques can be applied. A topological example, as presented in [125] (Chapter 5) concerns the case of the two–torus T^2

versus the two–sphere S^2, both of which are surfaces embedded in 3–space, and both possess two–dimensional tangent spaces (pointwise). Thus sitting in a small local coordinate patch, an observer cannot really notice much difference. However, their homotopy type and (singular) homologies are fundamentally different, since at level $k = 1$ we have $H_1(S^2) = 0$, whereas $H_1(T^2) = \mathbb{Z} \oplus \mathbb{Z}$.

In general, *the manifold itself forms an envelope of the entirety of its tangent planes and this envelope will in turn describe a homology type which reflects a particular topological structure*. This straightforward topological observation underscores the point of how rate distortion manifolds of distinct (fundamental) homology, homotopy and diffeomorphism types, provide an informational blueprint for the significant differences in bio–cultural/psycho–cognitive choices in Workspace informational processing. Such fundamental differences might be realized in a variety of cognitive situations where culture, as always, plays a significant role. Consider for instance the case in [99] which compares the main differences between the Asian mode of perception, on the one hand, widely framed and holistic, and the Western mode, on the other hand, more analytically and logically centered. In other words, culture bears an influence upon the 'topologies' of the respective information processing which partly explains the difference between the two perceptual characteristics.

6 Embeddings of contexts

This section takes up some of the development in the previous sections with a tentative outline of how we might formulate a finer geometric description of the Global Workspace in terms of a rate distortion manifold M:

(1) The Workspace and access to it may be considered in terms of nested sequence of rate distortion manifolds

$$M_1 \subset M_2 \subset \cdots \subset M_n \subseteq M.$$

This represents a hierarchy of cognitive processes based on nonlinear dynamical principles.

(2) Cooperating and competing contexts as unconsciousness networks to be integrated within the Workspace, participate within a higher dimensional dominant context describable by embeddings as in (1).

(3) Within each context there is a cooperating group of specialized processors where access to the Workspace can be represented through such a chain of inclusions/embeddings.

Let us start by taking an E–chart (U, ϕ) for a rate distortion manifold M relative to the model space E. Here (U, ϕ) is taken to be a chart representative of a 'context'. We consider (U, ϕ) to be sufficiently 'large' so as to admit a sequence of 'embeddings' of charts through the above sequence. Initially we have a hierarchial context given as before by a projection $\Pi : X \longrightarrow M$ from some high dimensional information source space X (typically, a 'culture' with rate–distortion features) to a lower dimensional information carrier, a rate distortion manifold M (typically, a 'structured' or 'canonical' cognitive system which could be a canonical model for X).

6.1 Cooperating contexts

Cooperating contexts (cf specialized unconscious processors) within a hierarchial structure are represented as a nested sequence of rate distortion manifolds $\{M_k\}$ given by

$$\begin{aligned} &(M_1; (V_1, \psi_1)) \subset (M_2; (V_2, \psi_2)) \cdots \\ &\quad \subset (M_m; (V_m, \psi_m)) \subseteq (M; (U, \phi)), \end{aligned} \tag{6.1}$$

each with their respective chart/atlas system (V_k, ψ_k), and where each M_k represents a processing stage within some cognitive (sub)system corresponding to a level of information k. These come complete with inclusions through their respective chart/atlas systems

$$V_1 \xrightarrow{\lambda_1} V_2 \xrightarrow{\lambda_2} \cdots \xrightarrow{\lambda_{m-1}} V_m \xrightarrow{\lambda_m} U, \tag{6.2}$$

satisfying, for $1 \leq i \leq m$, the composition

$$\psi_i = \phi \circ \lambda_m \circ \lambda_{m-1} \cdots \circ \lambda_i. \tag{6.3}$$

In this sequence, we may include at each level k retraction mappings (or projections when defined) $p_k : M \longrightarrow M_k$, for $k \leq m$, such that we have a commuting diagram

$$\begin{array}{ccc} X & \xrightarrow{\Pi} & M \\ & \searrow & \downarrow{\scriptstyle p_k} \\ & & M_k \end{array} \tag{6.4}$$

suggesting the influence of the higher (or faster) dimensional information source at subsystem level k.

Thus the nested sequence of embeddings could be interpreted as levels of information representing a broad dominant context hierarchy through executive functions and levels of cooperation such as (e.g. in [12] §4):

Perceptual Contexts $\Longrightarrow$ Conceptual Contexts $\Longrightarrow$ Goal Contexts

specified at each level k by M_k. The above sequence can also be appended with a string of (sub)–equivalence relations "$\sim_i$", when specified:

$$(M_1, \sim_1) \subset (M_2, \sim_2) \cdots \subset (M_m, \sim_m) \subseteq (M, \sim). \tag{6.5}$$

The sequence is thus seen to represent a parallel series of specialized processors (cf e.g. [12, 38]). Also, the adoption of the equivalence relations affords a useful interpretation (cf the notion of 'frames' in the cognitive sense). Each $(M_i, \sim_i)$ then leads in a straightforward way to a *groupoid* structure upon which we will elaborate below.

6.2 Filtration by Morse functions

The proof that an n–dimensional manifold M has the homotopy type of a CW–complex of dimension $\leqslant n$, relies on the use of Morse functions $f : M \longrightarrow \mathbb{R}$ which are ordered in a suitable sense (see e.g. [95, 113]). Specifically, we choose a sequence of numbers $c_1, c_2, \ldots, c_{n-1}$ for which the following is satisfied: if a_i^λ (resp. $b_j^{\lambda+1}$) are critical points of index λ (resp. $\lambda + 1$), we have $f(a_i^\lambda) < c_\lambda < f(b_j^{\lambda+1})$. Then the manifolds $M_\lambda = f^{-1}[0, c_\lambda]$ define a filtration

$$M_* : M_0 \subseteq M_1 \subseteq \ldots M_\lambda \subseteq \ldots \subseteq M_n = M. \quad (6.6)$$

6.3 Competing contexts

Next we consider another such nested sequence $\{N_\ell\}$ of rate distortion manifolds with their respective chart/atlas system (W_j, ζ_j) contained within that of $(M; (U, \phi))$:

$$\begin{aligned}&(N_1; (W_1, \zeta_1)) \subset (N_2; (W_2, \zeta_2)) \\ &\cdots \subset (N_n; (W_n, \zeta_n)) \subseteq (M; (U, \phi)),\end{aligned} \quad (6.7)$$

also complete with inclusions through their respective chart/atlas systems

$$W_1 \overset{\rho_1}{\longrightarrow} W_2 \overset{\rho_2}{\longrightarrow} \cdots \overset{\rho_{n-1}}{\longrightarrow} W_n \overset{\rho_n}{\longrightarrow} U \quad (6.8)$$

satisfying, for $1 \leq j \leq n$, the composition

$$\zeta_j = \phi \circ \rho_n \circ \rho_{n-1} \cdots \circ \rho_j. \quad (6.9)$$

Likewise, we have a projection of influence as in (6.4), and each (N_j, ζ_j) can be appended with an equivalence relation "$\sim_j$". The sequence then represents the hierarchy

Perceptual Contexts $\Longrightarrow$ Conceptual Contexts $\Longrightarrow$ Goal Contexts

The main difference between this sequence $\{N_j\}$ and the $\{M_i\}$, is that *at a sufficiently low level in the hierarchy*, the former represents a *competing sequence* to the latter within a dominating system $(M; (U, \phi))$ subject however to the projection $\Pi : X \longrightarrow M$. In a Workspace setting, the sequences $\{N_j\}, \{M_i\}$ each with limited capacity, contend for recognition in a central or main processor (as represented by the larger canonical model $(M; (U, \phi))$ with its rate distortion characteristics).

Some suggested conditions are in order (but most likely not exhaustive). These include the possibilities that over some range of indices $1 \leq \ell < \min\{i, j\}$, we have:

(i) $\dim M_\ell \geq \dim N_\ell$ [same level competition].

(ii) $M_\ell \cap N_\ell \neq \emptyset$ [same level competition].

(iii) For charts $\psi_\ell | V_\ell \cap W_\ell \neq \zeta_\ell | V_\ell \cap W_\ell$, when (ii) above occurs. The charts ψ_ℓ and ζ_ℓ are distinct functions on $V_\ell \cap W_\ell$ [competition].

In other words, the contexts near to, or at the bottom of, the hierarchy compete to dominate the Workspace by means of their intrinsic cognitive mechanisms as represented by the geometric structure of the $\{M_i\}$ and $\{N_i\}$ via their respective chart/atlas system. There are other possible interpretations. For instance, as in [125], stress can be viewed as a socially devised cultural characteristic involving a schemata of languages each with its own grammar/syntax. This may be represented in mathematical terms by the high dimensional information source X. The dimensional reduction (given by projection $\Pi : X \longrightarrow M$) along with the embeddings of the chain of the M_k into M, can then be viewed as the relevant interacting cognitive modules within some environment: for instance, how embedded psycho–social stress influences mind–body interactions. For 'noise' related purposes, we may also consider under appropriate conditions, the various embeddings $M_i \longrightarrow M_j$ as linked in a semimartingale process by Stratonovich operators $\mathfrak{E}_{ij} : TM_i \times M_j \longrightarrow TM_j$ as in (3.43), and likewise for the chain $N_i \longrightarrow N_j$.

These cognitive modules, as reflective of the mathematical description of the model, so adhere to the asymptotic limit theorems of information theory as we have outlined them. For many persons, information overload, 'noise' and 'heat engines', are the fiendish perpetrators of such stress ailments as the modules so describe. The all–or–none competing stimuli creating a bottleneck in the central processing of neural information is considered in [38, 128] as a Workspace explanation of inattentional blindness for which the above schemata of rate distortion manifolds may serve as a blueprint.

6.4 Higher order tangency

The above setting holds further prospects for applying state-of-the-art techniques from geometry and topology. In particular, we have mentioned that higher order rate distortion manifolds are likely to produce better approximations than those of lesser order. A formal explanation of these terms is as follows. If M is a smooth manifold in the traditional sense, then a classical example is how tangent vectors coalesce with osculating curves such as the local geodesics. More generally, one may consider the *higher order tangency* of submanifolds of M where the maps in question admit *osculating spaces* to certain orders. Following [102], the idea revolves around *p-th order tangent bundles* $TM^{[p]}$ whose typical fiber consists of a p-th order *osculating vector*. The latter can be related to the classical osculating spaces of order p of a submanifold of some affine space. These higher order tangent bundles comply with the exact sequence

$$0 \to TM^{[p-1]} \longrightarrow TM^{[p]} \longrightarrow S^p(TM^{[1]}) \to 0, \quad (6.10)$$

where S^p denote the p-th symmetric tensor product. Such osculating spaces seem relevant to 'higher order retinal tuning' in the context of institutional multitasking [131], a topic for further development.

6.5 Hierarchy of organization and complexity

Complex systems (whether they are genetic, neuronal or cultural–social) can be organized into various levels of complexity each equipped with their sovereign mechanisms for managing the prevailing environment or community. Recall also how neurons communicate via synapses by means of a synchronous assembly of encoded neurons, thus realizing a mental event with some degree of plasticity. At a higher order there exists a semantic memory which influenced by referential experiences within its environment, induces the development of personal memory as characteristic of humans and other higher order species. Quite often the question is to determine how the higher levels evolve from those lower without being directly reducible to them, in a way similar to how percolation techniques of lower to higher level operations are realized in the theory of graphs and networks. A similar problem has been discussed in [46] in the context of *memory evolutive systems* (MES) which depend on types and classes of selection procedures. An internal feature is that iterated complexification can induce a hierarchy of objects of strictly increasing orders of complexity each with its own characteristics which allow a switching across the constituent organisms. This can be explained in categorical terms, in particular, in terms of 'colimits' as we will briefly describe.

Firstly, let us give a short but quite abstract notion of a *graph* as a family of objects $\{A_i\}$ together with a collections of arrows $f : A_1 \longrightarrow A_2$ between objects. In the absence of the strict definition of a *category* (see e.g. [19, 91]), let us say for now, and informally, that a standard notion of a categorical structure can be defined on a graph in terms of objects and an internal rule of composition '$\circ$' associating to pairs $f : A_1 \longrightarrow A_2,\ g : A_2 \longrightarrow A_3$, the composition $g \circ f : A_1 \longrightarrow A_3$, satisfying the rule of associativity, together with the identity morphism $\mathrm{id}_A : A \longrightarrow A$.

If we regard objects as labeled in terms of ordered states $A < A'$, a transition functor $F(A, A') : F_A \longrightarrow F_{A'}$, represents a change in states $A \longrightarrow A'$, and satisfies

$$F(A, A'') = F(A, A') \circ F(A', A''). \qquad (6.11)$$

Following e.g. [46], if we have a system as represented by a graph, it is said to be *hierarchial* if the objects can be divided into specified complexity levels representative of the embeddings of contexts.

Further, we can speak of a pattern of linked objects A as a family of objects A_i with specified links (edges) between them. Consider another object B to which we can associate a collective link from A to B by a family of links $f_i : A_i \longrightarrow B$. We can picture then a cone with a base consisting of $A = \{A_1, A_2, \ldots\}$ and with B as the vertex. The pattern is said to admit a *colimit* denoted C, if there exists a collective link $A \longrightarrow C$ such that any other collective link $A \longrightarrow B$ admits a unique factorization through C. If such a colimit C exists, then locally C is well–defined by the nature of the pattern to which it is attached, and globally, C enjoys a universal property determined by the totality of the possible collective links of the pattern. In other words, C effectively binds the pattern objects while at the same time functions as the entire pattern in the sense that the collective links to B are in a one–to-one correspondence with those to C. Further, a category can be said to be *hierarchial* if its objects can be partitioned into different levels of complexity, with an object C of level $n + 1$ say, being the colimit of at least one pattern of linked objects of (strictly) lower levels $n, n - 1, \ldots$

In our situation the concept is particularly useful when the objects A_i comprise a pattern or network of rate distortion manifolds and the collective links $f_i : A_i \longrightarrow B$ are morphisms to B which may, for instance, model a central processor. The colimit C then functions as the binding agent for the respective channels of information.

Remark 6.1. The concept of a colimit in a category generalizes that of forming the union $A \cup B$ of two overlapping sets, with intersection $A \cap B$. However, rather than concentrating on the actual sets A, B, we place them in context with the role of the union as permitting the construction of functions $f : A \cup B \to C$, for any C, by specifying functions $f_A : A \longrightarrow C$, $f_B : B \longrightarrow C$ agreeing on $A \cap B$. Thus the union $A \cup B$ is replaced by a property which describes, in terms of functions, the relationship of this construction to all other sets. In practical terms [15, 22] it is how we might compare input and output. In this respect, a colimit has 'input data', viz a *cocone*. For the union $A \cup B$, the cocone consists of the two functions $i_A : A \cap B \to A$ and $i_B : A \cap B \to B$.

An *evolutive system* [46] is viewed as a family of categories indexed by a suitable parameter t (usually time), together with a family of transition functors. The internal organization of a complex component C can then be modeled in relationship to a pattern of linked objects such that the actions of C on any other component are determined completely by the collective links (of the pattern), thus characterizing C as the colimit. The above model can describe an evolutionary autonomous system (or organism) with a hierarchy of components managing organized exchanges within an environment. Thus a *hierarchial evolutive system* is then an evolutive system in which the state category at each value t is hierarchial and the transition functors preserve the levels. By means of such a network of learning, the system re–adapts to changing conditions within the environment, thus leading to a MES, a characteristic which can be related to the embedding of contexts. We re–iterate that these components could be realized as specific types of rate distortion manifolds which collectively model information relay within an interactive context.

6.6 Scheduling of paths

Methods of concurrency involving directed homotopies, scheduling, n–categories/n–complexes and related topics may well be suited to developing certain aspects of cognitive/institutional multi–tasking. The geometric perspective is outlined in [62, 105] and an application to the Global Workspace is discussed in [132].

A particular idea starts by recalling the notion of a *partially ordered space* (a *po–space* for short) with respect to a partial ordering "$\leqslant$" on M. A *local po–space* is a Hausdorff space M with a covering $\mathcal{U} = \{U_\alpha, \leqslant_\alpha, \alpha \in J\}$ where each $U_\alpha \subseteq M$ is open and $\leqslant_\alpha$ is a partial order on U_α. We may assume here that M is a rate distortion manifold corresponding to some cognitive process. There is some scope as to how "$\leqslant$" may be linked to the rate distortion theorem (locally) on the U_α, in terms of channel capacity, etc. We keep in mind that our rate distortion manifolds may be subjected to 'direction' as would be required within a setting of channeled consciousness; this would involve some further analysis and grounds for a separate discussion.

Remark 6.2. As pointed out in [103], given a smooth manifold M, a po–space structure on M may be defined in terms of ordering of Morse functions $f : X \longrightarrow \mathbb{R}$ as mentioned in §6.2. Briefly, for $x, x' \in M$, we decree an ordering by $x \leq x' \iff f(x) < f(x')$, or $x = x'$. The theory of po–spaces is one of several abstract methods employed for analyzing concept structures in theoretical computer science (others, such as *Chu spaces* [106] incorporate strict logical structures and are innately different to the 'thermodynamic' features of rate distortion manifolds).

Nevertheless, it is possible that a hierarchy of contexts may be executed concurrently and the ensuing transition states may be subjected to a 'schedule' which the (cognitive) organism may inter–impose via evolution. We give a brief mathematical description following [41, 51], but for now restricting to the non–directed case. Consider the path space $\mathcal{P}(M)$ of paths

$$\gamma : \mathbb{R}_{\geq 0} \longrightarrow M, \tag{6.12}$$

of finite length. Given a covering $\mathcal{U} = \{U_a : a \in A\}$ of M by open sets indexed by a set A, a *schedule* is an element of the monoid $SA = (A \times \mathbb{R}_{\geq 0})^*$, where elements are pairs of words of the same length $(a_1a_2 \cdots a_n, t_1t_2 \cdots t_n)$. We say that a path γ *fits the schedule* $(a_1a_2 \cdots a_n, t_1t_2 \cdots t_n)$, if

i) $\gamma(t) = \gamma(t_1 + \cdots + t_n)$, for $t \geq t_1 + \cdots + t_n$

ii) $\gamma([t_1 + \cdots + t_i, t_1 + \cdots + t_{i+1}]) \subset U_{a_{i+1}}$

iii) $\gamma([0, t_1]) \subset U_{a_1}$.

Also, there is an equivalence relation on schedules generated by

$$\begin{aligned}(a_1a_2 \cdots a_n, t_1t_2 \cdots t_n) &\simeq (a_1a_2 \cdots a_{i-1}a_{i+1} \\ \cdots a_n, t_1t_2 \cdots t_{i-1}t_{i+1} \cdots t_n)&, \text{ if } t_i = 0.\end{aligned} \tag{6.13}$$

The main result of [41] states that for certain coverings, the schedules may be assigned continuously to all paths up to the latter equivalence and this can be used towards *globalizing* locally continuous fibrations over a given space. It seems workable to apply this concept to the directed case, which we finesse for now. However, for our rate distortion manifolds the main point is that the open sets of the covering each should contain neighborhoods of points constrained by the estimate (4.4) of the Rate Distortion Theorem. A slight word of caution is necessary here: the rate distortion manifold (M, s_M) as we have described it, is already a canonical model for the 'semantic path space' (X, s_X), so the above path space $\mathcal{P}(M)$ has to be considered somewhat apart from the space X. It is appealing that this notion of path scheduling may be linked to the study of universal algorithms where there is an intention to establish lower bounds on the running times of computational procedures (cf [88]). Furthermore, 'paths in a space of paths' is a potentially useful concept since eventually one may wish to consider an approach similar to the Jamesian 'processes of processes' upon which we will comment later (see §8.7).

7 Further towards groupoids

7.1 Convoluted path space

It is suggested in [125]§3 that a pattern of sensory input mixed in some way with an internal ongoing activity induces a path $\gamma = (\psi_0, \psi_1, \ldots, \psi_n, \ldots)$, where each ψ_k may represent a composition of internal/external signals. Guiding this path into some kind of decision process, yields an output $h(\gamma)$ which belongs to one of two (disjoint) sets of system response depending on whether the pattern is recognized or not. If it is the case, the appropriate response–action may assumed as initiated. This is quite general, but for the sake of classifying cognitive modules it is suggested that commencing from the input level, one may actually classify the paths themselves in preference to specifying the output.

Typically, an input x representing an information source, is tied to an output y via some path; for example, a path representing the transition probability $p(x|y)$, or a channel of information. In another sense, one can define *equivalence classes* in a convoluted path space, such as (X, s_X), according to which a state ψ_k is path–connected to a source state ψ_S. In this way, two states $\psi, \hat{\psi}$ are said to be *equivalent*, denoted $\psi \mathcal{R} \hat{\psi}$, if they lie on the same path γ with source ψ_S seen as varying. In this way, the path space is decomposable into (relatively) disjoint sets of equivalence classes. Such an equivalence relation defines a category known as a *groupoid*, a 'small' category G with all morphisms invertible, represented by

$$\mathsf{G} \overset{r}{\underset{s}{\rightrightarrows}} M \tag{7.1}$$

where M denotes the space of *objects* and r, s denote the range (or target) and source maps, respectively (see e.g. [21, 143]). Such disjoint equivalence classes are applicable to disjoint 'cognitive modules' for which the equivalence relation is defined by the existence of a *high probability meaningful path* connecting two points. Later we will discuss a network groupoid in which the vertices of the network (or graph) will represent different information sources dual to a cognitive process. Certainly, the study of equivalence classes for dealing with e.g. response versus sensory input, is an attractive option to analyzing an overwhelmingly complex network since the key principle would be to reduce the latter to manageable configurations involving only the (equivalence) classes.

Specifically, in our case (M, s_M) is viewed as the canonical model of (X, s_X) which, as we proposed, could be replaced by the groupoid G under path equivalence. An action on M by G, induces an equivalence relation $\mathcal{R}$, together with a convolution product

$$(a * b)(\gamma) = \sum_{\gamma_1 \circ \gamma_2 = \gamma} a(\gamma_1) b(\gamma_2). \tag{7.2}$$

By using general means (see e.g. [31]) we can form a corresponding *convolution algebra* $\mathcal{C}(\mathsf{G})$ over G of which many special cases can be realized in a systems–response mechanism. Typically, in a response to an environmental stimulus, a 'response' function $h : \mathcal{C}(\mathsf{G}) \longrightarrow B$, mostly nonlinear, can be defined where B is such an extensible set as before, and which could be the underlying set of some semantic/syntactical algebra, or that of an algebra of some class of operators. The simple epidemic model as a 'generalized stochastic resonance' mentioned in §3.11 is an example that immediately comes to mind.

7.2 Geometric phase and holonomy

The concept of holonomy in a physical sense could be loosely described by the following scenario: imagine walking along a path of some gradient flow. You may observe that neighboring flow paths tend to veer off; but as you progress steadily further, other flow paths appear to approach asymptotically. The explanation is well–known to anyone who has taken a first course in differential geometry: holonomy is essentially the parallel translation of vectors around a closed path, thus leading to a representation of the space of closed paths into a group of prevailing symmetries. The classic example involves the *Poincaré first–return map* of a dynamical system (see e.g. [1, 77, 96]). The holonomy concept embraces the sense of phase transition throughout the physical and biological sciences in whatever the context and wherever the internal states of a system are tracked in relationship to the latter's spatiotemporal orientation. Notable physical examples include the Berry phase, whereby a slowly evolving quantum system in returning to its original state retains a memory of its motion via a geometric phase in the wave function, a phase as given by $\exp(\iota \int_\gamma A)$, where A is a suitable potential and γ is the path in question. Likewise in the Born–Oppenheimer approximation, as nuclei describe a closed path in a certain parameter space, the electronic wave function acquires such a phase. There is the more mundane example of a cat held upside down and then released from a reasonable height. The cat usually lands safely on its feet but with its orientation reversed [97], thus realizing holonomy as incorporated within a certain innate cognitive–physiological skill.

Parallel transport induced by a 'flat' connection/potential having zero field strength (curvature), but nevertheless having non–trivial holonomy, causes shifting interference patterns in electron beams in the vicinity of a solenoid (cf the Aharonov–Bohm effect and Wilson loop [5, 85]). So in a similar way, the key to understanding how seemingly disjoint cognitive modules interact lies within globalizing the iterates of such local procedures to create the associated holonomy groupoid (a technique described for topological groupoids in [4]). In the skeletal framework of graphs and networks, holonomy can be described in the context of symmetry groups.

7.3 Noise flow on a rate distortion manifold

There is a convenient approach to noise flow on a rate distortion manifold in terms of *groupoid actions*. One may consider a system of noise variables $\mathcal{B} = (\varsigma_1, \varsigma_2, \ldots, \varsigma_\ell)$ associated to an informational process associated modeled by a rate distortion manifold M. If the noise is network related, then it is reasonable to speak of 'equivalence classes of noise' and hence an associated groupoid B whose set of objects would be a network of paths in M. The essential point here is that given the groupoid B acts on M, the equivalence classes 'foliate' M in some way. This is simply the principle that a foliation on M corresponds to a groupoid, and conversely, the foliation is induced by the action of B:

$$\begin{aligned} &\Big\{ \text{Groupoid Action } \mathsf{B} \times M \longrightarrow M \Big\} \\ &\Longrightarrow \Big\{ \text{Foliation } (M, \mathcal{F}) \Big\}. \end{aligned} \tag{7.3}$$

Generally, this foliation will be *singular* as, for instance, the leaf dimensions may jump up and down. For a study of the general theory of foliations see [26] (cf [92, 96] which deal with groupoid actions). On the other hand, the noise equivalence classes may in practice be 1–dimensional, in which case we have a *noise–induced flow* $(M, \mathcal{F})$ on M that is essentially stochastic in nature, and most likely singular in the general case as well.

Relative to a metric $\mathcal{M}$ on M, we have the simplest type of Onsager relation $d\mathcal{M}/dt = L dS/d\mathcal{M}$ to which a noise term can be added:

$$\frac{d\mathcal{M}}{dt} = L \frac{dS}{d\mathcal{M}} + \sigma W(t), \tag{7.4}$$

where $W(t)$ is a function representing white noise and σ is a constant. In this way a stochastic differential equation is induced on $\mathcal{M}$ [134]:

$$d\mathcal{M}_t = L(\frac{\partial}{\partial t}, \frac{\partial S}{\partial \mathcal{M}})\, dt + \sigma(\frac{\partial}{\partial t}, \frac{\partial S}{\partial \mathcal{M}})\, dB_t, \tag{7.5}$$

where L and σ are now regarded as suitable functions of t and $ds/d\mathcal{M}$, and dB_t represents the noise structure derivable from the noise variables $\mathcal{B}$ as above. Thus the above groupoid action is manifestly the noise flow on M engendered by the 1-form component $\sigma(\frac{\partial}{\partial t}, \frac{\partial S}{\partial \mathcal{M}})\, dB_t$.

7.4 Global actions and groupoid atlases

We expect that more general descriptive mechanisms of cognitive modules may eventually involve features such as atlases modeled on groupoids themselves. Once instance involves an *orbifold* for which the associated orbifold atlas corresponds to a proper (Lie) groupoid (see [96] and Appendix §11.2). Thus a 'rate distortion orbifold' would then be a fitting term when a rate distortion space has to admit certain singularities. However, an even broader concept is that of a *groupoid atlas*[16]. The latter, loosely speaking, entails the pasting together of local groupoid actions with the net effect of a 'global' groupoid, a concept which may prove to be particularly significant for logically inscribing processors or sensors (the 'multi–agents') within a cognitive module or a communication network.

Following [16], one commences from a family of groupoids $\{\mathsf{G}_1, \mathsf{G}_2, \ldots\}$ where each groupoid has the same set of objects; this family is called a *single domain* or *multiple groupoid. A groupoid atlas* is then defined as a set with a covering by patches, each of which comprise a single domain with global action. An advantage of using this sort of atlas is that, in general, it admits a weaker structure compared with that of a conventional manifold since no condition of compatibility between arbitrary overlaps of the patches is necessary. This is an attractive option for studying cognitive modules geared to equivalence class representations. For instance, in a (single domain) global action by a group, the graph representation of intersecting orbits yields a configuration of various types of circuits (loops, etc.), and from such an action, a corresponding global action can be formulated so that group actions, in particular, can be more generally extended to groupoid actions that encode the actions of the various equivalence relations.

On taking a group G, the motivation for defining a groupoid atlas comes from considering a *global action* $\mathcal{A}$ which consists of a set $X_{\mathcal{A}}$ together with a family

$$\begin{aligned}&\{(G_{\mathcal{A}})_\alpha \curvearrowright (X_{\mathcal{A}})_\alpha : \alpha \in \Psi_{\mathcal{A}}\}\\ &= \{(G_{\mathcal{A}})_\alpha \times (X_{\mathcal{A}})_\alpha \longrightarrow (X_{\mathcal{A}})_\alpha : \alpha \in \Psi_{\mathcal{A}}\},\end{aligned} \tag{7.6}$$

of group actions '$\curvearrowright$' on subsets $(X_{\mathcal{A}})_\alpha \subseteq X_{\mathcal{A}}$, where the local groups $(G_{\mathcal{A}})_\alpha$ and the corresponding subsets $(X_{\mathcal{A}})_\alpha$ are indexed by an indexing set $\Psi_{\mathcal{A}}$ called *the coordinate system of* $\mathcal{A}$, satisfying the conditions:

(a) If $\alpha \leq \beta$ in $\Psi_{\mathcal{A}}$, then $(X_{\mathcal{A}})_\alpha \cap (X_{\mathcal{A}})_\beta$ is $(G_{\mathcal{A}})_\alpha$–invariant.

(b) For each pair $\alpha \leq \beta$, there is given a group homomorphism
$$(G_{\mathcal{A}})_{\alpha\leq\beta} : (G_{\mathcal{A}})_\alpha \longrightarrow (G_{\mathcal{A}})_\beta,$$
such that given elements $\sigma \in (G_{\mathcal{A}})_\alpha$, and $x \in (X_{\mathcal{A}})_\alpha \cap (X_{\mathcal{A}})_\beta$, we have $\sigma x = (G_{\mathcal{A}})_{\alpha\leq\beta}(\sigma)x$.

The categorical assignment

$$G_{\mathcal{A}} : \Psi_{\mathcal{A}} \longrightarrow \mathsf{Groups} \tag{7.7}$$

is called the *global group of* $\mathcal{A}$, and the set $X_{\mathcal{A}}$ is called the *enveloping set* or the *underlying set of* $\mathcal{A}$.

Suppose we have a group action $G \curvearrowright X$. Then we have a category $\mathsf{Act}(G, X)$ with object set X and $G \times X$ its arrow set. It is straightforward to show that $\mathsf{Act}(G, X)$ is actually a groupoid [16](see also Appendix I). Effectively, given an arrow (g, x), we have source and range defined respectively by $s(g, x) = x$, and $r(g, x) = g \cdot x$, represented by

$$x \overset{(g,x)}{\longrightarrow} g \cdot x. \tag{7.8}$$

The composition of (g, x) and (g', x') is defined when the range of (g, x) is the source of (g', x') such that $x' = g \cdot x$. This yields a composition $(g'g, x)$:

$$x \overset{(g,x)}{\longrightarrow} g \cdot x \overset{(g',gx)}{\longrightarrow} g'g \cdot x. \tag{7.9}$$

We have an identity at x given by $(1, x)$, and for any element (g, x) its inverse is (g^{-1}, gx). A key point in this construction is that the orbit of a group action then becomes a connected component of a groupoid.

The above account motivates the following. A *groupoid atlas* $\mathcal{A}$ on a set $X_{\mathcal{A}}$ consists of a family of local groupoids $(\mathsf{G}_{\mathcal{A}})$ defined with respective object sets $(X_{\mathcal{A}})_\alpha$ taken to be subsets of $X_{\mathcal{A}}$. These local groupoids are indexed by a set $\Psi_{\mathcal{A}}$, again called the *coordinate system of* $\mathcal{A}$ which is equipped with a reflexive relation denoted by $\leq$. Writing $(X_{\mathcal{A}})_{\alpha\beta} = (X_{\mathcal{A}})_\alpha \cap (X_{\mathcal{A}})_\beta$, this data is to satisfy the following conditions [16]:

(1) If $\alpha \leq \beta$ in $\Psi_{\mathcal{A}}$, then $(X_{\mathcal{A}})_{\alpha\beta}$ is a union of components of $(\mathsf{G}_{\mathcal{A}})$, that is, if $(X_{\mathcal{A}})_{\alpha\beta}$ and $g \in (\mathsf{G}_{\mathcal{A}})_\alpha$ acts as $g : x \longrightarrow y$, then $y \in (X_{\mathcal{A}})_{\alpha\beta}$.

(2) If $\alpha \leq \beta$ in $\Psi_{\mathcal{A}}$, then there is a groupoid morphism defined between the restrictions of the local groupoids to intersections
$$(\mathsf{G}_{\mathcal{A}})_\alpha|(X_{\mathcal{A}})_{\alpha\beta} \longrightarrow (\mathsf{G}_{\mathcal{A}})_\beta|(X_{\mathcal{A}})_{\alpha\beta}, \tag{7.10}$$
and which is the identity morphism on objects.

We can briefly exemplify matters as follows. Let us recall the projection of information sources $\Phi : X \longrightarrow Y$, from the higher (faster) X to the lower (slower) Y, and recall from (4.6) we defined a rate distortion manifold $\Gamma_{\mathrm{rd}} \subseteq \Gamma_\Phi$ on the graph of Φ, by

$$\Gamma_{\mathrm{rd}} := \{\, (x, \Phi(x)) \in \Gamma_\Phi : d(x, \Phi(x)) < \epsilon \,\}. \tag{7.11}$$

Let G_Z be a group whose elements are, for instance, matrix components suitably representing those of a culture or environment via a slowly interacting source $Z = \{Z_k\}$ [134].

We consider an action $G_Z \curvearrowright \Gamma_{\mathrm{rd}}$ along the previous lines, and subsequently obtain an action groupoid

$$\mathsf{Act}(\Gamma_{\mathrm{rd}}, G_Z) \rightrightarrows \Gamma_{\mathrm{rd}} \ . \quad (7.12)$$

The significance of this construction is that the components of the above action groupoid may be related to the mutual information splitting criterion

$$I(X|Y_1, \ldots, Y_m|Z_1, \ldots, Z_n), \quad (7.13)$$

an essential ingredient for representing such interactions as those of a biosocio–culture [128].

8 The underlying simplex of a rate distortion manifold and network groupoids

Here we devote some attention to how the simplicial methods of graphs and networks, which are some of mainstream tools of information theory, multi–agent systems and concurrency, can be related to the often 'continuous' structures of rate distortion manifolds.

8.1 Simplicial methods towards networks

Given our rate distortion manifold M, we may want to recover an underlying decomposition of M into 'networks' so as to incorporate useful graph–theoretic techniques. A standard topological way of achieving this is by introducing simplicial methods and in particular, by taking a *triangulation* of M (see Appendix II) that serves as a conceptual mechanism towards tracing the discrete nature of the various graphs and networks functioning as specialized processors within the Workspace. Specifically, a *triangulation* (K, ϕ) of a space M means we have a simplicial complex K together with a homeomorphism

$$\phi : |K| \longrightarrow M, \quad (8.1)$$

where $|K|$ denotes the *polyhedron* or *geometric realization* of M (see e.g. [117] and Appendix II). Observe that simplicial methods often deal with choices of an open covering and for a given rate distortion manifold M, such a covering may be achieved by a collection of length spaces $\{U_\alpha, d_\alpha\}$, each of can be taken to be isometric to some simplex.

The triangulation (K, ϕ) of M that we have described above leads to identifying M with an associated polyhedral space, and so there follows a number of 'discrete' possibilities leading from a 'continuous' to a 'discrete' coarse–graining approach. Other possibilities may include the comparison theorems of spaces of bounded curvature with their negative curvature characteristics such as the $CAT(k)$–spaces (*Cartan–Alexandrov–Toponogov spaces* where the (k) denotes that a value k is imposed as a curvature bound [24]). This leaves open the possibility that some class of rate distortion manifolds, realized in the category of $CAT(k)$ spaces, might, for instance, admit a 'hyperbolic structure' in a suitable sense.

A topological graph can be converted to a metric graph by an assignment of 'lengths' to edges, although for infinite graphs this may result in a topology change [24]. More specifically, a simplicial decomposition of M permits an internal, skeletal–like representation of M in terms of a graph or network Γ, and then subsequently to an associated groupoid model. We can start with a categorical representation $\mathsf{FreeCat}(\Gamma)$ of the graph Γ : regarding the vertices of Γ as objects, then between two vertices v, w, we take $\mathsf{FreeCat}(\Gamma)(v, w)$ to denote the set of paths or edges in Γ commencing at v and ending at w. To a path $v \mapsto w$, we can assign a sequence of edge–labels $(a_1, a_2, \ldots, a_n)$. The composition in $\mathsf{FreeCat}(\Gamma)$ is by the usual concatenation of paths where for each edge a between v and w, a reciprocal (or reverse) edge a^{-1} between w and v exists.

In forming path sequences, the latter can be reduced by removal from the sequence of any adjacent edges of the form (a, a^{-1}), or (a^{-1}, a). In this way the graph Γ leads to a groupoid structure, namely the free groupoid $\mathsf{FreeGpd}(\Gamma)$ over Γ (see e.g. [21, 146]). Thus many of the graph–theoretic and network analysis models relating to phase transition, percolation and epidemic processes, etc., can be reduced to a combinatorial groupoid setting for which there is already available a broad range of algebraic concepts that can be applied.

8.2 Modular networks and the giant component

Between disjoint cognitive modules one assumes that linkages occur randomly and the latter represent 'cross–talk' as a (non–zero) measure of mutual information. A descriptive method for studying this influence of cross–talk uses random graph theory (mainly following Erdős–Rényi [50]; we also refer to the exposition in [2]). One of the key concepts is that of a *giant component*, that is, a subnetwork that dominates the entire network (of cognitive modules) and which can capture up most of the smaller subnetworks.

More specifically, suppose we consider c elements of the equivalence class algebra of languages (that is, c disjoint cognitive modules) dual to some cognitive process as represented by the vertices of a graph. If a graph with c vertices has $\ell = \frac{1}{2}ac$ edges chosen at random, for $a > 1$, then it will have a giant connected component with approximately $g(a)c$ vertices with

$$g(a) = 1 + \frac{1}{a} W(-a \exp(-a)), \quad (8.2)$$

in which W denotes the Lambert W–function defined implicitly by $W(x)\exp(W(x)) = x$. An example depicted in [126], reveals a sharp phase transition occurring at $a = 1$ that initiates the Global Workspace as a 'tunable blackboard' defined by a set of cross–talk mutual information measures between interacting modules. The cross–talk connection corresponds to random linkages in the case of

[50] and the entropy H of the language dual to the cognitive process, will grow as some monotonic function of the giant component. Such a phase transition in a network affords some correlation between the size of the component and the richness of the language to which it is associated. Thus tuning the giant component by altering the network topology leads to further insight through a geometric representation.

Standard analysis of critical clustering reveals that uplifting the clustering coefficient increases the average number of edges necessary for the formation of a giant component. For instance, [126] on applying [124], shows that for a random network with parameter a, at cluster value C, there is a critical value given by

$$a_C = \frac{1}{1 - C - C^2}. \tag{8.3}$$

One easily sees that for $C = 0$, the giant component forms when $a = 1$. The case $C \geq \frac{1}{2}(\sqrt{5}-1)$, which is the Golden Section, shows that no giant component is definable, for any a. Thus not every network topology can support a giant component. As pointed out in [126], some cognitive network models cannot then represent consciousness and this poses some big questions ranging from the evolution of the latter to the actual nature of the sleeping state (cf [127]). Institutional or machine–based cognition no less necessitates the synchronization of information relay between giant components [134, 135].

A governing principle is to define an interaction parameter $\omega_C > 0$ that will define a regime of giant components of network elements linked via mutual information $\geq \omega_C$. Then following [134, 135], the idea is to *invert the argument*: namely, a given topology of the giant component will in turn define some value of ω_C, so that network elements interacting by mutual information less than ω_C will be blocked from conscious perception (see the example of inattentional blindness below). Thus ω_C is a syntactically dependent detection limit which depends on the giant component topology for an individual cognitive framework. Thus the variation of ω_C is one example of a topological shift. This opens up the possibility that the level sets of ω_C may be defined in terms of Morse theory. Accordingly, a parameter space may be characterized by the critical points of ω_C to ensure a fundamental shift in the high level cognitive topology.

8.3 Inattentional blindness

We briefly describe a situation from [128, 135] that has already been mentioned. An intensive focus on a task involving interactive cognitive modules may necessitate the giant component to be sustained at an optimum level within the topology of the network in question. In this way, a high limit may be placed on the magnitude of mutual information signals which can intrude into the Workspace. When the focus of attention on a single aspect of a complicated perceptual field or programmed environment precludes the detection of intervening events which may or may not be essential to the original task, a condition known as *inattentional blindness* occurs [115]. An example of this condition might be that of a person conducting an on–line business transaction while oblivious of occasional 'pop-ups'. In this scenario it may be that intervening signals fall below a threshold in syntax in order to intrude markedly on consciousness; alternatively, it fails to be an enduring competitor in the Workspace (cf §6.1). Further, it is expected that slower acting information sources represent the embedding sociocultural factors across the environment. In the context of institutional/directed cognition, the intense focus on economic and data–driven programs may often result in a blind–side to other essential factors (such as the sociological consequences of planned shrinkage, industrial expansion, etc.). Given a fixed topology of the Workspace, the condition of inattentional blindness thus emerges as a thermodynamic limit on the overall processing capacity [128, 135].

8.4 Network phase transitions via connections on graphs and groupoids

The finer study involves the nature of phase transitions within the simplicial network/graph structure underlying the geometry of a rate distortion manifold which we will proceed to describe. Typically, various types of percolation processes exhibit phase transitions. For instance, in [84] network percolation techniques are used to analyze phase transitions of dynamic neural systems such as those embedded within segments of cortical neuropil. But for a large class of networks there are available means for measuring phase transitions and differences in terms of parallel transport and holonomy which are analogous to the standard differential–geometric means employed on a differentiable manifold. We shall briefly discuss some of these.

Firstly, for graph–theoretic models there are certain combinatorial notions which can be used to replicate a 'differential' structure as realized on a standard differentiable manifold. Let $\Gamma = (V, E)$ be a graph with V denoting a finite vertex set, E an edge set with an oriented edge $e = (u, v)$ (accordingly, $e^{-1} = (v, u)$) such that $u = i(e)$ is the initial vertex and $v = t(e)$ is the terminal vertex. The *star of a vertex* $\mathsf{st}(v)$ is the set of edges emanating from v, that is

$$\mathsf{st}(v) = \{e : i(e) = v\}, \tag{8.4}$$

(see also Appendix II). Given that the star of a vertex is sometimes viewed as the combinatorial version of the tangent space to a manifold at a point, in [18] is defined a *connection* ∇ on a graph Γ as defined via a set of one–to–one functions $\nabla(u, v)$, one for each oriented edge $e = (u, v)$ of Γ satisfying:

(1) $\nabla(u, v) : \mathsf{st}(u) \longrightarrow \mathsf{st}(v)$

(2) $\nabla(u, v)(u, v) = (v, u)$

(3) $\nabla(v, u) = (\nabla(u, v))^{-1}$.

Assuming a graph Γ admits such a connection ∇, in [18] is defined the notion of a *3–geodesic* as a sequence of four vertices (u, v, w, z) with edges $\{u, v\}$, $\{v, w\}$ and $\{w, z\}$ for which $\nabla(v, w)(v, u) = (w, z)$. Subsequently, a *k–geodesic* is defined inductively as a sequence of $(k+1)$ vertices. The three consecutive edges $\{d, e, f\}$ of a 3–geodesic is referred to as *an edge chain*. A *closed geodesic* can then be specified as a sequence of edges $e_1, \ldots, e_n$ such that each consecutive triple $(e_\alpha, e_{\alpha+1}, e_{\alpha+2})$ is an edge chain for each $1 \leq \alpha \leq n$(modulo n). This leads to a convenient way of defining a 'totally geodesic subgraph' [18], that is, given (Γ, ∇), a subgraph $\Gamma_0 = (V_0, E_0) \subset \Gamma$, is said to be *totally geodesic* if all geodesics commencing at E_0 remain within E_0. In other words, for every two adjacent vertices u, v in Γ_0, we have

$$\nabla(u, v)(\mathsf{st}(u) \cap E_0) \subseteq E_0. \tag{8.5}$$

8.5 The covariant derivative of entropy along a network path

Suppose now the above vertices $(e_1, e_2, \ldots, e_{k+1})$ are interpreted and renamed as $k + 1$ information sources $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{k+1})$ in accordance with the APSE condition (see Remark 3.2), where the $\mathbf{X}_i$ act with the set of tuning parameters associated to a set of giant components. We consider a connection ∇ acting

$$\nabla(\mathbf{X}_i, \mathbf{X}_j) : \mathsf{st}(\mathbf{X}_i) \longrightarrow \mathsf{st}(\mathbf{X}_j), \tag{8.6}$$

with the indicated properties (for $1 \leq i, j \leq k + 1$) as before. With respect to the metric $\mathcal{M} = \mathcal{M}(\mathbf{X}_i, \mathbf{X}_j)$ applied to these information sources, the above 'connection' map in (8.6) implements on the underlying network, the covariant differentiation along the path $\mathbf{X}_i \longrightarrow \mathbf{X}_i$, just as in (3.30):

$$dH/d\mathcal{M} = \lim_{\mathbf{X}_j \longrightarrow \mathbf{X}_i} \frac{H(\mathbf{X}_j) - H(\mathbf{X}_i)}{\mathcal{M}(\mathbf{X}_i, \mathbf{X}_j)}. \tag{8.7}$$

Now relative to each $\mathbf{X}_i$, a maximized channel capacity C_i is assigned, in accordance with the estimate of (3.15), that is, $H(\mathbf{X}_i) \leq C_i$, holds for $1 \leq i \leq k + 1$, and in respect of the Rate Distortion Theorem along paths $\mathbf{X}_j \longrightarrow \mathbf{X}_i$. This apparent optimality in terms of the estimate (3.15) motivates decreeing the information sources $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{k+1})$ to be a k-geodesic (there is no loss in generality by supposing that these actually form a closed geodesic). In fact, [18] shows that the set of all such geodesics in a given network determines the connection ∇ which accordingly can be implemented as a form of covariant derivative along the remaining paths.

Remark 8.1. In order to realize how geodesics may arise in applied graphs and networks, there is the example of [73] where in the ever competing US telecommunications industry, extensive and rapid switching of networks between nodes (vertices), the large–scale use of fiber optics (reducing transmission costs) and network nodes ramifying to interconnecting subnetworks, has lead to the reduction of a vast pyramidial–like network to a structure with many geodesic subgraphs. Thus the pyramid transforms to a structure somewhat akin to Buckminster Fuller's 'geodesic dome'. The principles are analogous to those expected in huge networks of parallel computation which in turn contribute efficiency to the skeleton of some institutional cognitive mechanism.

8.6 The holonomy groupoid and path connections

Given $C = \{e_1, \ldots, e_n\}$ is any cycle in Γ, for which $t(e_\alpha) = i(e_{\alpha+1})$ modulo n, then the connection around C leads to a permutation

$$\nabla_C = \nabla_{e_n} \circ \cdots \circ \nabla_{e_1} \circ \nabla_{e_0}, \tag{8.8}$$

of $\mathsf{st}(u)$. The upshot is that the notion of a holonomy group at a vertex can be defined [18]: *the holonomy group* $\mathrm{Hol}(\Gamma_u)$ at a vertex u of Γ, is the subgroup of the permutation group of $\mathsf{st}(u)$ generated by the permutations ∇_C over all cycles C that pass through the vertex u. In this way, holonomy contributes to the geometry of the graph.

On the other hand, as noted in [82] every groupoid G leads to a *reflexive symmetric graph* (RSG). Loosely speaking, the objects form the vertices, arrows form the edges, inversion in the groupoid leads to symmetry, and the identity leads to the pointwise identity arrow. For instance, on a manifold M the set $\mathcal{P}(M)$ of (smooth) Moore paths $\gamma : [a, b] \longrightarrow M$, has the structure of a RSG with M the vertex set and $\gamma(a), \gamma(b)$ defined to be the domain and codomain of γ respectively. Taking $M_{(1)}$ to denote the *first neighborhood of the diagonal of* M [82], a connection ∇ on $\mathsf{G} \rightrightarrows M$ is a morphism of a RSG from $M_{(1)}$ to the underlying graph of G. Let $(x, y) \in M_{(1)}$, then similar to above, $\nabla(x, y)$ is an arrow $x \mapsto y$ in G, such that

i) $\nabla(x, x) = \mathrm{id}_x$

ii) $x \sim y \implies \nabla(y, x) = (\nabla(x, y))^{-1}$.

A *path connection* on G is a morphism of RSGs in the case of $\mathcal{P}(M) \longrightarrow \mathsf{G}$ that satisfies certain rules of reparametrization, representation and subdivision (we refer to [82] §6 which follows in part [123]).

Many groupoids with connection

$$(\mathsf{G} \rightrightarrows M, \nabla) \tag{8.9}$$

have the property that unique partial integrals exist along any map $[a, b] \longrightarrow M$. Thus we may say that (8.9) admits path integration. Consequently, a connection ∇ is then 'flat' along any path, that is, $\nabla(x, y) \circ \nabla(y, z) = \nabla(x, z)$.

For $u \sim v$ in $[a, b]$, let us set

$$\left(\int_\gamma \nabla\right)(u, v) = \nabla(\gamma(u), \gamma(v)). \tag{8.10}$$

Following [82] we define the holonomy $\mathrm{hol}_\nabla(\gamma)$ along a path $\gamma : [a, b] \longrightarrow M$, as the arrow $(\int_\gamma \nabla)(a, b)$ with domain

$\gamma(a)$ and codomain $\gamma(b)$. Thus a map

$$\mathrm{hol}_{\nabla} : \mathcal{P}(M) \longrightarrow \mathsf{G}, \tag{8.11}$$

is obtained. If (8.9) admits path integration, then the above map in (8.11) is considered to be a path connection.

There appears to be a close relation of these ideas to [61, 118] wherein are considered n–cell systems based on systems of n ordinary differential equations describing the dynamics of some (possibly) synchronized network. This is important because the synchronous coupling of a cell system to its close environment affects a change in the latter as well as in the collective organism whose task it is to square–up to those of the higher, multi–parallel, institutional types. The coupling and equivalence of cells leads to a natural groupoid structure of a resulting coupled cell network $\Gamma = (V, E, \sim_v, \sim_e)$ with its intrinsic equivalence classes $[v]_V$ and $[e]_E$. Here, the vertices or nodes of the network are taken to be representative of such cells. Synchrony of such cell systems may be dependent on groupoid symmetries which, as pointed out in [131] in the context of institutional cognition, can be broken by an impinging rapid crosstalk internal to the system while the latter attempts to manage a slower external crosstalk.

For each $v \in V$, a *cell phase space* P_v is defined. Usually P_v is a finite dimensional vector space and a *total phase space* is defined as $P = \prod_{v \in V} P_v$. A vector field may then be characterized in terms of a map $f : P \longrightarrow P$, that in principle should be related to the above permutation subgroups of $\mathsf{st}(u)$ thus leading to a suitable notion of *parallel transport* within the system.

We have at this stage arrived at a formalism for obtaining a network/graph theory underlying a typical rate distortion manifold, similar to taking an X–ray picture of an essential organism. Beyond the example of Remark 8.1 there are many possible applications such as in areas where one considers the passage from an iteration of local processes towards global structures. For instance, the situation described in Remark 8.1 is likely to have analogs in the study of social networks. These may be the 'small world' graphs having low density and which are highly clustered thus giving rise to the likelihood of networks of geodesic subgraphs [142]. 'Small world' relationships are studied in [109] in a similar way to how strong ties (families, cliques, etc.) with large clustering are bridged by weak ties [63]. The corresponding social networks are likely to involve more intricate topologies and statistical fluctuations, and where 'simplicial' Nash equilibria may provide optimal predictions within the resulting framework of games [120]. Additionally, the graph holonomy concept and the 'giant component' thus provide formal criteria in which to specify the essential phase transitions leading to higher orders of complexity. For small world networks in the context of Global Workspace Theory, steps in this direction have been taken in [60].

8.7 2–Groupoids and Stacks

To some degree the cognitive modules we have considered should afford a Jamesian characteristic of "processes of processes". The key is to take a step up in 'categorical dimension'. Loosely speaking, a *2–category* $\mathfrak{C}_2$ can be described in a 'cellular' sense: $\mathfrak{C}_2$ consists of a class of objects $\mathbb{O}$ (0–cells), a class of 1–morphisms $\mathbb{A}_1$ (1–cells), a class of 2–morphisms (2–cells) with 'horizontal' composition defined between 1–and 2–cells, along with a separate 'vertical' composition between 2–cells. In other words $\mathfrak{C}_2$ affords the extra mechanism of *morphisms between morphisms*. When the 2–morphisms of $\mathfrak{C}_2$ are invertible and the 1–morphisms invertible (up to homotopy), then $\mathfrak{C}_2$ shapes up as a *2–groupoid*. Suitable reference to this subject are e.g. [19, 76, 78, 91].

The Cartesian closed category Cat of small categories is a 2–category in which the 2–morphisms are the natural transformations for which the vertical composition is given via composition in the codomain category. Also, the category of groupoids Gpd is a (full) 2–subcategory of Cat. A 2–functor $\mathsf{F} : \mathfrak{C}_2 \longrightarrow \mathfrak{D}_2$ is an enriched functor in Cat that preserves the 2–category structure of $\mathfrak{C}_2$ on taking objects, 1–and 2–morphisms of $\mathfrak{C}_2$ to those in $\mathfrak{D}_2$.

In relationship to manifold structures, the Yoneda lemma says that any space or manifold M is uniquely determined by the categorical functor

$$\mathsf{Map}\ (\ , M) : \mathsf{Mnf} \longrightarrow \mathsf{Sets} \tag{8.12}$$

A *stack* S is a (2)–functor between categories of manifolds and groupoids (with categories)

$$\mathsf{S} : \mathsf{Mnf} \longrightarrow \mathsf{Gpd} \subset \mathsf{Cat} \tag{8.13}$$

where for any manifold M, we obtain a corresponding category $\mathsf{S}(M)$ in which all morphisms are isomorphisms, for any morphism $f : N \longrightarrow M$, we have a functor $f^* : \mathsf{S}(M) \longrightarrow \mathsf{S}(N)$, and for any $Z \xrightarrow{g} N \xrightarrow{f} M$, there is a natural transformation $T_{f,g} : g^* f^* \cong (g \circ f)^*$ which is associative on a trio of compatible morphisms.

As shown in e.g. [68, 86, 91], such a functor S also enjoys the properties of glueing together all of the objects and morphisms. Furthermore, S can itself admit an chart/atlas description generalizing that for a manifold, thus leading to a potentially useful concept for a further large–scale study of interactive cognitive modules in the same way as the groupoid atlas has been proposed. The above account is one categorically formal means of representing a rate distortion manifold as groupoid (or, to consider a stack of groupoids upon the former). In the 2–categorical sense, one then contemplates a next step up from 'meaningful paths' to 'meaningful membranes' towards realizing a higher order Global Workspace continuum (see Remark 8.3 below). In another, but related context, the idea of 'morphisms between morphisms' may be relevant to the passage from 1st order to 2nd order complexity of information in terms of 'referents', in so far that the 2nd order houses the sense of 'meaning' [10].

Remark 8.2. In view of earlier remarks concerning parallel processing, it would be reasonable to append to the model a chain of rate distortion submanifolds $M_1 \subseteq M_2 \subseteq \cdots \subseteq M_n = M$, on each of which there is a groupoid structure corresponding to an equivalence relation "$\sim_i$" (cf *path equivalence*), for $1 \leq i \leq n$:

$$\begin{aligned} (\mathsf{G}_1, \sim_1) &\rightrightarrows M_1 \\ (\mathsf{G}_2, \sim_2) &\rightrightarrows M_2 \\ \cdots\cdots &\ \cdots\cdots \\ (\mathsf{G}_n, \sim_n) &\rightrightarrows M_n \end{aligned} \tag{8.14}$$

A broader framework could be related to the 'stack' functor S as previously, where any such nested sequence (of information) would yield a corresponding nested groupoid sequence as stacks over the manifolds M_k.

Remark 8.3. We have mentioned that geometric concepts such as parallel transport and holonomy may be realized within graphs and networks. One instance of a 2–categorical approach to 2–parallel transport using simplicial methods is described in [11] by means of a 'sweeping functor'. There is also the related work of [23] which is relevant to surface holonomy.

9 Some applications towards cognition at–large

In previous sections we described the mathematical architecture of the possible rate distortion manifolds and network related ideas. Next we discuss the motivating informational background from the point of view of immunology-language and several classes of cognition with possible ramifications.

9.1 The Atlan–Cohen perspective

The immunology–information principle as outlined in [8] starts with sets of strings of amino acids in an antibody molecule poised to influence the quantity of information in the corresponding protein. Recall that protein synthesis as a channel of information is transcribed into the protein amino acid sequence which acknowledging the genetic code whereby DNA stores information in the neucleotide bases A(Adenine), C(Cytosin), G(Guanine), T(Thymine).

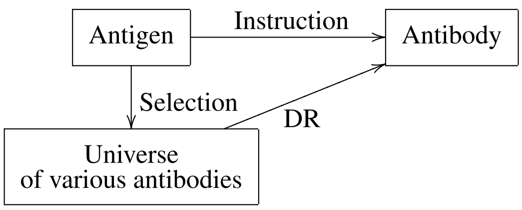
Biological interactive networks as a class of complex networks consist of local cellular communities organized and managed by their characteristic selection procedures. In such a partitioning of the structure, it is necessary to regulate the local properties of the organism rather than the global mechanism while genetic switches operate as transcription factors encoding and switching on other genes within this hierarchy. Moreover, one can include systems which by their intrinsic structure interact via noncommutative relationships. More specifically, inter–regulatory systems of genetic networks via activation or inhibition of DNA transcription can be modeled at several differing levels where various factors influence distinct states usually by some embryonic process or by the actual network structure itself. For each gene it is important to understand the dynamics of inter–regulatory genetic groups which of themselves create hierarchial systems with their own characteristics. A gene positively (or negatively) regulates another when the protein coding of the former activates (respectively, inhibits) the properties of the latter. In this way, genetic networks are comprised of interconnecting positive and negative feedback loops. The DNA binding protein is encoded by a gene at a vertex i say, activating a target gene j where the transcription rate of i is realized in terms of a function of the concentration x_j of the regulatory protein. Acting towards a given gene, the regulating genes are protein coded and induce a transcription factor. Subsequent modeling techniques can be drawn from a variety of mathematical sources : graph theory, stochastic differential equations, and Boolean networks are examples (specific approaches are realized by de Jong et al.[35]). An overall exposition of these ideas from the categorical viewpoint and that of higher dimensional algebra is presented in [15].

Immune networks had been proposed by Jerne [75] as networks of mutually interacting and cooperative 'idioytpes' and 'anti–idiotypes' as regulators of immune response towards antigenic approaches through which the antigen itself reveals a 'meaning'. In relationship to this hypothesis, a main premise of [8] is that an antigen should be viewed as a fundamental unit of information. However, it is postulated that noise prevails in the system, thus interfering with and faulting the transmission of information. Any 'meaning' then attributed to an antigen is dependent on the kind of immune response it generates and one which, as proposed in [8], operates via the molecular structure in some accordance with the Shannon–theoretic principles of information. Consequently, the system has several options in responding to an antigen: a finely tuned cognitive system organizes the information as it is induced by the antigen and devises the 'format' for internal processing and release into the biochemical language of the immune system.

One instance is where the system engages different response cytotoxic T–cells, where 'helper' T–cells secrete mixtures of cytokines while lymphocytes navigate several cell types. Subject–predicate type of communication occurs when an antigen cell communicates an immune sentence to a T–cell which is unable to recognize the antigen totally. Thus the T–cell antigen receptor (TCR) requires the antigen to register with a superficial major histocompatibility complex (MHC) whereby a peptide functions as the 'subject' (of the immune sentence), and the way in which the T–cell responds to the peptide in the MHC is said to define the *meaning* of the antigen. The various response/non–response options are germline predicate signals comprising of cell–interaction/adhesion molecules. The predicate signals assess the antigen–presenting cells (APC) and tissues, thus registering the potential threat posed by the antigen. The T–cells read off the context of the antigen subject through integration of the germline signals with antigen recognition and react accordingly by dispatching a team of cytokines and other molecules. In their overall function, they resemble typical cognitive processes – the response of the immune system to an antigen so reflects a function of the entire community (network). This is the essence of the immuno–cognitive function: assimilating a perceived signal with respect to a learned association with the environment, and then upon comparison, initiating a select response mechanism from a large repertoire of possibilities. As incorporated into higher animals, the immune system becomes patently a deeper cognitive organism due to the increased complexity of factors of social mechanisms and environmental management each exerting their characteristic cross–talk and tendency to noise. In certain respects, the breakdown of a given immune system so results from a disorder in transmission of information.

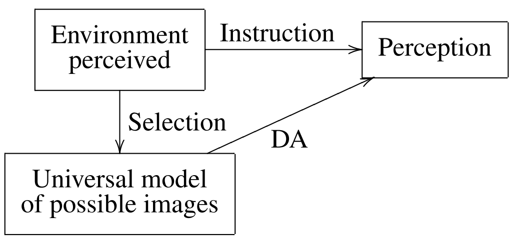
A possible scenario tied to the paradigm of [8] describes the activation of T–cell development and the immunological synapse via adhesion of the T–cell and APC which in theory could be initiated by the peptide MHC [40]. The viewpoint of [40] is that, in practice, the process of signaling response is influenced by certain classes of integrins in which the actin cytoskeleton provides a suitable structural mechanism for assimilating the signaling input. Since the immunological synapse is sensitive to the overall quality of the MHC peptide, the formation of this synapse depends upon the T–cell surface and the actin /myosin cytoskeletal systems in composing a cellular structure out of the transient interaction of the TCR and peptide MHC.

The ‘Collective Efficacy’ of [110] is one source of analogies between immune cognition and socio–environmental neighborhood cognition. Similarly, the diagram below based on [48](§3.3 Fig. 8) shows a mapping from immune recognition schemes in terms of a *determination through universals*



(9.1)

where ‘DR’ is short for ‘differential reproduction’. Associated to this interpretation via immunology, the corresponding ‘cognitive’ interpretation may be represented by [48](§3.4 Fig. 11):



(9.2)

where ‘DA’ is short for ‘differential amplification’.

9.2 Comparisons with neural networks

The underlying processes of institutional/directed cognition and intelligent machine operation can in part be compared with the functioning of neural networks, thalamacortical and olfactory systems, as examples. Recall that the acclaimed Hodgkin–Huxley model, together with several allied models provide a descriptive base for studying a variety of neuronal cell complexes in which informational patterns can be analyzed on codes based on the temporal properties of impulses: statistical intervals, frequencies, amplitude and phase variation. Accordingly, operative functions that will determine the number of possibilities, depend mainly on the statistical structure of the information sources and the specific nature of the codes in question. In a related setting, the theory of differentiable dynamics is applicable for modeling the effects of neuronal activity (such as spiking and bursting) in terms of homoclinic/periodic orbits in relationship to stable (or unstable) manifolds of critical elements, saddle node–bifurcations, hyperbolic sets, and the applications of the major theorems of Smale and others (see e.g. [52, 72]). In this respect, rate distortion manifolds are suitable models for analysis of such concepts while at the same time they afford the special features of adhering to the Shannon–McMillan theorems. We recall the Poincaré first return map relative to the phase portraits (see e.g. [72]) that originally led to the holonomy concept, an essential descriptive mechanism of neural and cognitive transition states as we have pointed out in the context of groupoids.

Rather than by individual cells, quanta of information can be considered as encoded by communities of the former. Typically, place cells are representative of encoding information within an environmental frame of reference whereupon a quorum of cells responds to the demands within a given location. Each constituent putatively breaks down its response in terms of an average, plus a variation in noise (neurons can be typically noisy and in turn can cause noisy synaptic inputs, oftentimes impeding transmission relay) thus contributing to sequences of spiking, in turn encoding information within the period of stimulus. Eventually, there results an overall cumulative response to the environment in relationship to the direction of motion, color, shape, form etc. as they are encoded into the appropriate regions of the visual cortex.

Recall that in the pioneering work on holography (and later wavelets) Gabor [56] postulated an ‘uncertainty’ – a quantum of information corresponding to a limit to which both frequency modulations and spatial information can be simultaneously measured. Pribram [107] in the context of neural networks and brain transition states, developed analogous ideas of holography/uncertainty, to some extent based upon the Gabor theory. Within neuronal systems, dendritic–processing employs analogous uncertainty principles in order to optimize the relay of information by micro–processing. Both time and spectral information (frequencies) are considered as stored in the brain which supposedly maintains a process of self–organization in order to minimize the uncertainty through a wide–scale regulatory system of phase transitions, the origin of which involves the various computational neuroscientific mechanisms of (hyper) polarizing action potentials, spiking, bursting and phase–locking, etc. These contribute to a multitude of network cells that register and react to an incoming perceptive signal. Thus it is claimed that cognitive processes up to consciousness may emerge from the neural level, but this emergence necessitates the integration of lower levels evolving from the successive cultural complexifications through phase transitions within a hierarchy of which the model of a colimits structured MES is one such example.

It is understood that the maintenance of a cell membrane potential depends not only upon inter–cellular communication, but also upon spiking and bursting rates: usually fast K^+ for transmission between nerve and muscle cells, and slow Ca^{2+} for contraction of muscle fibers [58]. Periodic inputs give rise to spike trains, but stochastic resonance through noise is needed in order to surpass a threshold for an action potential [57]. Granted a noisy environment, one expects a suitable noise level for the maximum signal transmission in correlation to the rate distortion theorem. It is pertinent to the question of neuronal computation by population coding, gating and phase–locking in the presence of stochastic resonance; altogether a different informational scenario to the language/immunology of cellular systems where maximum likelihood methods can involve substantial data accumulation leading to implement an 'electoral system' for predicting vectors by regarding the activity of a given cell as a vote for taking a preferred direction [104] and thus initiating its cognitive response, quite in tune with the Atlan–Cohen model.

9.3 The thalamacortical model

In an analogous way, the viewpoint of [12], as we have mentioned, is to regard the nervous system as a collection of specialized unconscious processors complete with its own squad of perceptual analyzers, output systems, etc. These are considered as performing cooperatively and efficiently within their locale, but since the system is characteristically decentralized, such qualities may not automatically function at a 'global' level. Thus within the system (or community) the interaction, control and coordination of squads of unconscious specialists depend on a central information exchange somewhat like a typical broadcasting system (such as the Global Workspace). Take for instance a cognitive basis for emotion, complete with its own language/grammar/syntax as a framework for individual adjustment to a challenging psycho–social environment and a mechanism for implementing various response categories towards the latter [125] – a further slant on the Atlan–Cohen perspective. Whereas some functions can be performed habitually, special operations require a combined team effort, the strategy and implementation of which is somehow relayed throughout the environment/community (cf §9.1). It appears to be a characteristic ubiquitous to a number of commonly studied neuro–cognitive and immuno–biological models. For instance, the neurobiological hypothesis of [13] is that intralaminar nuclei as a subset of the thalamus comprises the broadcasting network for the Global (Neuronal) Workspace. A main premise is that the reticular nucleus of the thalamus is instrumental for gating attention in an information–theoretic capacity and thus constitutes an agent towards consciousness.

In many regions within the various cortical zones, neuronal groups from one zone can arouse those in another so to produce a relatively organized re–projection of signals back to the former, thus creating a network of reverberating loops as are realized in the hippocampus, the olfactory system and cortical–thalamus. A riding assumption is that there is a certain synchronization of neurons through resonance and periodic oscillations of the neighboring population activity. Let us dwell on a particular scenario. Suppose $\mathfrak{X}$ and $\mathfrak{Y}$ denote surfaces consisting of neurons and receptor cells respectively, and let $f : \mathfrak{X} \longrightarrow \mathfrak{Y}$ be a mapping of points of $\mathfrak{X}$ to assigned points of $\mathfrak{Y}$ under f.

In the *maps/re–entry* model [42, 43, 44], such a mapping should be considered as a component of the cerebral anatomy which is equipped and genetically coded with such mapping networks, as for instance, the operational part of the visual cortex. Re–entry is a selective process whereby a multitude of neuronal groups interact rapidly by two–way signaling (reciprocity) where parallel signals are inter–relayed between maps; take for instance the field of reverberating/signaling cycles active within the thalamocortical meshwork. *A priori*, such a process is not a feedback system since there are many parallel streams operating simultaneously and re–entry channels serve to link, in a sense, the compositions $f_1 \circ f_2 \circ f_3 \cdots$ of distinct maps. In general these mappings are defined locally throughout, where a global mapping can be considered as defining a perceptual category. The maps/re–entry processes comprise a representational schemata for external stimuli on the nervous system, ensuring the context dependence of local synaptic dynamics at the same time mediating conflicting signals. Thus re–entrant channels between hierarchial levels of cortical regions assist the synchronous orchestration of neural processes. Impediments and general malfunctioning of information in the re–entry processes (possibly due to some biochemical imbalance) may then be part explanation for various mental disorders such as depression and schizophrenia. The association of short–term memory tied to consciousness within an architecture of thalamocortical reverberatory loops is proposed in [33]. Further support for the thalamacortical model as an essential component of the Workspace is provided in [38] in the context of a neuronal basis for inattentional blindness, the cognitive malfunction we had described earlier. From our perspective, the nested sequences of rate distortion manifolds considered in §6 and the processing via groupoids in (8.14) as descriptive mechanisms for such interactions, hence appear strikingly relevant.

The efficiency of re–entry is dependent on widespread variation in strength of connection, orientation and the potential convergenece/divergence of paths in the rate distortion sense. Suggestive of the 2–categories interpretation of the Jamesian sense of consciousness through processes of processes, *the dynamic core hypothesis* [44] concerns the strength–framework of neural interactions within a functional cluster, mainly prevalent in the thalamorcortical meshwork. A point here is that the dynamic core defines a neuronal state space (space of objects) and paths connecting points in this space represent a sequence of conscious states over time. We suggest that morphisms between the paths themselves should be admissible. Information relayed to the Workspace is proposed in [13] from the intralaminar nuclei comprising certain collections of thalamic regions. The reticular nucleus of the thalamus is considered in [13] as instrumental in gating attention. Under the premise that an orchestrated thalamus is a key component towards consciousness, the reticular nucleus is one leading factor to which a network–theoretic analysis/ rate distortion theory seems applicable.

Our discussion of groupoids concerning the reciprocity in relay of signaling (invertibility) in such networks, is a motivation for representing neuronal (groups) clusters by an appropriate categorical–algebraic structure (much weaker than the conventional notion of a 'group' in a mathematical sense). Such categorical representations in the terminology of [46] are called 'categorical neurons' (or *cat–neurons* for short). Consciousness loops [43], the Global Neuronal Workspace of [12] are among an assortment of models that have such a categorical representation. Among other things, there is proposed several criteria for studying the binding problem via the overall integration of neuronal assemblies and concepts such as *the archetypal core*: the cat–neuron resonates as an echo propagated to target concepts through series of thalamocortical loops. Analogous to how neurons communicate mainly through synaptic networks, cat–neurons interact in accordance with certain linking procedures and can be studied in the context of categorical logic which in turn may be applied to semantic modeling for neural networks [66, 67]. In this respect (neuro)groupoids with their invertibility property for all morphisms, provide the descriptive sub–mechanisms for reciprocity within the constituent assemblies.

9.4 Autopoietic systems and Distributed Cognition

The viewpoint of Maturana and Varela [93], as supported by several accounts in this paper, is that cognition is fundamentally a biological process and that living systems inhabit a cognitive domain through the autopoiesis of structurally coupled unitary (self-reproducing) systems that influence the organization and maintenance between both themselves and their environment over time. Many types of systems, be they biological or social, are realized through the autopoiesis of their various components and the totality of their interactive relations forming a medium in which these components realize their ontogeny. If anything, this may simply be for the sake of getting their survival mechanisms straight. It is through participation alone that an autopoietic system determines a social system by realizing the relations that are characteristic of that system, and it is reasonable to view their 'cellular' models as described in terms of the information spaces we have considered. The 'cellular' organization of cognition adjusts and adapts to the ever-changing thermodynamic phase transitions of the environment and subsequent levels of complexity; accordingly, the latter induces by reciprocation a re-adjustment within the former. Davia [34] suggests defining the range of thermodynamic conditions in which an organism can mediate transitions as a catalyst to be its "environmental survival space".

The descriptive and causal notions which can be described in terms of our groupoid (and other categorical) structures may be guided by the following principles [93] (Chapter III):

(1) Relations of constitution that determine the components produced constitute the topology in which the autopoiesis is realized.

(2) Relations of specificity that determine that the components produced be the specific ones defined by their participation in the autopoiesis.

(3) Relations of order that determine that the concatenation of the components in the relations of specification, constitution and order be the ones specified by the autopoiesis.

In this respect, concepts such as the Atlan–Cohen model and Institutional Cognition would appear to have partial overlap with autopoietic systems whereby the dynamics of their constituent cognitive cells can be modeled in terms of rate distortion manifolds as component representations of the Global Workspace architecture. Davia [34] argues that the concept of a soliton wave is ubiquitous to representing an autpoietic self–sustaining dynamic process. It is interesting to hypothesize that such wavefronts permeating through a given cognitive cell may actually be represented by a 'foliation' on a corresponding rate distortion manifold (such as a 'noise flow' as was discussed earlier).

9.5 Distributed and Institutional cognition

Closely related are systems of *Distributed Cognition* [70] (as discussed in [134] which dynamically inter–arrange and marshal their subsystems for task–implementation within the broader context of cognitive ethnography. In a similar way to autopoietic transformations, distributed cognition applies not only in relationship to individual human cognition, but extends to the broader institutional/machine–based cognition where humans undertake the task of controlling and navigating through multi–tasking machine worlds, implementing policies, etc. while embedded in the ambient 'memetic' environment of that culture. Hollan et al. [70] exemplify task–oriented activities of tightly–knitted groups in relationship to their working environment, and address the social organization and structure of activity that induces an information flow as part of the cognitive process necessary for the completion of a given operation. This may entail certain perceptual inferences within an evolved 'conceptual space' of the tasking environment (such as realized in the handling of digitally regulated flight instrumentation [70]). Concerning the interactions between distributed cognition, ethnography, experiment, work place and work materials, Hollan et al. [70] identify several widely applicable core principles such as:

– people establish and coordinate different types of structure in their environment
– it takes effort to maintain coordination
– people off–load cognitive effort to the environment whenever practical
– there are improved dynamics of cognitive load–balancing available in social organization.

The 'culture' of oceanic navigation, such as described in [74], with its exclusive range of techniques of measurement, skills, etc., itself becomes a cognitive process. This way of thinking about how such computational mechanisms are essentially cognitive, is discussed at philosophical/complexity levels in [119].

Somewhat related to the apparent corporate teamwork of distributed cognition are other information oriented systems of cognitive interaction. These may be viewed in a dynamical systems context which incorporates 'embodiment' within the context of cultural, linguistic factors, physical motion, and so on [30]. One such example is that of *social prosthetic systems* [83] which describe how deficiencies in individual (cognitive) capacity can be compensated via participation with the brain–fusion of socio–environmental networks. The argument is based on how supposed "selfish" genetic programming, aware of limitations on information handling, motivates reaching into the environment to attain to conceptual management within the latter. Loosely speaking, the brain uses the world and "enduring relationships" as extensions of itself [83].

Once such systems can be represented by their corresponding equivalence classes, configurations of interacting groupoids can be realized for which the discussion of §7.4 has relevant applicability. These can be compared with, and applied to, the network analysis and geodesic subgraph evolution via 'small world' partnerships as discussed in e.g. [2, 109, 142] and where 'weak ties' permit the formation of Global Workspaces and inter–communication between them [63, 135]. In this respect, the underlying graph of a groupoid and the concept of a groupoid atlas may well become essential techniques for delving further into the descriptive mechanisms of such systems (see [60]). Whereas the disciplines of neurophysiology/biology provide some explanation to the underlying mechanisms of human consciousness (but often curbed by the strictures of the 'mereological fallacy' [17]), it is of growing importance to further study the interactive–reciprocity of the individual body/brain with the environment, as in the way autopoietic and social prosthetic systems profess to do. Likewise, some brave new world of consciousness machines will interact with their embedding systems thus creating new strains of epidemics and cognitive failures [136].

One may also consider how related social factors on a more global scale can physically determine and shape the environment created through the cognitive mechanisms of its inhabitants. Within a framework of spatial syntax and information, this has been addressed in [69]. Such factors lead to multifarious forms of development (and those quite clearly tied to the influence of institutional cognition) and are manifest at many levels. For instance, we have the concepts of 'street', 'terrace', 'lane' and how these civil structures eventually do shape the physical appearance, the cultural character and ethos of a city while reflecting its order of wealth, industry, affluence, ethnic divisions, and so on. As much as this development might once have been viewed as positive over decades, centuries even, inevitably several 'institutional cognitive modes' that assisted the creation of the city in the first place, often are destined to go into reverse gear. Consequently, the features of urban atrophy begin to set in: derelict housing, the demise of public services (health, transport, education, etc.), planned shrinkage and an upsurge in societal epidemics (HIV, AIDS, obesity, depression, tuberculosis, etc.). A Markov game thus unfolds between city and suburbs [135, 137].

The 'wrench in the works' of social networking as foreseen by [63] is often the cause of certain epidemics as a recent report [29] on obesity suggests: from the embedding in a network, 'social distance', friendship (perhaps more so than within a family) and the network tolerance towards obesity appear at least as influential as heredity factors (such as an under–active thyroid gland). One might also argue that obesity is one of several epidemics realized at the negative end of social prosthetic systems, as much as toxic waste is to some 'thriving' chemical industry somewhere on the planet.

9.6 Red Queen versus the Pentagon Ratchet

Lewis Carroll's 'Red Queen' has been taken metaphorically to describe an evolutionary system which "keeps running" in order to co–evolve with ambient competing systems. The analogy seems to be ubiquitous to modes of institutional cognition, economic game theory, arms races and predator–prey type models where advanced capabilities in one system are aimed to decrease those in the other. For instance, how a slowly evolving cognitive system has to gear itself to the constant threat of infectious epidemics. In a similar way it can be viewed as a contest between the internal cultures of a system (corporation, whatever) on one hand, and associated external technologies, policies, legislation, outsourcing etc., on the other. As discussed in [135, 136] network (giant component) analysis reveals sudden and sharp phase transitions on passing critical points thus forecasting the most efficient co–evolutionary structure as gaining competitive strength. The Red Queen influences a multiple Workspace environment (e.g. one that is socio–economic, institutional or directed cognitive) and simultaneously interacts with a powerful mutual crosstalk creating a 'ratchet–down' effect. The latter has been coined the 'Pentagon Ratchet' [135, 136] a term suggestive of the legislative (re)allocation of major resources from the civilian into the military sectors. Likewise, the language of large-scale cognitive systems within an interactive environment may undergo a phase transition induced by intense crosstalk in reversal of their evolution. Thus the Ratchet gradually breaks down the competitive function of the Red Queen hence causing its sectors to become fragmented or to disintegrate altogether. Techniques involve the critical manifolds of differential game theory and explicit examples of renormalization modes leading to embeddings into state spaces, are exhibited in [135](§4.3) and [136](Chapter 4). Let us remark that the corresponding social network of the game may develop towards a 'small worlds' situation, and as previously mentioned, affording an enrichment of topological and statistical properties within a graph theoretic interpretation (cf [60, 120]).

The Red Queen (RQ) and Pentagon Ratchet (PR) are deemed to be interacting 'principal environments' for each other. In more general mathematical terms, let $\mathsf{G}_{\mathrm{RQ}}, \mathsf{G}_{\mathrm{PR}}$ denote the corresponding groupoids of path equivalences and $M_{\mathrm{RQ}}, M_{\mathrm{PR}}$ denote their respective set of objects ('acquired characteristics'). Then in the competition we may regard the groupoids as acting on each other's set of objects via crosstalk

$$\begin{aligned} \mathsf{G}_{\mathrm{PR}} \curvearrowright M_{\mathrm{RQ}} \longrightarrow M_{\mathrm{RQ}} \\ \mathsf{G}_{\mathrm{RQ}} \curvearrowright M_{\mathrm{PR}} \longrightarrow M_{\mathrm{PR}} \end{aligned} \tag{9.3}$$

thus yielding orbit spaces of generally lower dimension which, for instance, symbolize the curtailment/policy effect of one upon the other (the notion of a *groupoid action* is made specific in Appendix I §11.1). In view of the fundamental homology with 'thermodynamic' processes, the RQ along with small world networks are exemplified by certain distributed and institutional cognitive systems in [60].

There are several evolutionary scenarios tied to RQs and rate distortion theory that deserve mention. One such concerns a proposal by Eigen [47] of an evolutionary model which involves selections as a condensation in an information space. Some complications arise regarding the matter of genetic complexity since information has to be encoded in longer gene sequences by using replication with optimal fidelity. However, in order to do this, it is necessary to have a complex replication enzyme which just happens to be elusive, since such an enzyme will itself require a longer gene and the latter would violate an error threshold [71]. With the aim of resolving this paradox [141] employ a rate distortion argument coupled with a RQ coevolutionary rachet toward establishing an evolutionary condensation that results in an effective error–correction mechanism. We refer to [141] for complete details.

9.7 Optimal coding and physiology: examples

For most species, and whether for predator or prey in particular, interaural time difference (ITD) is a characteristic property geared to localizing sound sources as crucial to the survival mechanism. Case studies have revealed optimal coding strategies depend not only on sound frequency ranges, but also on evolutionary driven physiological factors such as cranial size and form. Within groups of coincidence–detector neurons encoding ITD, each constituent member may be tuned for ITD in relationship to the ambient physiological range whereby exact tuning is determined by a time interval of axonal conduction in the auditory system. For pure tones, there is for each ITD an interaural phase difference (IPD) whereby an optimal coding strategy is seen to depend significantly on the relative width of the physiological range of individual IPDs in comparison to their corresponding tuning curves [65]. At the same time, we expect such strategies are significantly influenced by the behavioral patterns of the environmental stimuli; in this respect (auditory) receptors attain to optimal rather than average performance for most survival purposes [90].

Such systems are expected not to be free of corruption by noise. On the other hand, we have noted that noise, particularly in the case of 'population' based phenomena', can engender a stochastic resonance which may favorably enhance and/or optimize the transmission of a weak signal via sensorimotor integration as shown, for instance, in certain cognitive studies of controlled visual stimulation [80] or 'randomly enhanced' human gaming strategies [144]. In [114] is considered the response of a neuron (in relationship to the cat primary visual cortex) on the linear filtering of the stimulus (luminance) values S by a linear receptive field L over space–time. In the usual network setting, a groupoid structure G can be revealed and the convolution $L \star S$ defined accordingly thus leading to a convolution algebra $\mathcal{C}(\mathsf{G})$ over a suitable class (of continuous) function on G. The output of the filter is then passed through a nonlinear function $h : \mathcal{C}(\mathsf{G}) \longrightarrow B$ such that the neurons response $R(t)$ is specified by $h(L \star S)$. Similar principles may be applied for explaining how activation receptors on registering a certain stimulus, transmit pulsations to the sensory cortex and assimilate the resulting meshwork of convoluted signals [54].

10 Conclusion

In this paper we have described a structural framework upon which rate distortion manifolds as representing certain cognitive modules, can be constructed by a variety of state-of-the-art mathematical concepts. We expect that implementing these concepts will lead to more exact, conceptually–centered, information–based models of cognition-at-large. The associated techniques as we have presented them, provide a method for the construction of a variety of information spaces structured by the Shannon coding and rate distortion theorems besides the means of describing globalization through local procedures. We have shown that the flexible, less rigid structures afforded to us by the notion of an atlas-manifold topology (or more generally, a groupoid atlas) along with simplicial/graph theoretic methods, can be adapted to a wide range of cognitive situations operative within the Global Workspace. This affords greater elegance and meaning to how these processes can be modeled without recourse to the traditional rigid, data–driven techniques which quite often can obscure some deeper underlying meaning.

In several instances we have employed the groupoid method as a category theory technique that allows one to reduce a vast labyrinthine configuration of networks to their corresponding sets of equivalence classes. The latter are computationally more user–friendly and create their own kind of dynamical systems via groupoid actions, (path) holonomy, etc. In particular, we should observe that the techniques we have outlined in the manifold/groupoid setting, are those suited to the description of 'local-to-global processing', seen for instance in the case of scheduling of paths and in the construction of the holonomy groupoid. In this way, the dynamics of cognitive processes (particularly those of the distributed and institutional type) can be aptly encoded in terms of groupoid actions as revealed, for instance, in the coevolutionary contest between the Red Queen and the Pentagon Rachet. Likewise, we have seen how symmetry breaking of the network groupoid of linked cognitive modules cultivates a giant component which eventually emerges as a phase transition. In this respect, the fundamental homology describes close analogies between evolutionary modes (e.g. punctuated equilibria) influencing most cognitive processes, and the underlying dynamics of certain statistical–physical systems; more specifically, how alterations in the information network topology can induce phase–transitional states.

The geometry/topology of a rate distortion manifold thus represents the shape and form of information flow with respect to its syntatic–semantic content within the cultural environment of the particular Workspace through which it passes. In so far that the message transmits the channel, the former may be susceptible to cultural and evolutionary impingement. In a related way, a computational scheme of a cognitive process, may itself be deemed as a form of cognition. This leads us to questions of 'higher categorical' cognition rather befitting the 'processes of processes' as was alluded to earlier–clearly a matter that warrants further investigation.

Acknowledgement

We wish to thank the Overview Editor, Dr. Maria Ganzha, for her consideration of this paper.

11 Appendix I: Groupoids and their actions

Since groupoids and related actions have been pinpointed in the text, we provide the basic definitions and refer to [21, 31, 92, 143] for further detils. Recall that a *groupoid* G is, loosely speaking, a small category with inverses over its set of objects $\mathrm{Ob}(\mathsf{G})$. More specifically :

A groupoid consists of a set G with a distinguished subset denoted $\mathsf{G}^{(0)} = \mathrm{Ob}(\mathsf{G}) \subset \mathsf{G}$, called *the set of objects* of G, together with maps

$$r, s : \ \mathsf{G} \underset{s}{\overset{r}{\rightrightarrows}} \mathsf{G}^{(0)} \tag{11.1}$$

called the *range* and *source maps* respectively, together with a law of composition

$$\begin{aligned} \circ : \mathsf{G}^{(2)} = \{(\gamma_1, \gamma_2) \in \mathsf{G} \times \mathsf{G} : s(\gamma_1) = r(\gamma_2)\} \\ \longrightarrow \mathsf{G} \end{aligned} \tag{11.2}$$

on the set of 'arrows' $\mathsf{G}^{(2)}$, such that the following hold:

(1) $s(\gamma_1 \circ \gamma_2) = s(\gamma_2)$, $r(\gamma_1 \circ \gamma_2) = r(\gamma_1)$, for all $(\gamma_1, \gamma_2) \in \mathsf{G}^{(2)}$.

(2) $s(x) = r(x) = x$, for all $x \in \mathsf{G}^{(0)}$.

(3) $\gamma \circ s(\gamma) = \gamma$, $r(\gamma) \circ \gamma = \gamma$, for all $\gamma \in \mathsf{G}$.

(4) $(\gamma_1 \circ \gamma_2) \circ \gamma_3 = \gamma_1 \circ (\gamma_2 \circ \gamma_3)$.

(5) Each γ has a two–sided inverse γ^{-1} with $\gamma\gamma^{-1} = r(\gamma)$, $\gamma^{-1}\gamma = s(\gamma)$.

Often one denotes by $\mathsf{G}_x^y = s^{-1}(x) \cap r^{-1}(y)$ the set of morphisms in G from x to y, and G_x^x denotes the isotropy group at $x \in \mathsf{G}^{(0)}$.

Example 11.1. An equivalence relation $\mathcal{R}$ on a set X can constitute a groupoid in the following way. Specifically, $\mathcal{R} \subset X \times X$ is identifiable with the set of ordered pairs (x, y) satisfying $x\mathcal{R}y$, whereby the morphisms are

$$\mathcal{R}_y^x = \begin{cases} \{(x, y)\} \ , \text{ if } x\mathcal{R}y, \\ 0, \text{ otherwise.} \end{cases} \tag{11.3}$$

The composition is given by

$$\begin{aligned} \circ : \mathcal{R}_y^x \times \mathcal{R}_z^y \longrightarrow \mathcal{R}_z^x, \\ (x, y) \circ (y, z) = (x, z), \end{aligned} \tag{11.4}$$

where (x, x) is the identity and $(x, y)^{-1} = (y, x)$. Accordingly, the orbit $\mathcal{R}(x)$ is the equivalence class of $x \in X$.

Conversely, a groupoid G may induce an equivalence relation $\mathcal{R}$ on the set X, for which the equivalence classes $\mathcal{R}(x)$ are the orbits $\mathsf{G}(x)$, for all x in X. This is subject to a forgetful functor $F : \mathsf{G} \longrightarrow \mathcal{R}$, such that $F(g) = (y, x)$ if and only if $g \in \mathsf{G}_x^y$.

Example 11.2. Clearly, any group is a groupoid whereby the object set consists of the single element $\{e\}$ *the identity* (e.g the fundamental group $\pi_1(M)$ of a manifold M, is a groupoid). Thus groupoids may be seen as consisting of 'multiple identities'. Indeed, any manifold M can be viewed as a groupoid over itself where all morphisms are units (that is, the arrow set of M is M itself). We also have the *pair groupoid* $M \times M \rightrightarrows M$ where the natural projections from each factor comprise the range and source maps.

11.1 Groupoid actions

Let X be a topological space admitting an action '$\curvearrowright$' of a group G. Specifically $\curvearrowright$: $X \times G \longrightarrow X$, with $\curvearrowright (x, g) = xg$ and $x(g_1 g_2) = (xg_1)g_2$, for all $x \in X$ and $g \in G$. Here we have a natural groupoid $\mathsf{G} = X \times G$ with $\mathsf{G}^{(0)} = X \times \{\mathbf{1}\}$, and for which the following conditions hold:

(1) $r(x, g) = x,\ s(x, g) = xg$, for all $(x, g) \in X \times G$.

(2) $(x, g_1)(y, g_2) = (x, g_1 g_2)$ if $xg_1 = y$.

(3) $(x, g)^{-1} = (xg, g^{-1})$, for all $(x, g) \in X \times G$.

Consider a groupoid $\mathsf{G} \rightrightarrows B$ over its set of objects $B = \mathsf{G}^{(0)}$. Let M be a topological space and $f : M \longrightarrow B$ a continuous map. Consider the set

$$\begin{aligned} \mathsf{G} \curvearrowright M = \{\ (g, u) \in \mathsf{G} \times M : sg = f(u)\ \} \\ \subset \mathsf{G} \times M. \end{aligned} \tag{11.5}$$

An *action* of G on (M, f, B) is a continuous map $\mathsf{G} \times M \longrightarrow M$, given by $(g, u) \mapsto gu$ satisfying:

(1) $f(gu) = rg$, for all $(g, u) \in \mathsf{G} \curvearrowright M$.

(2) $h(gu) = (hg)u$, for all $(h, g) \in \mathsf{G} \times \mathsf{G},\ (g, u) \in \mathsf{G} \curvearrowright M$.

(3) $\widetilde{f(u)}u = u$, for all $u \in M$, where $\sim$ denotes the corresponding groupoid isomorphism.

We call $\mathsf{G} \curvearrowright M \longrightarrow M$ the *action groupoid*. For $u \in M$, the subset $\mathsf{G}[u] = \{gu : g \in \mathsf{G}\}$, is the *orbit* of u under G . These concepts generalize the notion of a group action on a topological space.

11.2 Proper groupoids and orbifolds

Firstly, G is said to be a *Lie groupoid* when $\mathsf{G}^{(0)}$ and $\mathsf{G}^{(2)}$ have the structures of differentiable (Hausdorff) manifolds, the map s is a differentiable submersion (with Hausdorff fibers), and all other structure maps are differentiable. A Lie groupoid is said to be *proper* if it is Hausdorff and the map $(s, r) : \mathsf{G}^{(2)} \longrightarrow \mathsf{G}^{(0)} \times \mathsf{G}^{(0)}$ is proper (that is, each inverse image of a compact subset is compact).

An orbifold atlas of dimension n of a topological space Q is a collection of pairwise compatible orbifold charts

$$\mathcal{U} = \{(U_i, G_i, \phi_i)\}_{i \in \mathcal{I}}, \tag{11.6}$$

of dimension n on Q, where the $G_i \subset \text{Diff}(U_i)$ are finite subgroups, such that $\bigcup_{i \in \mathcal{I}} \phi_i(U_i) = Q$. Two orbifold atlases of Q are equivalent if their union is an orbifold atlas. Then an *orbifold* of dimension n is a pair $(Q, \mathcal{U})$ when Q is a (second countable) Hausdorff topological space and $\mathcal{U}$ is a maximal orbifold atlas of dimension n of Q. For further details see [96]. In particular, there is an associated pseudogroup of transitions $\Psi(\mathcal{U})$ and an effective proper groupoid $\Gamma(\mathcal{U}) = \Gamma(\Psi(\mathcal{U}))$ associated to $\Psi(\mathcal{U})$ (see [96] §5.6).

12 Appendix II: Briefly simplicial complexes and triangulations

Let K be a simplicial complex, that is, K contains a set of objects $V(K)$ called *vertices* and a set of non–empty subsets of $V(K)$ called *simplices*. If $\sigma \subset V(K)$ is a given simplex and $\kappa \subset \sigma,\ \kappa \neq \emptyset$, then κ is also a simplex. The geometric realization (or *polyhedron*) of K, denoted $|K|$, comprises the set of all functions $V(K) \longrightarrow [0, 1]$, such that:

1. If $\alpha \in |K|$, the set $\{v \in V(K) : \alpha(v) \neq 0\}$ is a simplex of K.
2. $\sum_{v \in V(K)} \alpha(v) = 1$.

If $s \in K$, we let $|s|$ denote the set

$$|s| = \{\alpha \in |K| : \alpha(v) \neq 0 \Rightarrow v \in s\}, \tag{12.1}$$

and

$$\langle s \rangle = \{\alpha \in |K| : \alpha(v) \neq 0 \Leftrightarrow v \in s\}. \tag{12.2}$$

We call $\alpha(v)$ the *v–th barycentric coordinate of* α and $p_V(\alpha) = \alpha(v) : |K| \longrightarrow [0, 1]$ is the corresponding v–th barycentric projection. A metric $d(\ ,\ )$ can be defined on $|K|$ by setting

$$d(\alpha, \beta) = \Big[\sum_{v \in V(K)} (p_v(\alpha) - p_v(\beta))^2 \Big]^{\frac{1}{2}}, \tag{12.3}$$

with the resulting initial topology for barycentric projections.

Often it is useful to pass to a subdivision of a given simplicial complex K. A *subdivision* of K is a simplicial complex K' satisfying:

a) the vertices of K' are identified as points of $|K|$;

b) if s' is a simplex of K', there exists a simplex s in K such that $s' \subset |s|$; and

c) the map $|K'|\longrightarrow|K|$ extending the map of vertices of K' to their corresponding points of $|K|$, is a homeomorphism.

Let K_p denote the set of p–simplices of K. If $\sigma = \{v_o, \ldots, v_p\} \in K_p$, then its *barycenter* $b(\sigma)$ is the point

$$b(\sigma) = \sum_{0\leq i\leq p} \frac{1}{p+1} v_i \in |K|. \tag{12.4}$$

Accordingly, we may speak of a *barycentric subdivision* $\widehat{K}$ of K as the simplicial complex with vertices the barycenters of the simplices of K, and whose simplices are finite non–empty collections of barycentres of simplices totally ordered by the face relations of K.

Suppose now $\mathcal{U} = \{U_\lambda : \lambda \in \Lambda\}$ is an open covering of M. We assign an ordering to the indexing set Λ of the cover and then let

$$\Lambda^{(n)} = \{(\lambda_0, \lambda_1, \ldots, \lambda_{n-1}) \in \Lambda : \lambda_0 \leqslant \lambda_1 \leqslant \ldots \leqslant \lambda_{n-1}\}. \tag{12.5}$$

The *nerve* $N\mathcal{U}$ of the cover $\mathcal{U}$ is defined as follows. Firstly, let

$$(N\mathcal{U})_n := \bigsqcup_{\nu\in\Lambda^{(n+1)}} U_\nu, \tag{12.6}$$

where $U_\nu = U_{\lambda_0} \cap \cdots \cap U_{\lambda_n}$. In this way an element of $(N\mathcal{U})_n$ consists of an $(n+2)$–tuple $(x, \lambda_0, \ldots, \lambda_n)$, where $x \in U_\nu$ and $\nu = (\lambda_0, \ldots, \lambda_n) \in \Lambda^{(n+1)}$. Then the nerve of $\mathcal{U}$ is given by $N\mathcal{U} := \lim_n (N\mathcal{U})_n$.

12.1 Triangulations

A *triangulation* (K, ϕ) of a space M means we have a simplicial complex K together with a homeomorphism $\phi : |K|\longrightarrow M$. For any vertex v in K, we define its *(open) star* by

$$\mathsf{st}(v) = \{\alpha \in |K| : \alpha(v) \neq 0\}. \tag{12.7}$$

Alternatively,

$$\mathsf{st}(v) = \bigcup\{\langle s\rangle : v \text{ is a vertex of } s\}, \tag{12.8}$$

that is, the union of interiors of all simplices having s as a vertex. Note that $\mathcal{U} = \{\mathsf{st}(v) : v \in K\}$ provides an open covering of $|K|$. References to these topics are [20, 76, 117].

13 Appendix III: Differentiable structures on path space

13.1 Plots and iterated integrals

Let us recall the state space E^Γ over the alphabet Γ. In general, we do not expect E^Γ to have a differentiable structure in the conventional sense of classical calculus, but one of several concepts of abstract 'differentiable spaces', might be applicable. One such structure uses an abstract notion of 'plots' [27], permitting a 'differentiable space' structure on E^Γ in terms of the following conditions. We consider a collection of maps $f : \mathbb{R}^n\longrightarrow E^\Gamma$ (where n can be arbitrarily large), called *plots*, such that:

1. If $f : \mathbb{R}^n\longrightarrow E^\Gamma$ is a plot, and if $g \in C^\infty(\mathbb{R}^m, \mathbb{R}^n)$ is a smooth map in the usual sense, then $f \circ g$ is a plot.
2. If $g_\alpha : \mathbb{R}^n\longrightarrow\mathbb{R}^n$ is a collection of embeddings whose images cover $\mathbb{R}^n$, and $f : \mathbb{R}^n\longrightarrow E^\Gamma$ is a map such that $f \circ g_\alpha$ is a plot, then f is also a plot.
3. Every map $f : \mathbb{R}^0\longrightarrow E^\Gamma$ is a plot.

Given another such differentiable space Y, a map $\psi : E^\Gamma\longrightarrow Y$ is said to be *differentiable* if for every plot f in E^Γ, $f \circ \psi$ is a plot in Y.

Relevant here is that the information (path) space $X = \mathcal{P}(E^\Gamma)$, supposedly as a length space, with its rate distortion measure s_X, could be assumed as endowed with a differentiable space structure as well. Effectively, we can view a path in E^Γ as a plot of the type $\gamma : I\longrightarrow E^\Gamma$. For every set map $\alpha : U\longrightarrow X = \mathcal{P}(E^\Gamma)$, there is a corresponding suspension map

$$\begin{aligned} \lambda_\alpha : I \times U &\longrightarrow E^\Gamma \\ (t, \xi) &\mapsto \alpha(\xi)(t). \end{aligned} \tag{13.1}$$

Then $X = \mathcal{P}(E^\Gamma)$ can be viewed as a differentiable space when assigned plots of the type $\alpha : U\longrightarrow X$, such that λ_α is a plot of E^Γ.

Suppose that $\mathcal{A}$ denotes some suitable (alphabetical or events) algebra and that E^Γ admits some choice of algebra $\mathcal{A}$–valued 1–forms $w_1, \ldots, w_r \in \Omega^1(E^\Gamma, \mathcal{A})$, then once given a path $\gamma : I\longrightarrow E^\Gamma$ of sufficiently differentiability, [27] introduces the notion of *iterated integrals*

$$\int_\gamma w_1, \ldots, w_r := \int_0^1 f_1 dt_1, \ldots, f_r dt_r, \tag{13.2}$$

where $f_i(t) = w_i(\gamma(t), \gamma'(t))$, or in terms of the pull–back, $\gamma^* w_i = f_i(t)dt$. Subsequently, this defines a map

$$\int_\gamma : (X, s_X)\longrightarrow\mathcal{A}. \tag{13.3}$$

This makes the same sense if we replace (X, s_X) by its canonical model (M, s_M):

$$\int_\gamma : (M, s_M)\longrightarrow\mathcal{A}. \tag{13.4}$$

Higher degree (differential) forms can be treated accordingly.

13.2 Fröhlicher spaces

The above notion of 'differentiability' via plots and iterated integrals is a relatively weak one that may be suited to rate distortion theory. There are other possibilities that provide an approach to calculus on spaces more 'pathological' than standard differentiable manifolds (such as to be found in e.g. [55, 81]). For instance in [55], a *Fröhlicher space* X consists of a triple $(X, \mathcal{C}_X, \mathcal{F}_X)$ where X is a set, $\mathcal{C}_X$ is a subset of all mappings $\mathbb{R} \longrightarrow X$, and $\mathcal{F}_X$ is the set of all functions $X \longrightarrow \mathbb{R}$, satisfying the properties:

(1) A function $f : X \longrightarrow \mathbb{R}$ belongs to $\mathcal{F}_X$ if and only if $f \circ c \in C^\infty(\mathbb{R}, \mathbb{R})$, for all $c \in \mathcal{C}_X$.

(2) A map $c : \mathbb{R} \longrightarrow X$ belongs to $\mathcal{C}_X$ if and only if $f \circ c \in C^\infty(\mathbb{R}, \mathbb{R})$, for all $f \in \mathcal{F}_X$.

A morphism of $\varphi : X \longrightarrow Y$ of Fröhlicher spaces X, Y is said to be *smooth* when the following three equivalent conditions hold:

(1) For each $c \in \mathcal{C}_X$, the composition $\varphi \circ c \in \mathcal{C}_Y$.

(2) For each $f \in \mathcal{F}_X$, the composition $f \circ \varphi \in \mathcal{F}_X$.

(3) For each $c \in \mathcal{C}_X$, and for each $f \in \mathcal{F}_X$, the composition $f \circ \varphi \circ c \in C^\infty(\mathbb{R}, \mathbb{R})$.

In short, the space X admits an admissible family of curves $\mathcal{C}_X$ together with an admissible family of functions $\mathcal{F}_X$ whose respective elements satisfy the above conditions.

References

[1] R. Abraham and J. E. Marsden (1978), *Foundations of Mechanics* (2nd Ed.) Benjamin/Cummings Publ., Reading MA.

[2] R. Albert and A.–L. Barabási (2002), Statistical mechanics of complex networks, *Rev. Mod. Phys.* 74, 47–97.

[3] D. Alonso, A. J. McKane and M. Pascual (2007), Stochastic amplification in epidemics, *J. R. Soc. Interface* 4, 575–582.

[4] M. E. – S. A. Aof and R. Brown (1992), The holonomy groupoid of a locally topological groupoid, *Top. Appl.* 47, 97–113.

[5] J. Anandan (1992), The geometric phase, *Nature* 360, 307–313.

[6] R. Ash (1990), *Information Theory*, Dover Publications, New York.

[7] H. Atlan (1972), *L'Organisation biologique et la théorie de l'information*, Hermann, Paris.

[8] H. Atlan and I. R. Cohen (1998), Immune information, self–organization and meaning, *International Immunology* 10, 711–717.

[9] H. Atlan (1979), *Entre le cristal et la fumée*, Seuil, Paris.

[10] H. Atmanspacher and R. G. Jahn (2003), Problems of reproducibility in complex mind matter systems, *J. of Scientific Exploration* 17(2), 243–270.

[11] R. Attal (2004), Combinatorics of non–abelian gerbes with connection and curvature, *Ann. Fond. Louis de Broglie* 29 (4), 609–633.

[12] B. J. Baars (1988), *A Cognitive Theory of Consciousness*, Cambridge University Press.

[13] B. J. Baars and J. Newman (1994), A neurobiological interpretation of the Global Workspace theory of consciousness, in *Consciousness in Philosophy and Cognitive Neuroscience* (A. Revonsuo and M. Kamppinen, eds.) Erlbaum, Hilsdale NJ.

[14] B. J. Baars and S. Franklin (2003), How conscious experience and working memory interact, *Trends in Cognitive Science* 7, 166–172.

[15] I. C. Baianu, R. Brown, G. Georgescu, and J. F. Glazebrook (2006), Complex nonlinear biodynamics in categories, higher dimensional algebra and Lukasiewicz–Moisil topos: transformations of neuronal, genetic and neoplastic networks, *Axiomathes* 16 Nos. 1–2, 65–122.

[16] A. Bak, R. Brown, G. Minian and T. Porter (2006), Global actions, groupoid atlases and related topics, *J. of Homotopy and Related Structures* 1, 1–54.

[17] M. Bennett and P. Hacker (2003), *Philosophical Foundations of Neuroscience*, Blackwell Publishing, London.

[18] E. D. Bolker, V. W. Guillemin and T. S. Holm, How is a graph like a manifold?, (to appear) http://arxiv:math.CO/0206103

[19] F. Borceux (1994), *Handbook of Categorical Algebra (1 and 2)– Encyclopedia of Mathematics and its Applications* 50 & 51, Cambridge University Press.

[20] R. Bott and L. W. Tu (1982), *Differential Forms in Algebraic Topology*, Grad. Texts in Math. 82, Springer Verlag.

[21] R. Brown (2006) *Topology and Groupoids*, BookSurge LLC.

[22] R. Brown, R. Paton and T. Porter (2004), Categorical language and hierarchical models for cell systems, in *Computation in Cells and Tissues - Perspectives and Tools of Thought*, Paton, R.; Bolouri, H.; Holcombe, M.; Parish, J.H.; Tateson, R. (Eds.) Natural Computing Series, Springer Verlag, 289–303.

[23] R. Brown and İ. İçen (2003), Towards a 2–dimensional notion of holonomy, *Advances in Math.* 178, 141–175.

[24] D. Burago, Y. Burago and S. Ivanov (2001), *A Course in Metric Geometry*, Grad. Studies in Math. 33, Amer. Math. Soc., Providence RI.

[25] D. J. Chalmers (1996), *The Conscious Mind–In search of a fundamental theory*, Oxford University Press.

[26] A. Candel and L. Conlon (2000), *Foliations I*, Graduate Studies in Mathematics 23, Amer. Math. Soc, Providence RI.

[27] K. T. Chen (1977), Iterated path integrals, *Bull. Amer. Math. Soc.* 83 No. 5, 831–879.

[28] D. Chigirev and W. Bialek (2004), Optimal manifold representation of data : an information theoretic approach, in *Advances in Neural Information Processing Systems* 16 161–168, MIT Press, Cambridge MA.

[29] N. A. Christakis and J. H. Fowler (2007), The spread of obesity in a large social network over 32 years, *New England J. Medicine* 357 (4), 370–379.

[30] A. Clark (1997), *Being there: Putting brain, body and world together again*, MIT Press, Cambridge MA.

[31] A. Connes (1994), *Noncommutative Geometry*, Academic Press.

[32] T. Cover and J. Thomas (1991), *Elements of Information Theory*, Wiley, New York.

[33] F. Crick and C. Koch (1990), Toward a neurobiological theory of consciousness, *The Neurosciences*, 263–275.

[34] C. J. Davia (2006), Life, catalysis and excitable media: A dynamic systems approach to metabolism and cognition, in (J. A. Tuszynski, ed.) *The Emerging Physics of Consciousness*, Springer Berlin–Heidelberg, 255–292.

[35] H. de Jong and M. Page (2000), Qualitative simulation of large and complex genetic regulatory systems, in *Proc. of Fourteenth European Conference on Artifical Intelligence, ECAI 2000*, W. Horn (ed.), IOS Press, Amsterdam, 141–145.

[36] A. Dembo and O. Zeitouni (1993), *Large Deviation Techniques and Applications*, Jones and Bartlett Publ. Boston–London.

[37] S. Dehaene and L. Naccache (2001), Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework, *Cognition* 79, 1–37.

[38] S. Dehaene and J.-P. Changeux (2005), Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness, *PloS Biology* 3 (5), 1–18.

[39] F. Dretske (1981), *Knowledge and the Flow of Information*, MIT Press, Cambridge MA.

[40] M. L. Dustin and J. A. Cooper (2000), The immunological synapse and the actin cytoskeleton: molecular hardware for T–cell signaling, *Nature Immunology* 1 No. 1, 23–29.

[41] E. Dyer and S. Eilenberg (1988), Globalizing fibrations by schedules, *Fund. Math.* 130, 125–136.

[42] G. Edelman (1992), *Brilliant Air, Brilliant Fire – On the Matter of the Mind*, Basic Books, New York.

[43] G. Edelman (1989), *The Remembered Present*, Basic Books, New York.

[44] G. Edelman and G. Tononi (2000), *A Universe of Consciousness*, Basic Books, New York.

[45] J. Eells and L. Lemaire (1983), *Selected topics in harmonic maps*, CBMS Regional Conf. Series, Amer. Math. Soc., Providence RI.

[46] A. C. Ehresmann and J.-P. Vanbremeersch (2007), *Memory Evolutive Systems: Hierarchy, Emergence, Cognition*, Elsevier.

[47] M. Eigen (1996), *Steps towards life: A perspective on evolution*, Oxford University Press, New York.

[48] D. Ellerman (2007), Adjoints and emergence: applications of a new theory of adjoint functors, *Axiomathes* 17, 19–39.

[49] M. Emery (1989), *Stochastic Calculus in Manifolds*, Springer Verlag, Berlin Heidelberg.

[50] P. Erdős and R. Rényi, A. (1960), On the evolution of random graphs. Reprinted in *The Art of Counting* 1973 (pp. 574–618). Cambridge, MA: MIT Press. Also in *Selected Papers of Alfred Rényi* (pp. 482–525) (1976). Budapest: Akademiai Kiado.

[51] L. Fajstrup (2005), Dipaths and dihomotopies in a cubical complex, *Adv. in Appl. Math.* 35, 188–206.

[52] U. Feudel et al. (2000), Homoclinic bifurcation in a Hodgkin–Huxley model of thermally sensitive neurons, *Chaos* 10 No. 1, 231–239.

[53] R. Feynman (1996), *Feynman Lectures on Computation*, Addison–Wesley, Reading MA.

[54] W. J. Freeman (1999), Consciousnness, intentionality and causality, *J. Consciousness Studies* 11 (1999), 143–172.

[55] A. Frölicher and A. Kriegl (1988), *Linear Spaces and Differentiation Theory*, J. Wiley, Chichester.

[56] D. Gabor (1946), Theory of communication *J. IEE (London)* 93 (III), 429–457.

[57] L. Gammaitoni, P. Hänggi, P. Jung and F. Marchesoni (1998), Stochastic resonance, *Rev. Mod. Phys.* 70 No. 1, 223–287.

[58] W. Gerstner and W. Kistler (2002), *Spiking Neuron Models*, Cambridge University Press.

[59] T. Giordano, I. Putnam and C. Skau (2004), Affable equivalence relations and orbit structure of Cantor dynamical systems, *Ergodic Theory Dynam. Systems* 24 No. 2, 441–475.

[60] J. F. Glazebrook and R. Wallace (2009), Small Worlds and Red Queens in the Global Workspace: an information–theoretic approach, *Cognitive Systems Research*, to appear.

[61] M. Golubitsky and I. Stewart (2006), Nonlinear dynamics of networks: the groupoid formulism, *Bull. Amer. Math. Soc.* 43, No. 3, 305–364.

[62] E. Goubault (2003), Some geometric perspectives in concurrency theory, *Homology, Homotopy and Applications* 5 (2), 95–136.

[63] M. S. Granovetter (1973), The strength of weak ties, *Amer. J. Sociol.* 78, 1360–1380.

[64] S. Gudder (2004), Noncommutative probability and applications, *Real and Stochastic Analysis–New Perspectives* (M. Rao, ed.), pp 199–238, Birkhäuser, Boston–Basel–Berlin.

[65] N. S. Harper and D. McAlpine (2004), Optimal neural population coding of an auditory spatial cue, *Nature* 430, 682–686.

[66] M. J. Healy and T. P. Caudell (2006), Ontologies and worlds in category theory: Implications for neural systems, *Axiomathes* 16 Nos. 1–2, 165–214.

[67] M. J. Healy and T. P. Caudell (2004), Neural networks, knowledge and cognition: A mathematical semantic model based upon category theory. Techninal report EECE-TR-04-20, Dept. of Electrical and Computer Engineering, University of New Mexico.

[68] J. Heinloth (2004), Some notes on differentiable stacks, *Math. Inst. Sem.* (Y. Tschinkel, ed.), 1–32, Univ. Göttingen.

[69] B. Hillier (2003), The knowledge that shapes the city, in *Proc. 4th International Space Syntax Symposium*, London.

[70] J. Hollan, E. Hutchins and D. Kirsh (2000), Distributed cognition: Toward a new foundation for human–computer interaction research, *ACM Transactions on Computer Interaction* 7 (2), 174–196.

[71] E. Holmes (2005), On being the right size, *Nature Genetics* 37, 543–546.

[72] F. C. Hoppensteadt and E. M. Izhikevich (1997), *Weakly Connected Neural Networks*, Appl. Math Sciences 126, Springer Verlag.

[73] P. W. Huber (1987), The Geodesic Network – 1987 Report on competition in the telephone industry, *U.S. Dept. of Justice*, U. S. Gov. Printing Office, Washington DC.

[74] E. Hutchins (1994), *Cognition in the Wild*, MIT Press, Cambridge MA.

[75] N. K. Jerne (1974), Towards a network theory of the immune system, *Ann. Imunology (Inst. Pasteur)*, 125C, 373–389.

[76] H. Kamps and T. Porter (1996), *Abstract Homotopy and Simple Homotopy Theory*, World Scientific, Singapore.

[77] A. Katok and B. Hasselblatt (1995), *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press.

[78] G. M. Kelly and R. Street (1974), Review of the elements of 2–categories, in *Category Seminar (Proc. Sem. Sydney, 1972/1973)*, *Lect. Notes in Math.* 420, Springer, Berlin, 75–103.

[79] A. I. Khinchin (1957), *Mathematical Foundations of Information Theory*, Dover Publ.

[80] K. Kitajo, D. Nozaki, L. M. Ward and Y. Yamamoto (2003), Behavorial stochastic resonance within the human brain, *Phys. Rev. Lett.* 90 No. 21, 218103, 1–4.

[81] A. Kock (2006), *Synthetic Differential Geometry*, Cambridge Univ. Press.

[82] A. Kock (2006), Connections and path connections in groupoids, Aarhus Univ. Math. Preprint No. 10.

[83] S. M. Kosslyn (2007), On the evolution of human motivation: The role of social prosthetic systems, in (Platek et al. eds) *Evolutionary Cognitive Neuroscience*, MIT Press.

[84] R. Kozma, M. Puljic, P. Balister, B. Bollobas and W. J. Freeman (2004), Neuropercolation: a random cellular automata approach to spatio–temporal neurodynamics, *Lecture notes in Computer Science* 3305, 435–443.

[85] N. P. Landsman (1998), *Mathematical Topics between Classical and Quantum Mechanics*, Springer Verlag, New York.

[86] G. Laumon and L. Moret-Baily (2000), Champs algébriques, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge., vol. 39, Springer-Verlag, Berlin.

[87] J. Lee (2000), *Introduction to topological manifolds*, Graduate Texts in Math. 202, Springer-Verlag, New York.

[88] M. Li and P. Vitányi (1993), *An Introduction to Kolmogorov complexity and its applications*, Texts and Monographs in Computer Science, Springer-Verlag, New York.

[89] D. Lind and B. Marcus (1995), *An Introduction to Symbolic Dynamics and Coding*, Cambridge Univ. Press.

[90] C. K. Machens, T. Gollisch, O. Kolesnikova and A. V. Herz (2005), Testing the efficiency of sensory coding with optimal stimulus ensembles, *Neuron* 47 (3), 447–456.

[91] S. Mac Lane and I. Moerdijk (1992), *Sheaves in Geometry and Logic: A first Introduction to Topos Theory,* Springer Verlag, New York.

[92] K. C. H. Mackenzie (1987), *Lie Groupoids and Lie Algebroids in Differential Geometry*, London Math. Soc. Lect. Notes 124, Cambridge University Press.

[93] H. R. Maturana and F. J. Varela (1980), *Autopoiesis and Cognition–The Realization of the Living*, Boston Studies in the Philosophy of Science Vol. 42, Reidel Pub. Co., Dordrecht.

[94] J. Mayor and W. Gerstner (2005), Noise–enhanced computation in a model of cortical column, *NeuroReport* 16 (11), 1237–1240.

[95] J. Milnor (1963), *Morse Theory*, Princeton University Press, Princeton NJ.

[96] I. Moerdijk and J. Mrčun (2003), *Introduction to Foliations and Lie Groupoids*, Cambridge Studies in Adv. Math. 91, Cambridge Univ. Press.

[97] R. Montgomery (1990), Isoholonomic problems and applications, *Commun. Math. Phys.* 128 (3), 565–592.

[98] T. Murayama and M. Okada (2003), One step RSB scheme for the rate distortion function, *J. Phys. A. Math. Gen.* 36, 11123–11130.

[99] R. Nisbett (2003), *The Geography of Thought: How Asians and Westerners think differently ... and why*, Free Press, New York.

[100] J. Peebles, I. F. Putnam and I. F. Zwiers, A survey of orbit equivalence for Cantor minimal dynamics, to appear.

[101] G. Pistone and C. Sempi (1995), An infinite–dimensional geometric structure on the space of all the probability measures equivalent to a given one, *Annals of Statistics* **25** (5), 1543–1561.

[102] W. F. Pohl (1962), Differential geometry of higher order, *Topology* 1, 169–211.

[103] T. Porter (2008), Enriched categories and models for spaces of evolving states, *Theoretical Computer Science* 405, No. 1-2, 88–100.

[104] A. Pouget, P. Dayan and R. Zemel (2000), Information processing within population codes, *Nature Rev.* Vol 1, 125–132.

[105] V. Pratt (1991), Modeling concurrency with geometry, in *Proceedings of the 18th ACM Symposium on Principles of Programming Languages*.

[106] V. Pratt (2003), Chu spaces as a semantic bridge between linear logic and mathematics, *Theoretical Comp. Sci.* 294, 439–471.

[107] K. H. Pribram (1991), *Brain and Perception: Holonomy and Structure in Figural Processing*, Lawrence Erlbaum Assoc., Hillsdale NJ.

[108] I. Prigogine (1980), *From Being to Becoming: Time and Complexity in the Physical Sciences*, W. H. Freeman and Co., San Francisco.

[109] G. Robins, P. Pattison and J. Woolcock (2005), Small and other worlds: Global structures from local processes, *Amer. J. Sociology* 110 (4), 894–936.

[110] R. Sampson (2004), Networks and neighborhoods, in *Network Logic: Who governs in an Interconnected World?* (H. McCarthy, P. Miller and R. Skidmore, eds.), Demos, London.

[111] E. Schneidman, W. Bialek and M. J. Berry II (2002), An information theoretic approach to the functional classification of neurons, *Advances in Neural Information Processing Systems* 15 197–204, MIT Press, Cambridge MA.

[112] C. E. Shannon (1948), A mathematical theory of communication, *Bell Syst. Tech. J.* 30, 379–423, 623–656.

[113] V. V. Sharko (1993), Functions on manifolds, algebraic and topological aspects, *Amer. Math. Soc. Translations* 131, Amer. Math. Soc., Providence RI.

[114] T. Sharpee et al. (2006), Adaptive filtering enhances information transmission in visual cortex, *Nature* 439, 936–942.

[115] D. J. Simons and C. F. Chabris (1999), Gorillas in our midst: Sustained inattentional blindness for dynamic events, *Perception* 28, 1059–1074.

[116] N. Sourlas (1989), Spin–glass models as error–correcting codes, *Nature* 339, 693–695.

[117] E. H. Spanier (1966), *Algebraic Topology*, McGraw Hill.

[118] I. Stewart, M. Golubitsky and M. Pivato (2003), Symmetry groupoids and patterns of synchrony in coupled cell networks, *Siam J. Appl. Dynamics* 2 No. 4, 609–646.

[119] R. Sun (2009), Theoretical status of computational cognitive modeling, *Cognitive Systems Research* 10 No. 2, 124–140.

[120] G. Szabó and G. Fáth (2007), Evolutionary games on graphs, *Phys. Rep.* 446, 97–216.

[121] J. B. Tenenbaum, V. de Silva and J. C. Langford (2000), A global geometric framework for nonlinear dimensionality reduction, *Science* 290, 2319–2323.

[122] J. T. Townsend, B. Solomon and J. S. Smith (2001), The perfect Gestalt: Infinite dimensional Riemannian face spaces and other aspects of face perception. In M. J. Wenger and J. T. Townsend (Eds.) *Computational geometric and process perspectives of facial cognition: Contexts and challenges*, 39–82, Mahwah, NJ: Lawrence Erlbaum Associates.

[123] J. Virsik (1971), On the holonomity of higher order connections, *Cah. Top. Géom. Diff. Cat.* 12, 197–212.

[124] E. Volz (2004), Random networks with tunable degree distribution and clustering, *Phys. Rev. E* 70, 056115.

[125] R. Wallace (2005), *Consciousness: A Mathematical Treatment of the Global Neuronal Workspace*, Springer, New York.

[126] R. Wallace (2005), A global workspace perspective on mental disorders, *Theoretical Biology and Medical Modelling* 2, 49.

[127] R. Wallace (2005), The sleep cycle: a mathematical analysis from a global workspace perspective. http://cogprints.org/4517/

[128] R. Wallace (2007), Culture and inattentional blindness: a global workspace perspective, *Journal of Theoretical Biology*, 245, 378–390.

[129] R. Wallace (2006), Public policy, institutional cognition and the geographic diffusion of multiple–drug resistant HIV in the United States, preprint.

[130] R. Wallace (2006) New mathematical foundations for AI and Alife: Are the necessary conditions for animal consciousness sufficient for the design of intelligent machines?, preprint.

[131] R. Wallace (2006), Institutional cognition, preprint.

[132] R. Wallace (2006), Varieties of animal consciousness: the spectrum of mathematical Global Workspace models and its implications, preprint.

[133] R. Wallace (2002), Adaptation, punctuation, and information: a rate distortion approach to non-cognitive 'learning plateaus' in evolutionary process, *Acta Biotheoretica* 50, 101–116.

[134] R. Wallace (2008), Toward formal models of biologically inspired, highly parallel machine cognition, *International Journal of Emergent, Parallel and Distributed Systems* 23, 367–408.

[135] R. Wallace and M. Fullilove (2008), *Collective Consciousness and Its Discontents: Institutional distributed cognition, racial policy, and public health in the United States*, Springer, New York.

[136] R. Wallace and R. G. Wallace, *Psychopathica Automatorum: A cognitive neuroscience perspective on highly parallel computation and its dysfunctions*, to appear.

[137] D. Wallace and R. Wallace (1998), *A Plague on Your Houses*, Verso Press, New York.

[138] R. Wallace and R. G. Wallace (1998), Information theory, scaling laws and the thermodynamics of evolution, *J. Theor. Biology* 192, 545–559.

[139] R. Wallace and D. Wallace (2004), Structured psychosocial stress and therapeutic failure, *Journal of Biological Systems* 12, 335–369.

[140] R. Wallace, D. Wallace and R. G. Wallace (2003), Towards cultural oncology: the evloutionary information dynamics of cancer, *Open systems and Information Dynamics* 10, 159–181.

[141] R. Wallace and R. G. Wallace (2008), On the spectrum of prebiotic chemical systems: An information theory treatment of Eigen's paradox, *Origins of Life and Evolution of Biospheres* 38 No. 5, 419–4551.

[142] D. J. Watts (1999), *Small Worlds: The Dynamics of networks between Order and Randomness*, Princeton Univ. Press, Princeton, NJ.

[143] A. Weinstein (1996), Groupoids: unifying internal and external symmetry, *Notices of the Amer. Math. Soc.* 43, 744–752.

[144] R. L. West, T. C. Stewart, C. Lebiere and S. Chandrasekharan (2005), Stochastic resonance in human cognition: ACT-R versus game theory, associative neural networks, recursive neural networks, Q-learning and humans, in *Proc. 27th Annual Meeting of The Cognitive Science Soc.*, 2353–2358.

[145] J. Zhang and P. Hästö (2006), Statistical manifold as an affine space: A functional equation approach, *J. Math. Psychology* 50(1), 60–65.

[146] R. Živaljević (2008), Combinatorial groupoids, cubical complexes, and the Lovász conjecture, *Discrete and Computational Geometry* 41(1), 135–161.