Making Sense of Internal Logic: Theory and a Case Study

Ichiro Tsuda and Motohiko Hatakeyama

Series #492. September 2000

HOKKAIDO UNIVERSITY
PREPRINT SERIES IN MATHEMATICS

#468 A. Higuchi, K. Matsue and T. Tsujishita, Deductive hyperdigraphs – a method of describing diversity of coherences, 35 pages. 1999.

#469 M. Nakamura and T. Ozawa, Small solutions to nonlinear Schrödinger equations in the Sobolev spaces, 26 pages. 1999.

#470 M. Nakamura and T. Ozawa, Small solutions to nonlinear wave equations in the Sobolev spaces, 27 pages. 1999.

#471 K. Ito and Y. Kohsaka, Stability of a stationary solution for evolving boundaries of symmetric three-phases driven by surface diffusion, 36 pages. 1999.

#472 Y. Giga and M.-H. Sato, A level set approach to semicontinuous viscosity solutions for Cauchy problems, 32 pages. 1999.

#473 R. Yoneda, Characterization of Bloch space and Besov spaces by oscillations, 31 pages. 1999.

#474 T. Nakazi, Norm inequalities for some singular integral operators, 13 pages. 1999.

#475 A. Inoue and Y. Kasahara, Asymptotics for prediction errors of stationary processes with reflection positivity, 15 pages. 1999.

#476 T. Sano, On affine parallels of generic plane curves, 8 pages. 1999.

#477 T. Nakazi, On an invariant subspace whose common zero set is the zeros of some function, 11 pages. 1999.

#478 M.-H. Giga and Y. Giga, Generalized motion by nonlocal curvature in the plane, 68 pages. 2000.

#479 M.-H. Giga and Y. Giga, Crystalline and level set flow – Convergence of a crystalline algorithm for a general anisotropic curvature flow in the plane, 16 pages. 2000.

#480 A. Arai and M. Hirokawa, Stability of ground states in sectors and its application to the Wigner-Weisskopf model, 16 pages. 2000.

#481 T. Nakazi, Two dimensional Q-algebras, 11 pages. 2000.

#482 N. H. Bingham and A. Inoue, Tauberian and Mercerian theorems for systems of kernels, 16 pages. 2000.

#483 N. H. Bingham and A. Inoue, Abelian, Tauberian and Mercerian theorems for arithmetic sums, 29 pages. 2000.

#484 I. A. Bogaevski and G. Ishikawa, Lagrange mappings of the first open Whitney umbrella, 22 pages. 2000.

#485 A. Arai and H. Kawano, A class of deformations of the Schrödinger representation of the Heisenberg commutation relation and exact solution to a Heisenberg equation and a Schrödinger equation, 22 pages. 2000.

#486 T. Nakazi, Functions in N_+ with the positive real parts on the boundary, 21 pages. 2000.

#487 Y. Shibukawa, Classification of the R-operator, 36 pages. 2000.

#488 A. Inoue, Asymptotic behaviour for partial autocorrelation functions of fractional ARIMA processes, 20 pages. 2000.

#489 S. Ohtani, Construction of unramified Galois extensions over maximal abelian extensions of algebraic number fields, 14 pages. 2000.

#490 T. Nakazi and T. Yamamoto, The real part of an outer function and a Helson-Szegö weight, 13 pages. 2000.

#491 A. Yamagami, On Gouvêás conjecture on controlling the conductor, 11 pages. 2000.

Making Sense of Internal Logic: Theory and a Case Study

Ichiro Tsuda and Motohiko Hatakeyama
Applied Mathematics and Complex Systems Research Group,
Department of Mathematics, Graduate School of Science,
Hokkaido University,
Sapporo, 060 Japan

Abstract

Motivated by the *interfaciology* proposed by Otto Rössler, we have attempted to construct a framework of internal logic of the mind and brain. We propose a functional equation as an abstract form representing mental processes. We consider a method by which such internal logic can be interpreted and understood by an (external) observer. For this purpose, we propose a theory for cognitive experiments. Applying this theory to simple deductive inference processes exhibited by animal subjects in an experimental setting, with the assumption that syllogism is expressed as a composite mapping corresponding to the product operation of two implications $A \to B$ and $B \to C$, an interpretation of the neural activity associated with the behavior in these experiments is obtained. This theory is consistent with the internal description hypothesized by Rob Rosen.

1 Introduction

In this paper we are concerned with the question of how an external observer can understand the internal logic of a subject. In order to study this question, we wish to develop a general theory that can be used to construct and implement an interface between the observer and subject that can allow for such an understanding [1].

As a first step in this search for an interface between the observer and the subject in cognitive experiments, we assume the presence of neuronal correlates of cognition, especially with regard to the correlation between neural activity and its implication at the behavioral level, with the hope that a certain type of cognitive experiment on humans or animals leads us to identify such an interface. We have constructed a theory of a supposed internal process of inference. In this theory, the *discrimination* based on some equivalence class is introduced. By using this concept, the correspondence between the space of attributes of external stimuli and neuronal activity is derived and interpreted as the interface in question [2]. In this paper, we briefly outline this theory, which is consistent with the internal description hypothesized by Rob Rosen [3] in his relational biology.

In our model of the experiments, the environment is expressed in terms of some dynamical variable or variables. Here we denote these by X_t. Let us consider two extreme cases for internal dynamics of the brain (Fig. 1). In one case these dynamics consist of a copy of the external world. In this case, the internal dynamics are identical to the environmental dynamics, X_t. Thus the dynamics of the brain constituting the *description* are of *fixed-point* nature. The other extreme case is that in which the internal dynamics describe the external world in the form of a *constant function.* This represents a rigid description, independent of any dynamics of the environment. The former extreme here is the completely adaptive case, and the latter is the autistic case. Clearly actual brain dynamics must be somewhere between these two extremes. Such an intermediate dynamics could be chaotic at the description level. We propose a functional equation for the *description*, which we believe may represent a meta-dynamics of mental processes.

–Fig. 1–

2 A theory for cognitive experiments

In order to formulate an effective interface, we have searched a sufficiently simple yet meaningful cognitive experiment. As one candidate for such an experiment, we considered the type recently carried out by Sakagami and Niki [4] and Sakagami and Tsutsui [5].

They performed a set of experiments investigating multidimensional visual discrimination tasks with monkeys. In these experiments, they recorded the prefrontal cell activity while monkeys performed a *go/no go attention task.* The subjects were trained to make a *go* or *no go* response, depending on, for instance, the color, the direction of motion, or the shape of compound visual stimuli, namely moving patterns of different colors and shapes. Moreover, the subjects had to "understand" the meaning of the color of the fixation spot that is presented at the center of the screen at each trial. That is, for each trial, the subject is presented a pattern. Each pattern is in general characterized by the attributes of motion, color and shape, but for a given trial only one of these is "important". The important attribute is indicated by the color of the fixation spot. For instance, if this color is purple, then the important attribute is color (we call this situation the *color condition*), if it is blue this attribute is motion (the *motion condition*), and if it is red this attribute is shape (the *shape condition*). The subject is then required to act in some way, according to the state of the important attribute, namely the kind of condition. The basic formulation of the experiment can be understood from the following. At the beginning of a trial, the fixation spot is presented. With the important attribute thereby fixed, the subject is presented with a pattern. For each attribute, certain states correspond to the *go* response, and certain attributes correspond to *no go* response. For example, red means *go* and right-directed motion means *no go.* Thus if a red, right-moving pattern is presented, the correct response is *go* in the color condition but *no go* in the motion condition.

We formulated this experiment as a case study. Let us denote a set of conditions by $X = x_c, x_m, x_s$, where x_c represents the color condition, x_m the motion condition and x_s the shape condition. Let us denote a set of target stimuli by $\mathbf{Y} = Y_c \times Y_m \times Y_s$, where Y_c is the set of color attributes of the target stimuli, Y_m the set of motion attributes, and Y_s

the set of shape attributes. Finally, a set of responses is denoted by $Z = \{z_g, z_{ng}\}$, where z_g is a *go* response and z_{ng} a *no go* response.

Then, a learned response to stimuli can be expressed by a mapping $f : X \times \mathbf{Y} \to Z$. It is reasonable to assume that this mapping can be described by the following *renewal mapping system*:

$$f(x, (y_c, y_m, y_s)) = \begin{cases} f_c(y_c) & (x = x_c) \\ f_m(y_m) & (x = x_m) \\ f_s(y_s) & (x = x_s), \end{cases} \tag{1}$$

where $f_c : Y_c \to Z$ is a response to a color attribute in the color condition, $f_m : Y_m \to Z$ a response to a motion attribute in the motion condition, and $f_s : Y_s \to Z$ a response to a shape attribute in the shape condition. This is a kind of description at the cognitive level. On the other hand, a physiological description is represented by a mapping of the form $F : X \times \mathbf{Y} \to$ *neural activity*. The mapping F represents the activity of neurons or neuron assemblies resulting from stimuli in the case that the correct response is given.

We define a *discrimination* as a decomposition of a set A derived from the equivalence relation induced by a mapping $f : A \to B$ for any sets A and B. Sakagami *et al.* found several types of neurons which respond conditionally or unconditionally to a certain specific attribute of stimuli. Applying these results, it is possible to compare the theoretical classification of the stimulus space $X \times \mathbf{Y}$ with the classification based on neuronal activity.

Let us now consider candidates for deductive inference which could occur during the performance of a cognitive task. There could exist two elicited higher-order functions of f (by the Curry method) , which we denote here as g and h. The map g is defined as a renewal mapping system $\mathbf{Y} \to Z$ depending on the conditions X, whereas the map h is defined as a renewal mapping system $X \to Z$ depending on the stimuli $\mathbf{Y}$. Since for each trial in the above experimental setup an important attribute is first recognized by the subject, the existence of g is more plausible than h.

Two possible decompositions of g correspond to two kinds of attention, *focal attention* and *divided attention*. Here, focal attention is attention to a particular attribute, and divided attention is attention to all attributes. One decomposition can be expressed as a composite mapping, $\mathbf{Y} \to Y \to Z$. In other words, a particular attribute of the stimulus is first selected, and then behavior is determined from this attribute. The second decomposition can be expressed as another composite mapping, $\mathbf{Y} \to \mathbf{Z} \to Z$. In other words, the processing of a stimulus to its behavioral implication is first carried out parallelly for each attribute, and then a particular behavior is selected. The causality in the process of the former decomposition is summarized as *selection process* $\Longrightarrow$ *transformation process*, and that of the latter is summarized as *transformation process* $\Longrightarrow$ *selection process*. Thus, the former decomposition represents the focal attention and the latter divided attention.

Taking into account the above considerations, we can determine the internal processes that occur at the level of each type of neuron. The details will be published elsewhere.

3 The internal description must be dynamic

Let us assume the environment to be dynamic and described by $X_{n+1} = F(X_n)$, where n is a discrete time step. In general, F can represent some nonlinear, chaotic dynamics. The brain inevitably possesses a *description* of F, which is here denoted by h_n. We assume that this description changes to follow another dynamical system $\tilde{F}$. Then we obtain the functional equation

$$h_{n+1} \circ F = \tilde{F} \circ h_n. \tag{2}$$

The dynamics of this description are schematically depicted in Fig. 2.

–Fig. 2–

Let us consider the two special cases. In the case $\tilde{F} = F$, what is an invariant description? This condition implies that the same dynamical system can be used to model both the external world and its internal description. A trivial solution in this is

$$h^*(X) = X. \tag{3}$$

Here, the description provides a copy of the external world. This is the first extreme case discussed in the Introduction (see also Fig. 1).

The second special case we consider is that in which the dynamics of the description has a fixed point, that is, the case $\tilde{F}(X) = X$. Then we obtain

$$h_{n+1}(X_{n+1}) = h_n(X_n). \tag{4}$$

This implies the presence of a fixed point of the description, irrespective of the external world. This is the second extreme case discussed in the Introduction (see also Fig. 1).

The functional equation (2) for the internal description of the brain contains a great variety of solutions. When $F(X_n) = X_n$ and $\tilde{F} = h$, this functional equation takes on a form similar to that of the functional map of Kataoka-Kaneko [6].

In this paper we have treated the case of a learned state only, and thus the present theory does not provide a true dynamics of description. We are just at the starting point. In order to treat such a dynamics in the brain more realistically, we have to treat the learning process itself. In such a situation, the dynamics of the functional equation are expected to provide useful information.

Acknowledgment

One of the authors (IT) would like to express his sincere thanks to Otto Rössler for having continually given him inspiration in the study of chaos, endophysics, and dynamic interfaces. He also thanks Reimara and Michale Rössler for stimulation. We express our sincere thanks to Masamichi Sakagami, Hideaki Saito and Minoru Tsukada for fruitful discussions on an interpretation of the series of experiments conducted by Sakagami *et al.* We also thank Hans Diebner for useful discussions on the manuscript.

References

[1] Rössler, O. E. *Endophysics: The World as an interface.* (World Scientific, 1998).

[2] Hatakeyama, M. and Tsuda, I. , *in preparation.*

[3] Rosen, R. *Life Itself: A comprehensive inquiry into the nature, origin, and fabrication of life.* (Columbia University Press, New York, 1991).

[4] Sakagami, M. and Niki, H. Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: relation to relevant task conditions. *Exp. Brain Res.* **97** (1994) 423-436.

[5] Sakagami, M. and Tsutsui, K. The hierarchical organization of decision making in the primate prefrontal cortex. *Neuroscience Res.* **34** (1999) 79-89.

[6] Kataoka, N. and Kaneko, K. Functional dynamics I: articulation process. *Physica D* **138** (2000) 225-250.

Figure captions

Fig. 1 The extreme cases of brain dynamics are depicted. One extreme is the case in which these dynamics constitute a copy of the external world. Conventional neural network studies are concerned with this case. The other extreme is the case of autism, in which the brain dynamics are fixed, irrespective of the external world. The actual situation must be somewhere between these two cases.

Fig. 2 The dynamics of the internal description are modeled by a functional transformation. A certain instability at the level of the description insures the development of the level of internal description, depending on the external dynamics.

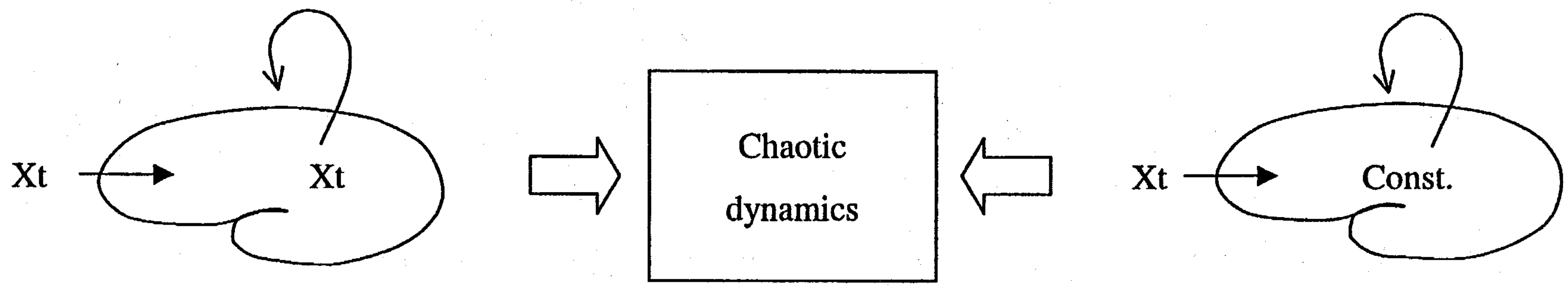
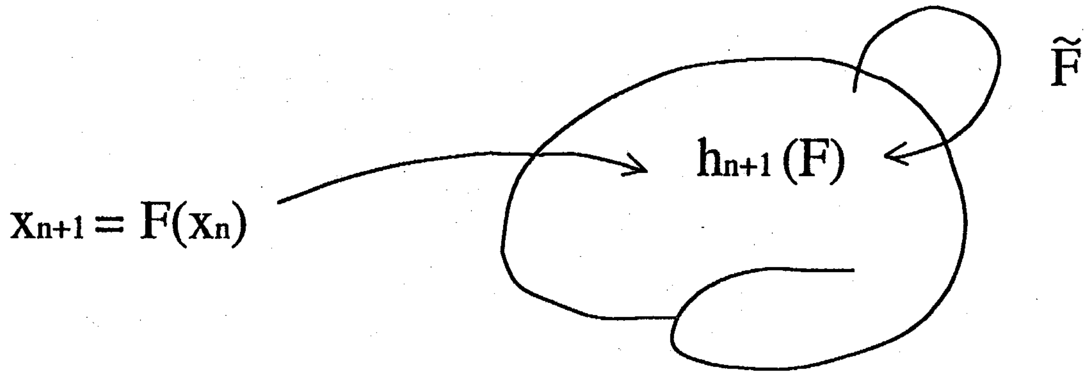


Fig 1. Tsuda & Hatakeyama



$$h_{n+1} \circ F = \tilde{F} \circ h_n$$

$$\begin{tikzcd}
x \arrow[r, mapsto, "F"] \arrow[d, mapsto, "h_1"'] & F(x) \arrow[r, mapsto, "F"] \arrow[d, mapsto, "h_2"'] & F^2(x) \arrow[r, mapsto] \arrow[d, mapsto, "h_3"'] & \cdots \\
h_1(x) \arrow[r, mapsto, "\tilde{F}"] & \tilde{F} \cdot h_1(x) \arrow[r, mapsto, "\tilde{F}"] & \tilde{F}^2 \cdot h_1(x) \arrow[r, mapsto] & \cdots
\end{tikzcd}$$

Fig. 2 Tsuda & Hatakeyama