AISB 2020 Convention, St Mary's University, London, April 7-9 2021

http://aisb20.wordpress.com

Symposium "Representation and Reality in Humans, Other Living Beings and Intelligent Machines: New Trends

Organizers: Raffaela Giovagnoli (PUL), Gianfranco Basti (PUL), Robert Lowe (University of Gothenburg)

Invited speakers: Gordana Dodig-Crnkovic (Chalmes University), Susan Stepney (University of York, Lorenzo Magnani (University of Pavia), Michael Barber (St Louis University), Zoran Konkoli (Chalmes University), Giuseppe Vitiello (University of Salerno), Dean Petters (Sheffield Hallam University)

INTRODUCTION

Our symposium is the continuation of a series of events starting with the Symposium "computing nature" organized at the AISB/IACAP World Congress 2012. We would like to offer an occasion to discuss new directions in the development of understanding of representation as a means of sense-making and communication. It is closely related to the question of what capacities can be plausibly computed and what are the most prominsing approaches that try to solve the problem.

The authors discuss the various faces of the relationship between reality and representation in humans, other animals and machines from different perspectives (cognitive, computational, philosophical, logical and machine-centered) and deep insights into the topic. Among other issues, the symposium focuses on:

- The discussion of current developments of the classical debate between representationalism and anti-representationalism with the question of *in what sense* it can be argued that cognition relies on *representations mirroring* reality, with its assumptions and constrains and in what way it is an *adaptive form of dynamics* based on the interaction of an agent with the environment.

- A fruitful strategy to analyze the problem of representation from a philosophical perspective that implies the comparison between human and machine capacities and skills. Searle presented an interesting theory of representation based on the mind's capacity to represent objects and to the linguistic capacities to extend the representation to social entities.

- Evolutionary aspects of the development of increasingly complex capacities in (embodied, embedded) living organisms to process information in the interaction with the environment and as a consequence development of new morphological structures - process of morphogenetic is and meta-morphogenetic which we want to elucidate from multiple disciplinary and interdisciplinary perspectives, from philosophy to neuroscience and computational approaches and cognitive science.

On the use of collaborative interactions for embedded sensing applications: Memristor networks as intelligent sensing substrates

Vasileios Athanasiou [1] and **Zoran Konkoli** [2]

Abstract. A novel sensing approach has been investigated in which environment-sensitive memristor networks are used as intelligent sensing substrates. A substrate collects pieces of environment-related information over time and encodes this information into its state. The stored information can be extracted by monitoring how the substrate responds to an external drive signal. An advantage of this indirect sensing approach is that the drive signal can be optimised to make the inference process efficient: even small pieces of information (which might go unnoticed in the traditional sensing setup) are collected. To demonstrate the main ideas an instance of a binary classification problem has been investigated. A separability index has been used as a measure of the substrate quality. By simulating the dynamics of a large number of memristor networks and computing their separability indices, it has been found that heterogeneous networks with delayed feedback elements make good sensing substrates.

1 Introduction

The most important architectural feature of a traditional sensing device is that data acquisition and data analysis processes are separated and must be engineered as two distinct modules, rendering the devices unnecessarily large and energy-inefficient. This study explores the possibility of constructing sensing devices, in which the data acquisition and the data analysis steps are performed simultaneously. Such sensors can be used in situations where it is necessary to: have embedded information processing, pre-process information *in situ* and reduce the necessary communication bandwidth and need for centralised analysis. There are further advantages to decentralising computational power such as low power consumption, less post-processing and bio-compatibility [11]. Moreover, from the technological point of view, such sensing solutions could be more flexible and with fewer engineering constraints needed to implement them.

As the overarching design principle for constructing decentralised sensors, we explore the option of using intelligent sensing substrates that both accumulate and analyze environment-related information *in situ* and in real-time. [8] An intelligent sensing substrate should be an environment-sensitive dynamical system, the state of which may be queried. Scattered pieces of information, which might go unnoticed or lost in the traditional sensing setup, can be collected gradually and stored in the substrate state. The substrate state is inferred by interfacing it with a readout layer. Ideally, the "intelligence" of the sensing device should reside in the substrate and not in the readout layer: the readout layer should be a simple inference unit.

[1] Chalmers University of Technology, Sweden, email: vasath@chalmers.se
[2] Chalmers University of Technology, Sweden, email: zorank@chalmers.se

A "good" substrate should be able to adopt a different configuration for every distinct environmental condition to which it has been exposed (the separability property). Conversely, a "bad" substrate is completely unresponsive, staying in the same state regardless of the environmental condition. A further desirable substrate property is that the environment-related information encoded in the substrate should be easy to decode. If an environmental condition always drives the system to the same region of the state space, then the substrate exhibits a clustering property. This can be used to infer this particular environmental condition, because the associated decision boundary in the state space represents a surface of a closed region. Taken together, the separability and clustering properties determine the amount of environment-related information that can be inferred (the sensing capacity of a substrate). The best substrates feature both these properties and can be used to build an efficient sensing device by using a simple readout layer.

We posit that memristor networks are natural sensing substrates. The memristor is one of the simplest electronic components which accumulate information about their past [4]. It is a non-linear, passive, two-terminal component. The voltage-current response of such an element resembles that of a normal resistor. However, the resistance value can change over time depending on the current that has been passing through the memristor. The resistance value can only vary within a finite interval $[R_{min}, R_{max}]$, where the bounds R_{min} and R_{max} are material dependent constants. The state space of a multi-memristor substrate is a collection of points $(R_1, R_2, \cdots, R_{N_R})$ where N_R is the number of memristances in the system, with each memristance bound by the respective interval $[R_{min}, R_{max}]$. If it can be arranged for different environmental conditions to drive each memristance to either of the bounds, then the inference process can be simple. This allows for a very natural encoding scheme in which distinct environmental conditions can be associated with distinct sequences of R_{min} and R_{max}. One might think of this as a binary encoding scheme, in which environmental conditions are labelled with binary words of length N_R with logical 0 (1) representing R_{min} (R_{max}).

Indeed, in the earlier study [2], it was demonstrated *in silico* that a single memristor can be used to distinguish between two classes of environmental signals, representing a static and a varying environment. It was shown that there is a way to operate the element so that the memristance value is driven towards R_{max} or R_{min}, depending on whether the element is exposed to either of the environmental conditions. This was done by optimising (training) an external drive signal using a supervised learning technique. What makes the sensor efficient is that the analysis part is memory-less, as it just means checking whether the memristance values cluster close to R_{max} or R_{min}. If this were not the case, then an auxiliary analysis tool would need to be used, such as an artificial neural network.

The overarching goal is to identify topological features that are a trademark of efficient memristor networks and to understand whether the cooperative behaviour between memristor components can be exploited to gain additional sensing functionality. The separability index is used to compare memristor networks with different topological features in three ways: by considering more nodes and by adding more connections between the nodes (network complexity), by adding delayed feedback element around memristors (element complexity) and by combining these two approaches.

The work is organised as follows. Section 2 explains the sensing procedure in detail. In section 3 the substrate model is introduced and an explanation is provided regarding how the collaboration between network elements is expected to increase the sensing capacity. Section 4 introduces the GA scheme used to train the drive signal for a environment classification problem. Section 5 presents a strategy of searching for optimal network designs. The best performing systems are presented in the results section 6. A summary of the main findings appears in section 7. This section contains a comprehensive synthesis of numerical results. Section 8 sums up important aspects of the work and points to possible extensions.

2 The sensing setup

Figure 1 illustrates the sensing setup of interest. The most important mathematical primitives that are necessary to formalize the sensing problem are shown [7, 8]. Depicted is a hand-drawn example of a binary classification problem. The key primitive is the drive signal $u(t)$. The drive signal can be optimised to improve the sensing capacity of the device. The reservoir (substrate) state should be strongly correlated with the environmental signals q_1 and q_2 that represent the two states of the environment. In the figure, a two-dimensional state space is assumed, described by two memristances R_1 and R_2.

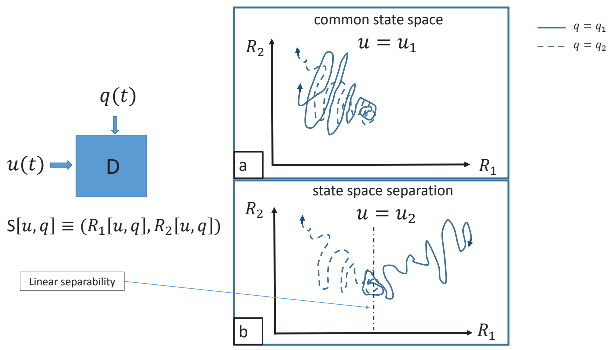

Figure 1. Exploiting a reservoir D in the SWEET sensing setup context. Hand-drawn examples appear in a) and b): a) illustrates when the reservoir is driven by a signal u_1 and state space separation is not feasible. In other words, the trajectories of the state are driven to common regions of the configuration space. In b), state space separation is feasible when driven by signal u_2, with the trajectories driven to different regions.

The reservoir D is "driven" by both the drive signal $u(t)$ and an environmental signal $q(t)$. The key point is that while the environmental signal cannot be controlled the drive signal can. The state of the reservoir changes in time and forms a trajectory in the two dimensional space as shown in Figs. 1a and 1b. Figure 1a shows two typical state trajectories when the reservoir is driven by a drive signal u_1 that is not optimal. Regardless of whether $q = q_1$ or $q = q_2$ the reservoir adopts roughly the same state. In this case, the trajectories in the figure are driven to common regions of the configuration space and the environmental condition cannot be inferred.

Conversely, in Fig. 1b, under another drive signal u_2, the environmental signals can be inferred because the trajectories are driven to different regions. This is the property of state space separation because the trajectories are driven to different regions of the substrate state space. For such substrates there is no need to use sophisticated artificial intelligence (AI) systems to infer the state of the environment. For example, if a substrate has a response that resembles the one in Fig. 1b, then a linear decision boundary could be used to infer the environment: if the state is to the right (left) of the linear boundary, then, the environmental condition is q_1 (q_2). For such a device, most of the computation takes place in the substrate and the process of reading the reservoir state should be seen as a supplement with low computational complexity imposing modest engineering overhead. It is important to realize that the behavior depicted in Fig. 1b does not occur naturally. One is more likely to encounter the behavior shown in Fig. 1a. In fact, one has to work hard to find a drive signal that results in the behavior depicted in Fig. 1b.

In the figure and throughout the text, a special notation has been used. To emphasize that a quantity s depends on the history of u and q up to a time instance t one writes $s(t) = F[u, q](t)$ where F is a functional that describes this dependence.

In a sensing classification problem, the task of a sensor would be to classify k environmental conditions $c_1, c_2, \cdots, c_k$. Every condition is represented as a class of typical environmental signals $c_i = \{q_1^i, q_2^i, \cdots, q_{N_i}^i\}$ where $i = 1, 2, \cdots k$. If the device is exposed to an environmental signal belonging to class c_i, e.g. q_j^i with $j \in \{1, 2, \ldots, N_i\}$, then the readout layer should report a corresponding class label l_i. In the general case of considering many environmental conditions, state space separation can be achieved if the trajectories are driven to different regions of the state space $\Omega_1, \Omega_2, \cdots, \Omega_k$ when the network is exposed to the different environmental conditions $c_1, c_2, \cdots c_k$. [6, 5, 13]

3 The substrate model

Figure 2 depicts the substrate model of interest: the network of environment-sensitive memristor elements. A memristor network used for classification consists of N_R environment-sensitive memristors with resistances $R_1, R_2, \cdots, R_{N_R}$. The state of the network under a drive signal u and an environmental condition q_j^i is written as:

$$S[u, q_j^i](t) \equiv (R_1[u, q_j^i](t), R_2[u, q_j^i](t), \cdots, R_{N_R}[u, q_j^i](t)) \quad (1)$$

and evolves in time depending on the environmental condition the device has been exposed to and the drive applied. The sensor consists of a memristor network substrate equipped with a readout layer that is used to assess the state of the network and infer the class of environmental condition that the network is exposed to.

Multi-memristors substrates are interesting since two or more memristors in the same network can collaborate to improve the state space separation. This is possible since in a network the voltage difference across a memristor depends on the resistances of memristors somewhere else in the network. Therefore, the state of one memristor depends on the states of all the other memristors and they respond to the external drive signal as a group. These dependencies are a trademark of complex dynamical systems. In this study we view them in a very special way: These interactions realize an indirect communication between memristor elements and can distribute the task of the state space separation among the memristor elements.

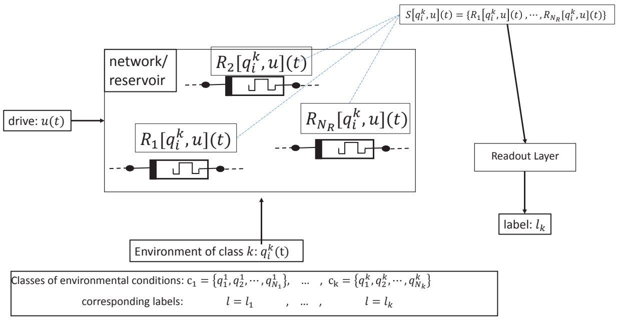

Figure 2. Sensing as the classification of an environmental conditions q_i^k with the label l_k. The network is exposed to the environmental condition q_i^k. The state $S[q_i^k, u](t)$ is given as input to the readout layer. Based on the information stored in the state the readout layer should infer the label l_k to the environmental condition q_i^k. The drive signal $u(t)$ is used to improve the state space separation.

To describe the dynamics of a memristor, a model is chosen that is realistic but not too complicated to work with. The model used in the study is built on the original Pershin-Di Ventra memristor model [14], augmented with the dependence on the environmental signal as described in [2]. If the voltage ΔV is being applied across the memristor model, the value of the memristance $R(t)$ changes as

$$\dot{R}(t) = f(\Delta V(t), \beta(t))\Theta(R(t), \Delta V(t)) \tag{2}$$

with

$$f(\Delta V, \beta) = \beta \Delta V + \frac{1}{2}\,(\alpha - \beta) \cdot \left(|\Delta V + V_{thr}| - |\Delta V - V_{thr}|\right) \tag{3}$$

and

$$\Theta(R, \Delta V) = \begin{cases} 0, & \text{if } \Delta V = 0 \\ \theta(R_{max} > R), & \text{if} \Delta V > 0 \\ \theta(R > R_{min}), & \text{if} \Delta V < 0 \end{cases}$$

where $\theta(R_{max} > R)$ is zero unless the condition in the argument is satisfied, and likewise for $\theta(R > R_{min})$. Additionally, V_{thr} is the threshold voltage, and R_{min} and R_{max} are the minimum and the maximum memristance values introduced earlier. The parameter α (β) describes the rate of memristance change in the regions $|\Delta V| < V_{thr}$ ($|\Delta V| > V_{thr}$).

By assumption, the memristor "feels" the environment through the parameter β:

$$\beta(t) = q(t) \tag{4}$$

where $q(t)$ is a time-series that describe the environmental condition. The rest of the parameters are assumed to be independent of the environment, but this may not be the case in reality. The threshold voltage might be environment dependent too. For a device used for ion sensing, the presence of ions in the solution usually screens the surfaces from the external electrical field. As a result, the voltages felt by the device will be different from the external applied voltages. A more detailed discussion regarding the validity of the model can be found in [2].

4 Learning the drive signal

To recognize the environmental condition c_i which the substrate model is exposed to, it is needed to find the optimal drive signal u^* so that the degree of state space separation is maximised. To measure the degree of state space separation, we use a separability index ν.

The separability index ν is obtained by computing a typical distance between trajectories. This typical distance is obtained if one averages the trajectories over time first and then compute the Euclidean norm. The algorithm used to compute the typical distance between two trajectories $S[u, q_j^i]$ and $S[u, q_{j'}^{i'}]$, $i \neq i'$

$$d_{j,j'}^{i,i'} = \frac{1}{\sqrt{N_R}} \|\bar{S}[u, q_j^i] - \bar{S}[u, q_{j'}^{i'}]\| \tag{5}$$

where

$$\bar{S}[u, q_j^i] \equiv (\bar{R}_1[u, q_j^i], \bar{R}_2[u, q_j^i], \cdots, \bar{R}_{N_R}[u, q_j^i]) \tag{6}$$

with

$$\bar{R}_m[u, q_j^i] = \frac{1}{T}\int_0^T dt R_m[u, q_j^i](t) \tag{7}$$

The above equations result in the following compact expression

$$d_{j,j'}^{i,i'} = \sqrt{\frac{1}{N_R}\sum_{m=1}^{N_R}\left(\bar{R}_m[u, q_j^i] - \bar{R}_m[u, q_{j'}^{i'}]\right)^2} \tag{8}$$

The typical distances between trajectories are combined into an overall measure. The separability index is calculated as the geometric mean over all the typical distances:

$$\nu[u; c_1, c_2, \cdots, c_k] = \left(\prod_{i=1}^{k}\prod_{i'=i+1}^{k}\prod_{j=1}^{N_i}\prod_{j'=1}^{N_{i'}} d_{j,j'}^{i,i'}\right)^{\frac{1}{N_D}} \tag{9}$$

where N_D denotes the number of all possible distances in the set of the training data. This number is given by:

$$N_D = \sum_{i=1}^{k}\sum_{i'=i+1}^{k}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_{i'}} 1 \tag{10}$$

The geometric mean has been chosen for calculating the separability index against the arithmetic mean because we prefer that all the distances are fairly large. By using the geometric mean, a very small distance contributes to a small final product. If we used the arithmetic mean, a very small distance would not appropriately contribute to a small sum given that another large distance exists. The optimisation problem is expressed mathematically as:

$$u_* = \arg\max_u \nu[u; c_1, c_2, \cdots, c_k] \tag{11}$$

The optimal drive signal is found by using a genetic algorithm (GA) optimisation where the separability index is used as a fitness function. We have developed the genetic algorithm in a previous work [3]. GA is a strong optimisation technique that can be used to solve problems regardless of their complexity [15, 16, 1]. GA is used in this work instead of gradient based optimisations because gradients of the fitness have not been calculated analytically. Additionally, GA is advantageous in cases where there is no prior knowledge of the fitness function form. For example, it is not possible to know beforehand if the suggested fitness function (separability index) is convex.

The drive signal $u(t)$ is represented as a Fourier series:

$$u(t) = a_0 + \sum_{i=1}^{N_c} a_i \sin(i\omega t) + \sum_{i=1}^{N_c} b_i \cos(i\omega t) \tag{12}$$

In the GA scheme the signal u is encoded as the set of $2\,N_c + 2$ parameters

$$P_u = \{a_0, a_1, ..., a_{N_c}, b_1, ..., b_{N_c}, \omega\} \tag{13}$$

When there are delay feedback mechanisms in the network, their delay times are optimised too. In principle one could try to optimise the full list of the time delays: $\tau, \tau_2, \tau_3, \cdots$ However, the first time delay τ is kept fixed since we noticed that the optimal frequency ω and the time delays $\tau_2, \tau_3, \cdots$ are strongly dependent on the value of τ. If τ is changed, other parameters adjust to τ so that there is no overall change in the systems behavior. For example, in a case of a network with two delay times the set of parameters to optimise is given by $P_{opt} = P_u \cup P_\tau, P_\tau = \{\tau_2\}$. In a case with three time delays we use the set $P_{opt} = P_u \cup P_\tau, P_\tau = \{\tau_2, \tau_3\}$ etc. In general, we have these two cases:

- $P_{opt} = P_u$, when there are no more than one delay feedback mechanisms,
- $P_{opt} = P_u \cup P_\tau$, when there are more than one delay feedback mechanisms.

The goal is to find the optimal choice of parameters P^*_{opt} for a fixed training dataset $c_1, c_2, \cdots, c_k$ (and a fixed network) by maximizing the separability index:

$$P^*_{opt} = \arg\max_{P_{opt}} \nu[P_{opt}; c_1, c_2, \cdots, c_k] \tag{14}$$

The genetic algorithm is used to identify a globally optimal solution within the given bounds for P_{opt} and for the specific network under consideration. In somewhat simpler terms, the genetic algorithm works as follows. The network of interest is simulated[3] in a time interval $[0, T]$ for a wide range of P_{opt} and environmental conditions, which produce the typical trajectories discussed earlier. From those trajectories the separability index ν can be calculated (Eq. 9). The values for ν are sorted and finally the solution P^*_{opt} is found as the one with the largest separability index ν.

Since the index ν is maximized by using the training data, we will refer to it as the training data index:

$$\nu_{\text{train}} = \nu[P^*_{opt}; c_1, c_2, \cdots, c_k] \tag{15}$$

Furthermore, to evaluate the ability of the network to generalize, one should test its performance on some other set of environmental conditions that the network has not been trained for. In practical terms this is hard to do since we are not explicitly training any readout layer. To estimate the ability to generalize we compute the index ν with the optimised parameters P^*_{opt} but with a different set of environmental conditions $c'_1, c'_2, \cdots, c'_k$:

$$\nu_{\text{test}} = \nu[P^*_{opt}; c'_1, c'_2, \cdots, c'_k] \tag{16}$$

The test data set $c'_1, c'_2, \cdots, c'_k$ is generated with similar features as the training data set $c_1, c_2, \cdots, c_k$ but with a much bigger number of signals per class. The related separability index will be referred to as the test data index ν_{test}. In the following, the symbol ν will be used to denote both ν_{train} and ν_{test}.

5 Strategies for increasing network complexity

To study the collaboration between memristor elements in a systematic way, a range of memristor networks is investigated with an increasing degree of complexity. As it is shown in Fig. 3, we choose three ways of increasing that complexity.

[3] In this work the simulator was developed in Java programming language by using modified nodal analysis and the concepts of simulating memristor networks given in [9].

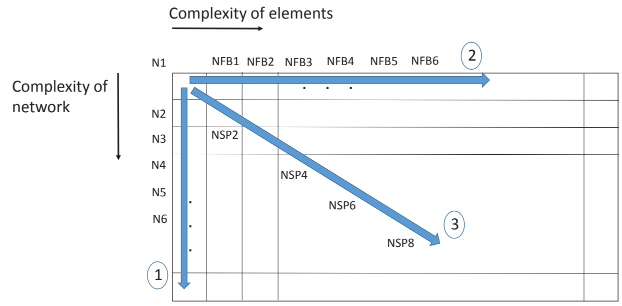

Figure 3. Three essential ways of increasing the network complexity. In direction 1, we increased the complexity of the network topology. In direction 2, we added to the networks elements with increased complexity. In direction 3, we increased the complexity of elements by providing heterogeneous memristor elements and simultaneously we increased the complexity of the network topology.

Direction 1: An example of increasing the complexity of the memristor network topology is shown in Fig. 4. The network N1 is a simple one-memristor network studied earlier [2]. The networks N2, N3, N4, N5 and N6 are samples of more complicated networks with an increasing number of memristors and various combinations of serial and parallel couplings.

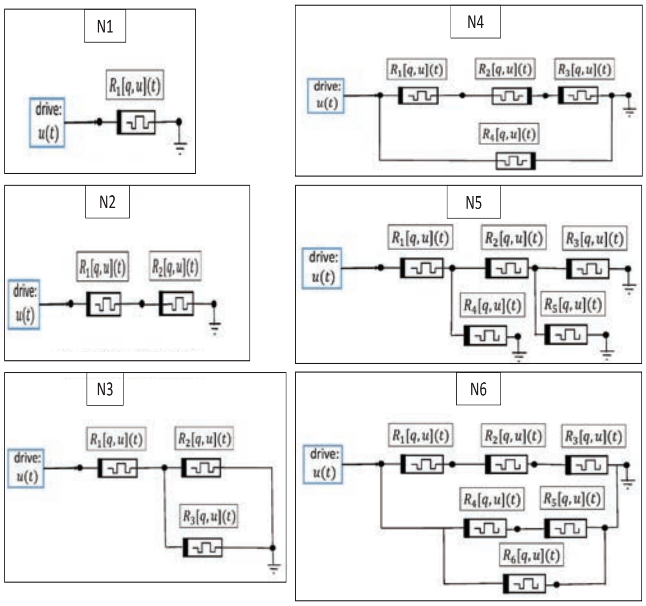

Figure 4. The memristor networks we used to examine how the increased network topology can favor the state separation.

Direction 2: An example of adding elements with an increasing degree of complexity is shown in Fig. 5. The network NFB1 consists of a memristor feedback (MFB) unit, being essentially a memristor equipped with a delay feedback mechanism with time delay τ and a summation (SUM) unit. The value of memristance $R_1(t)$ is given as input to the delay feedback loop. This loop has an internal queue which stores the earlier memristance values: $R_1(t), R_1(t - dt), R_1(t - 2dt) \cdots, R_1(t - \tau)$. The last memristance value in this list, $R_1(t - \tau)$, is converted to the voltage $V_{\text{fb}}(t - \tau)$ according to an internal resistance to voltage mapping. The voltage $V_{\text{fb}}(t - \tau)$ is the output of the delay feedback mechanism. The drive signal $u(t)$ and the voltage $V_{\text{fb}}(t - \tau)$ are given as input to the SUM unit which adds them up and feeds as a modified input at the time t.

One expects that the delay feedback mechanism adds memory properties to the network. The memristance $R_1(t)$ depends on the

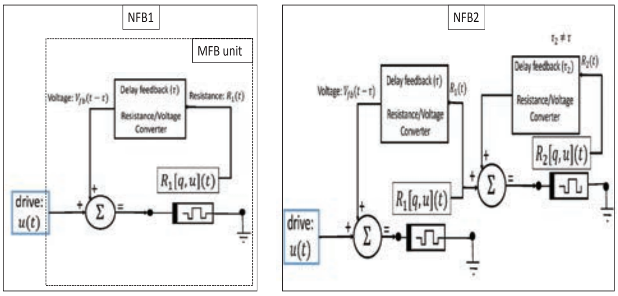

Figure 5. The networks NFB1 and NFB2 were used to examine whether increasing the complexity of the network element favors the state space separation. An MFB unit is added in series to the network NFB1 so as to form the network NFB2. Similarly, by adding an MFB unit in series to the NFB2, the network NFB3 was formed, by adding an MFB unit in series to the NFB3, the network NFB4 was formed, etc., until the formation of the network NFB6.

time series of the drive signal u, the time series of the environmental condition q and on all the past values of the memristance R_1 because of the memristor memory properties. Additionally, when adding the delay feedback mechanism, the current state has an extra dependence on the past values $\{R_1(t-\tau), R_1(t-2\,\tau), R_1(t-3\,\tau), \cdots\}$.

Additionally, in Fig. 5, the network NFB2 is formed by adding in series an MFB unit to network NFB1. The delay feedback mechanism of the added MFB unit has a delay $\tau_2 \neq \tau$. The memristance $R_2(t)$ has an extra dependence on the memristances $R_1(t), R_1(t-1\,\tau), R_1(t-2\,\tau), \cdots$ because of the connection of the two MFB units. Additionally, the memristance $R_2(t)$ has an extra dependence on the past memristances $R_2(t-1\,\tau), R_2(t-2\,\tau), \cdots$ because of the delay feedback with time τ_2. Therefore, the memristance $R_2(t)$ has a memory of both the memristances R_1 and R_2.

The freedom to choose the time delays of the two MFB units could allow for completely different time series of voltage differences across the memristors. These different memristor inputs ought to favor the collaboration between memristors because every memristor would have different tasks of state separation.

In the same way by adding MFB units in series, we construct the networks NFB3 (three MFB units in series, delay times: τ, τ_2, τ_3), NFB4 (four units in series, delay times: $\tau, \tau_2, \tau_3, \tau_4$), up to NFB6 (six units in series, delay times: $\tau, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6$). These networks are not shown.

Direction 3: The complexity of the elements can be also increased by taking advantage of the memristor variability [17, 12]. If two memristors have different parameters β, then, these differences can be exploited so that each memristor can be used for processing different features of the environmental conditions, *e.g.* occurring at different scales.

Additionally, in an experimental scenario, two or more memristors might sense simultaneously different instantaneous values of q because they are placed at different positions. In such a case, the parameter β of each memristor would be affected differently.

We model the variability by assuming:

$$\beta(t) = (1 + m_0)\, q(t) + m_0 \tag{17}$$

where m_0 is unique for every memristor element and randomly chosen between 0 and 1.

In direction 3, the complexity of network topology increases by combining memristor elements with series and parallel couplings. The purpose is to build networks where the different memristor elements experience different local voltage differences. The question is whether such structures, with variability in local voltage drives could favor the collaboration in the network. The networks NSP2, NSP4, NSP6 and NSP8, being shown in Fig. 6, are examples.

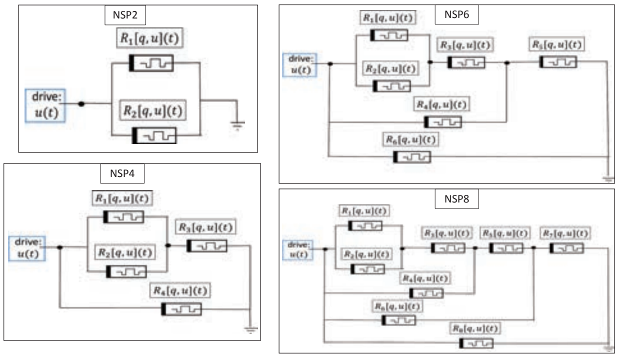

Figure 6. The networks NSP2, NSP4, NSP6 and NSP8 were used in direction 3 to investigate whether considering both inhomogeneous memristor elements and increased network topology favors the state space separation.

6 Results

An elementary two-class problem is investigated where two environmental states are allowed: stable or varying. The training signals that describe each of these classes are shown in Fig. 7. Each class is represented by 10 signals. As test data, 1000 environmental signals have been generated for each class. The signals have been generated by assuming a certain periodicity and then by sampling the period and the coefficients of the Fourier series. The series with four coefficients have been used. The coefficients have been sampled from two uniform distributions with variances σ_1 (varying) and σ_2 (stable) such that $\sigma_1 \gg \sigma_2$.

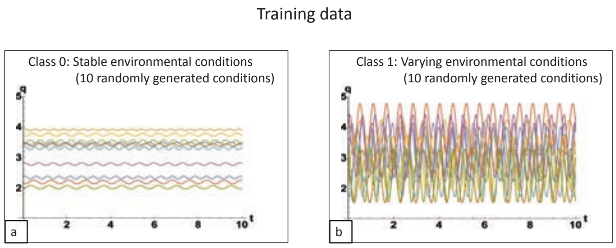

Figure 7. a) The training data for the class 0: stable environmental conditions. Ten environmental conditions were generated randomly. b) The training data for the class 1: varying environmental conditions. Ten environmental conditions were generated randomly.

Circa $5,000$ virtual experiments per network are conducted to identify the optimal mode of operation for each network. For simpler networks, histograms of memristance values are built to visualize the state space separation. However, the visualization of the histograms is not easy when the state space dimensionality is large. For more complicated networks ($N_R > 2$), we compute only the separability index ν. All separability index values are expressed in units of $\Delta M = R_{max} - R_{min}$ where ΔM is the maximum distance between the trajectories:

$$0 \leq \nu \leq \Delta M \tag{18}$$

The relative memristance $\nu/\Delta M$ is always bound between 0 and 1.

The drive signals of all simulated networks are optimised with the same number of parameters, ten in total: α_0, α_1, α_2, α_3, α_4, β_1, β_2, β_3, β_4, ω. One could increase the number of the parameters and possibly achieve a better separability index. But this would be a drawback since the optimisation would need to search for an optimum solution in a larger space of parameters. The goal is to achieve larger separability indices not by introducing smarter signals but by making the reservoir, *i.e.* the memristor network, smarter.

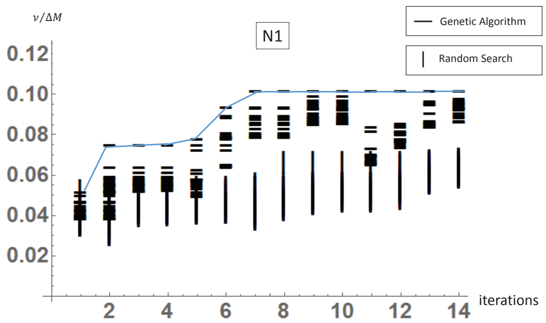

Figure 8. The performance of the genetic algorithm is compared to a random search algorithm. At every iteration the best solutions of each algorithm are presented. The index ν denotes the separability index and the value ΔM is the maximum value that the index ν can obtain.

Starting with the simplest network topology (N1), the results for every iteration of the genetic algorithm optimisation are shown in Fig. 8. One can see that to increase the separability index, it is important to find an appropriate drive signal. At every iteration of the algorithm the separability index was calculated for 300 candidate solutions and 30 solutions with the largest separability index were used for genetic operations at the next iteration. The best drive signal was found after six iterations. The GA performs clearly better than the random search algorithm. At every iteration of the random search algorithm, 300 random solutions were evaluated. The twenty best after each iteration are shown. After thirteen iterations, the random search algorithm has not found a drive signal with equally good separability index as the one found by using the genetic algorithm.

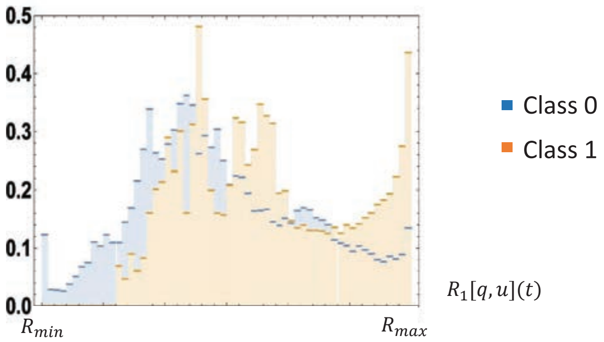

Figure 9. The optimum state separation for the network N1 found by the optimisation algorithm. The two distributions depict the histograms of the memristance values collected during the simulations of the system under the two classes of environmental conditions of the training data.

We recorded the values of the memristances (states of the system) under the two different environments, class 0 and class 1, and the best drive. The histograms in Fig. 9 show how the values are distributed. The trajectories are driven to a region closer to R_{min} under the environment 0 and closer to R_{max} under the environment 1. Clearly, the external drive tries to separate the states to the best of its abilities, but cannot achieve a perfect state space separation because the distributions in the histogram are still overlapping.

Direction 1: To examine whether the process of increasing the complexity of the network topology leads to larger separability indices, the separability index ν has been computed for the networks N1, N2, N3, N4, N5 and N6. The values of ν are shown in Fig. 10. On the vertical axis we plot the relative memristance.

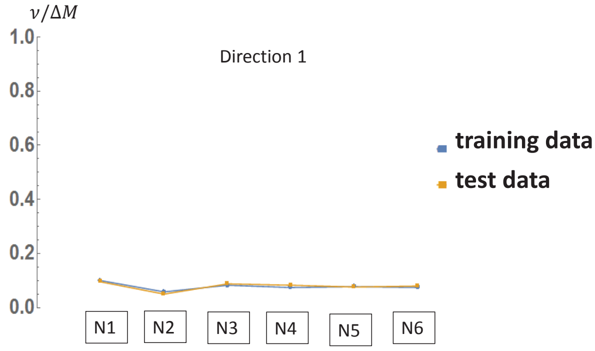

Figure 10. The separability index for the training data and the test data for the networks N1, N2, N3, N4, N5 and N6. All networks are homogeneous: They are made of a single memristor element.

The key result: The separability index ν did not increase (improve) when the complexity of the network topology was increased. There was not enough collaboration between the memristor elements, even when the networks with up to 6 memristor elements were considered (from N2 up to N6). To identify whether a genuine collaboration exists between the elements, all the memristor elements were chosen to be identical (the results for heterogenous networks are discussed below).

Direction 2: A few networks with (heterogenous) delayed-feedback elements have been studied: NFB1 (one feedback element), NFB2 (two feedback elements), NFB3, NFB4, NFB5 and NFB6.

Network NFB1 features only one feedback element, the MFB unit explained earlier. The optimal state såpace separation achieved is shown in Fig. 11. By coarse graining and dividing the state in two intervals, left half region $[R_{min}, (R_{min} + R_{max})/2]$ and right half region $[(R_{min}+R_{max})/2, R_{max}]$, the following can be observed: Under the environment class 0, the 60% of the trajectories were driven to the left interval, while the rest of the trajectories (40%) were driven to the right interval. However, under the environment class 1, the 100% of the trajectories were driven to the right half. Therefore, the state space separation is not perfect, but it is better than the one achieved with the (feedback-free) network N1.

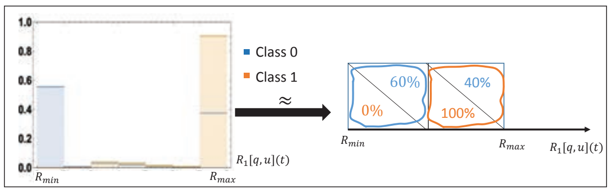

Figure 11. The optimum state space separation for the network NFB1. The two distributions have been obtained using the optimum drive signal. The optimum state space separation is achieved with 60% (100%) of the state driven to the left (right) half region of the state space for environmental conditions 0 (1).

The state space separation for network NFB2 was studied to see whether the state space separation of NFB1 can be improved further by adding another MFB unit. The histograms in Fig. 12 show how the states are distributed for the two environmental conditions. By coarse graining and dividing the state in four regions: up-right, up-left, down-right and down-left, we found the following:

- Under the environment 0, 60% of the trajectories were driven to the up-right square region of the state space, 30% were driven to the down-left square region and 10% to the up-left square.
- Under the environment 1, 70% of the trajectories were driven to the up-left square region of the state space, 20% were driven to the down-right square region and 10% to the down-left square.

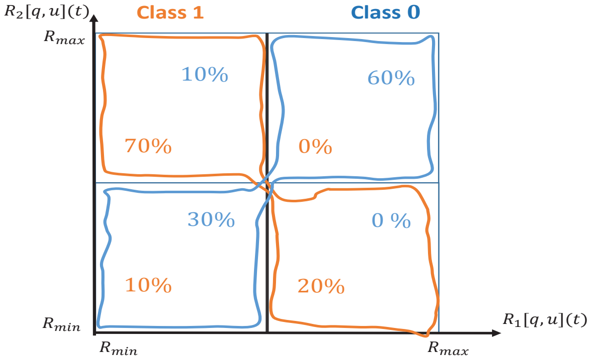

Figure 12. The optimum state space separation for the network NFB2 obtained for the training data. We coarse-grain the state space into four regions: 90% (90%) of the trajectories are driven to the up-right or down-left (up-left or down-right) quarters of the state and the network is exposed to the environmental conditions of class 0 (1).

Therefore, if the state was read in the up-right or in the down-left square (in the up-left or in the down-right square), then a read-out layer could be constructed to classify the environment as class 0 (1). Still, the state space separation is not ideal in the up-left and down-left regions. For example, an ideal case would be that under the environment class 0 (1), 100% of the trajectories were driven to the up-right or down-left square (up-left or down-right square) of the state space.

Networks NFB3, NFB4, NFB5 and NFB6 contain three or more MFB units. Their separability indices are shown in Fig. 13.

- The separability index ν_{train} was better for NFB1 than for N1. The state space separation increased by adding a delay feedback mechanism to a single memristor element. The index ν_{test} was roughly the same for NFB1 and N1.
- Both ν_{train} and ν_{test} increased when adding MFB units in series to network NFB1 resulting in NFB2, NFB3 and NFB4. There is more collaboration among elements when adding MFB units.
- When constructing the networks NFB5 and NFB6, the separability index ν was similar to the one of the network NFB4. This means that when adding more MFB units in the direction from NFB4 to NFB6 there was little additional information about the environment.
- The largest allowed value of $\nu/\Delta M$ equals 1. The observed maximum value of $\nu/\Delta M$ is roughly 0.7.
- Four MFB units (NFB4) were adequate to reach the maximum separability index ν, while connecting five or six MFB units in series (NFB5 or NFB6) would be a waste of resources. Therefore, the graph in Fig. 13 shows the minimum amount of resources to reach the maximum separability index.

- ν_{test} increased when adding MFB units in the direction from NFB1 to NFB6.
- ν_{test} was lower than ν_{train} for all the networks NFB1 - NFB6.
- ν_{test} did not increase as much as ν_{train} in the direction NFB1-NFB3.
- ν_{test} increased more than ν_{train} in the direction NFB3-NFB4.
- Figure 13 suggests that the state space separability is easier to achieve in large dimensions: networks NFB4-NFB6 have the largest separability index ν_{test}. It seems that in large dimensions the system has a better ability to generalize: network NFB4-NFB6 exhibit a relatively small difference between ν_{test} and ν_{train}.

Direction 3: The separability indices for the networks NSP2, NSP4, NSP6 and NSP8 are shown in Fig. 14.

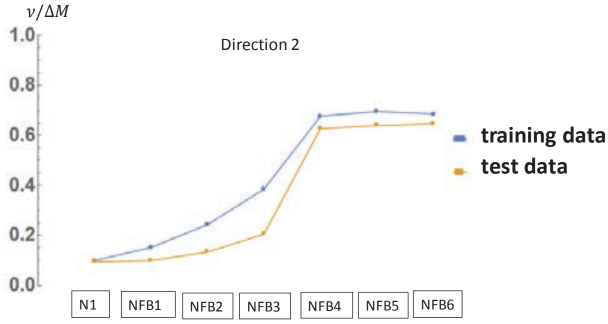

Figure 13. The maximum separability index for the training data and the testing data for the networks N1, NFB1, NFB2, NFB3, NFB4, NFB5 and NFB6.

- The training data separability index ν_{train} was found slightly larger for NSP2 than N1. This means that two heterogeneous parallel memristors worked better than one memristor. However, we found that by just putting in parallel memristors did not result in increasing ν_{train}. We tried three heterogeneous memristors in parallel and the separability index was worse than the one for N1.
- Increasing the size of the networks resulted in larger ν_{train}.
- The same behavior was not observed for ν_{test}. Firstly, ν_{test} for NSP4 was smaller than the ones for NSP2 and N1. Secondly, ν_{test} for NSP6 was larger than the one for NSP4.

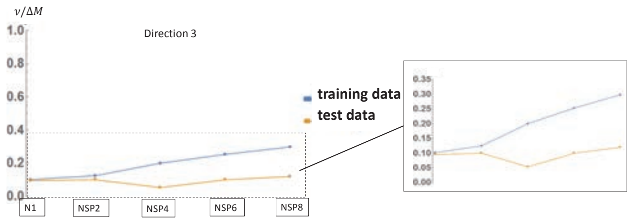

Figure 14. The maximum separability index for the training data and the testing data for the networks N1, NSP2, NSP4, NSP6 and NSP8.

7 Discussions

Judging by the values of ν_{test} and ν_{train}, it seems that memristor networks have a good ability to generalize. As expected, all networks performed worse on the test data with $\nu_{\text{test}} < \nu_{\text{train}}$. However, for some networks, ν_{test} was remarkably close to ν_{train} even with such a huge difference between the size of the two data sets, being in the 1:100 ratio: the training data set contains 20 time series, and the test data set contains 2000. Further, the test data index ν was more sensitive to the increase of the state space dimension (NFB4, NFB5 and NFB6): For a larger state space dimension there were more chances that the trajectories would be driven to different regions of the state space under the two environmental conditions. This is also the reason why the difference $\nu_{train} - \nu_{test}$ decreases as the dimension of the state space increases following the direction NFB4, NFB5 up to NFB6.

Both ν_{test} and ν_{train} roughly follow the same trend when computed for different networks. Thus the findings discussed in the previous subsections ought to hold for any data set. An interesting finding concerns perhaps a naive expectation that elements with a larger degree of complexity should lead to better substrates (direction 2 in Fig. 3). It is true that when adding elements with increased complexity there is an increase in the separability index ν, but only for more complex networks. Figure 13 illustrates this behavior. An increase in ν is marginal between N1 (a single memristor with no feedback) and NFB1 (a single memristor with feedback). It is true that for more complex networks there is a substantial increase in ν (e.g. going through NFB2, NFB3, NFB4). However, it seems that this trend does not continue *ad infinitum* (e.g. moving to NFB5 or NFB6 does not further increase ν). The question is whether the observed behavior is generic or just an artefact of following a particular way of adding elements into the network.

Our hypothesis is that by just adding more complex elements into the network will not necessarily lead to a better sensor. To obtain a better sensor it is necessary to ensure that there is an advantageous collaboration between the elements, it is important how elements work as a group. We showed a way of achieving this by optimising the time delays of the MFB units. When the MFB units have different time delays then each one can react to different features of the environmental conditions, i.e. to collaborate and divide the burden of state separation. This further suggests that heterogeneity of the networks is important: it is essential to allow for different time delays of the MFB units. Note that the memristor elements of all the MFB units were considered identical. Thus the heterogeneity of the time delays was the main reason for achieving collaboration between the memristor elements.

When increasing the complexity of the network topology (direction 1 in Fig. 3) there was no improvement in the separability index ν. This means that in this direction we did not find a network structure where the collaboration between the memristor elements favors a larger separability index ν. This shows that the state space separability is a non trivial requirement. Note that the elements of the networks were identical. Our hypothesis is that the increased complexity of the network topology would have a better effect on the separability index when considering heterogeneous memristor elements.

8 Conclusions

A separability index ν has been used as a measure of substrate quality. A good separability index should measure the degree of correlation between the information embedded in the time-series signal (environment) and the system state (sensor). Several difficulties related to a suitable definition of this index have been identified and solved. We found a way of defining a separability index without considering a specific analysis (readout) layer. Further, the separability index has been constructed to detect a genuine collaboration between elements. If an element is added into a network, the dimensionality of the state space increases and it should be easier to separate trajectories for different environmental conditions. However, there are ways of adding elements into the network that will not lead to a larger trajectory separation. For example, if an identical element is added into the network that copies the functionality of an existing element, the "intelligence" of the substrate will not increase. In such a case, a good separability index should not report a false improvement.

As an alternative to the separability index suggested here, the mutual information could be used to quantify how much information about the environment is stored in the system state. However, the computational cost associated with estimating the mutual information can be high, rendering the approach impractical in some cases. [10] An important advantage of the approach advocated here is that the computational cost of calculating ν is relatively modest when compared to the mutual information approach.

By computing the separability index for a range of memristor networks, with an increasing degree of complexity, we were able to identify which architectural features of such networks guarantee good substrate quality. The presence of feedback loops and increased degree of network heterogeneity have the most impact on the sensing capacity of the substrate. The presence of feedback loops leads to faster responses and more precise decision boundaries. Adding identical elements does not improve the sensing capacity. Such elements tend to interact in the same way with the environment without providing additional clues about it. If heterogeneous elements are linked into a network, each element can recognise a different environmental feature.

This work can be extended in various ways. The approach suggested here could be applied in situations where the substrate response is not simulated but measured in the actual experiment. However, finding the right drive could be problematic. Further, the network space has been explored using the trial and error method. However, one could develop an automatic GA optimisation procedure to identify the best performing networks and then study them to reverse-engineer useful design principles. Such a GA could be naturally implemented by using the separability index as the fitness function. While the numerical experiments were conducted on memristor networks, the principles established are universal and provide useful guidelines for designing similar sensing devices with other types of hardware. The non-linear behaviour of the memristor could be mimicked by other non-linear elements, such as artificial neurons or other non-linear electronic components. In this study, ideas from various disciplines were exploited including theoretical and applied computer science (neuromorphic computation, computing capacity), the theory of complex dynamical systems (state space structure), signal engineering (filters) and machine learning (supervised learning). The results of this work may be of considerable interest to the wider information processing community.

ACKNOWLEDGEMENTS

The authors wish to acknowledge a series of discussions with all partners in the Horizon 2020 RECORD-IT consortium. These discussions contributed greatly in formulating the ideas presented in this work. We would like to thank our colleagues Per Lundgren, Per Rudquist, and Dag Winkler for their suggestions, which helped us improve the text.

REFERENCES

[1] O. Abedinia, M. S. Naderi, A. Jalili, and B. Khamenehpour, 'Optimal tuning of multi-machine power system stabilizer parameters using genetic-algorithm', in *2010 International Conference on Power System Technology*, pp. 1–6, (Oct 2010).

[2] V. Athanasiou and Z. Konkoli, 'On using reservoir computing for sensing applications: exploring environment-sensitive memristor networks', *International Journal of Parallel, Emergent and Distributed Systems*, (2017).

[3] V. Athanasiou and Z. Konkoli, 'On sensing principles using temporally extended bar codes', *IEEE sensors Journal*, (2020 in press).

[4] L. Chua, 'Memristor-the missing circuit element', *IEEE Transactions on Circuit Theory*, **18**(5), 507–519, (Sep. 1971).

[5] T. E. Gibbons, 'Unifying quality metrics for reservoir networks', in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, (July 2010).

[6] E. Goodman and D. Ventura, 'Spatiotemporal pattern recognition via liquid state machines', in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp. 3848–3853, (2006).

[7] Z. Konkoli, 'The state weaving environment-echo tracker (sweet/record-it) sensing setup and algorithm (uspto, patent pending, 62/319972)', (2016).

[8] Z. Konkoli, 'The sweet algorithm: generic theory of using reservoir computing for sensing applications', *International Journal of Parallel, Emergent and Distributed Systems*, (2016).

[9] Z. Konkoli and G. Wendin, 'A generic simulator for large networks of memristive elements', *NANOTECHNOLOGY*, **24**, (2013).

[10] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger, 'Estimating mutual information', *Physical Review E*, **69**(6), 066138, (2004).
[11] G. Monte, V. Huang, P. Liscovsky, D. Marasco, and A. Agnello, 'Standard of things, first step: Understanding and normalizing sensor signals', in *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, pp. 118–123, (Nov 2013).
[12] R. Naous, M. Al-Shedivat, and K. N. Salama, 'Stochasticity modeling in memristors', *IEEE Transactions on Nanotechnology*, **15**(1), 15–28, (Jan 2016).
[13] D. Norton and D. Ventura, 'Improving the separability of a reservoir facilitates learning transfer', in *2009 International Joint Conference on Neural Networks*, pp. 2288–2293, (June 2009).
[14] Y. V. Pershin and M. Di Ventra, 'Experimental demonstration of associative memory with memristive neural networks', *Neural Networks*, **23**(7), 881–886, (2010).
[15] Heidar Ali Shayanfar, Hossein Shayeghi, Oveis Abedinia, and Aref Jalili, 'Design rule-base of fuzzy controller in multimachine power system stabilizer using genetic algorithm', in *Proceedings of the 2010 International Conference on Artificial Intelligence, ICAI 2010, July 12-15, 2010, Las Vegas Nevada, USA, 2 Volumes*, pp. 43–49, (2010).
[16] H. Shayeghi, H. A. Shayanfar, and O. Albedinia, 'Fuzzy pss design for a multi-machine power system using improved genetic algorithm', *Computer Science and Engineering*, (2012).
[17] S. Smaili and Y. Massoud, 'Analytic modeling of memristor variability for robust memristor systems designs', in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 794–797, (June 2014).

Morphological Computation and Learning to Learn In Natural Intelligent Systems And AI

Gordana Dodig-Crnkovic [1,2]

Abstract. At present, artificial intelligence in the form of machine learning is making impressive progress, especially the field of deep learning (DL) [1]. Deep learning algorithms have been inspired from the beginning by nature, specifically by the human brain, in spite of our incomplete knowledge about its brain function. Learning from nature is a two-way process as discussed in [2][3][4], computing is learning from neuroscience, while neuroscience is quickly adopting information processing models. The question is, what can the inspiration from computational nature at this stage of the development contribute to deep learning and how much models and experiments in machine learning can motivate, justify and lead research in neuroscience and cognitive science and to practical applications of artificial intelligence.

1 INTRODUCTION

This paper explores the relationships between the info-computational network based on morphological computation and the present developments in both the sciences of the artificial (with the focus on deep learning) as well as natural sciences (especially neuroscience, cognitive science and biology), social sciences (social cognition) and philosophy (philosophy of computing and philosophy of mind).

Deep learning is based on artificial neural networks resembling neural networks of the brain, processing huge amounts of (labelled) data by high-performance GPUs (graphical processing units) with a parallel architecture. It is (typically supervised) machine learning from examples. It is static, based on the assumption that the world behaves in a similar way and that domain of application is close to the training data. However impressive and successful, deep-learning intelligence has an Achilles heel, and that is lack of common sense reasoning [5][6][7]. It bases recognition of pictures on pixels, and small changes, even invisible for humans can confuse deep learning algorithm and lead to very surprising errors.

According to Bengio, deep learning is missing out of distribution generalization, and compositionality. Human intelligence has two distinct mechanisms of learning – quick, bottom up, from data to patterns (System 1) and slow, top-down from language to objects (System 2) which have been recognized earlier [8][9][10]. The starting point of old AI (GOFAI) was System 2, symbolic, language, logic-based reasoning, planning and decision making. However, it was without System 1 so it ended in symbol grounding problem. Now deep learning has grounding for its symbols in the data, but it lacks the System 2 capabilities in order to get to the human-level intelligence and ability of learn and meta-learning, that is learning to learn.

The step from big-data based System 1 to manipulation of few concepts like in high level reasoning is suggested to proceed via concepts of agency, attention and causality.

It is expected that agent perspective will help to put constraints on the learned representations and so to encapsulate causal variables, and affordances. Bengio proposes that "meta-learning, the modularization aspect of the consciousness prior [7] and the agent perspective on representation learning should facilitate re-use of learned components in novel ways (even if statistically improbable, as in counterfactuals), enabling more powerful forms of compositional generalization, i.e., out-of-distribution generalization based on the hypothesis of localized (in time, space, and concept space) changes in the environment due to interventions of agents." [5]

This step, from System 1 (present) to System 2 (higher level cognition will open new and even more powerful possibilities to AI. It is not the development into the unknown, as some of it has been attempted by GOFAI, and new developments in cognitive science and neuroscience. In this article we will focus on the connections to another computational model of cognition, natural infocomputation [3][4].

1 LEARNING ABOUT THE WORLD THROUGH AGENCY

When discussing cognition as a bioinformatic process of special interest, we use the notion of *agent*, i.e. a *system able to act on its own behalf* [11]. Agency in biological systems in the sense I use here has been explored in [12][13]. The world as it appears to an agent depends on the type of interaction through which the agent acquires information [11].

Agents communicate by exchanging messages (information) which helps them to coordinate their actions based on the information they possess and then they share through social cognition.

We start from the definition of *agency and cognition as a property of all living organisms,* building on Maturana and Varela [13][14] and Stewart [16]. The next question will be how artifactual agents should be built in order to possess cognition and eventually even consciousness. Is it possible at all, given that cognition in living organisms is a deeply biologically rooted process? Along with reasoning, language is considered high-level cognitive activity that only humans are capable of. Increasing levels of cognition evolutionary developed in living organisms, starting from basic automatic behaviours such as found in bacteria to insects (even though they have nervous

[1] Department of Computer Science and Engineering at Chalmers University of Technology and the University of Gothenburg, Sweden. Email: **dodig@chalmers.se**

[2] School of Innovation, Design and Engineering, Mälardalen University, Sweden. Email: **gordana.dodig-crnkovic@mdh.se**

system and brain, they lack the limbic system that (in amniota = limbed vertebrates = reptiles, birds and mammals) controls emotional response to physical stimuli, suggesting they don't process physical stimuli emotionally) to increasingly complex behaviour in higher organisms such as mammals. Can AI "jump over" evolutionary steps in the development of cognition?

The framework for the discussion is the *computing nature* in the form of *info-computationalism*. It takes *the world (Umwelt)* for an agent to be *information* with its *dynamics* seen as *computation*. Information is observer relative and so is computation. [11][17][18]

Cognition has been studied as information processing in such simple organisms as bacteria [19], [20] as well as cognitive processes in other, more complex multicellular life forms. While the idea that *cognition is a biological process in all living organisms* has been extensively discussed [14][16][21], it is not clear on which basis cognitive processes in all kinds of organisms would be accompanied by (some kind of, some degree of) consciousness. If we in parallel with "minimal cognition" [22] search for "minimal consciousness", what would that be? Opinions are divided at what point in the evolution one can say that consciousness emerged. Some would suggest as Liljenström and Århem that only humans possess consciousness, while the others are ready to recognize consciousness in animals with emotions (like amniota) [23][24]. From the info-computational point of view it has been argued that cognitive agents with nervous systems are the step in evolution which first enabled consciousness in the sense of internal model with the ability of distinguishing the "self" from the "other" [4][25].

2 LEARNING IN THE COMPUTING NATURE

For naturalist, nature is the only reality [26]. Nature is described through its structures, processes and relationships, using a scientific approach [27][28]. Naturalism studies the evolution of the entire natural world, including the life and development of human and humanity as a part of nature. Social and cultural phenomena are studied through their physical manifestations. An example of contemporary naturalist approach is the research field is social cognition with its network-based studies of social behaviors.

Computational naturalism (pancomputationalism, naturalist computationalism, computing nature)[29][30][31][3][4] is the view that the entire nature is a huge network of computational processes, which, according to physical laws, computes (dynamically develops) its own next state from the current one. Among prominent. representatives of this approach are Zuse, Fredkin, Wolfram, Chaitin and Lloyd, who proposed different varieties of computational naturalism. According to the idea of computing nature, one can view the time development (dynamics) of physical states as information processing (natural computation). Such processes include self-assembly, self-organization, developmental processes, gene regulation networks, gene assembly, protein-protein interaction networks, biological transport networks, social computing, evolution and similar processes of morphogenesis (creation of form). The idea of computing nature and the relationships between two basic concepts of information and computation are explored in [11][17][18].

In the computing nature, cognition is a natural process, seen as a result of natural bio-chemical processes. All living organisms possess some degree of cognition and for the simplest ones like bacteria cognition consists in metabolism and (my addition) locomotion. [11] This "degree" is not meant as continuous function but as a qualitative characterisation that cognitive capacities increase from simplest to the most complex organisms. The process of interaction with the environment causes changes in the informational structures that correspond to the body of an agent and its control mechanisms, which define its future interactions with the world and its inner information processing. Informational structures of an agent become semantic information first in the case of highly intelligent agents.

Recently, empirical studies have revealed an unexpected richness of cognitive behaviors (perception, information processing, memory, decision making) in organisms as simple as bacteria. [18][19][32]. Single bacteria are too small, and sense only their immediate environment. They live too short to be able to memorize a significant amount of data. On the other hand bacterial colonies, swarms and films extends to a bigger space, have longer memory and exhibit an unanticipated complexity of behaviors that can undoubtedly be characterized as cognition [33][34][35]. Fascinating case are even simpler agents like viruses, on the border of the living [36][37]. Memory and learning are the key competences of living organisms [33].

Apart from bacteria and archaea [38] all other organisms without nervous system cognize (perceive their environment, process information, learn, memorize, communicate), such as e.g. slime mold, multinucleate or multicellular Amoebozoa, which has been used as natural computer to compute shortest paths. Even plants cognize, in spite of being typically thought of as living systems without cognitive capacities [39]. However, plants too have been found to possess memory (in their bodily structures that change as a result of past events), the ability to learn (plasticity, ability to adapt through morphodynamics), and the capacity to anticipate and direct their behavior accordingly. Plants are argued to possess rudimentary forms of knowledge, according to [40] p. 121, [41] p. 7 and [42] p. 61.

Consequently, in this article we take primitive cognition to be the totality of processes of self-generation/self-organization, self-regulation and self-maintenance that enables organisms to survive using information from the environment. The understanding of cognition as it appears in degrees of complexity in living nature can help us better understand the step between inanimate and animate matter from the first autocatalytic chemical reactions to the first autopoietic proto-cells.

4 LEARNING AS COMPUTATION IN NETWORKS OF AGENTS

Informational structures constituting the fabric of physical nature for an agent are networks of networks, which represent semantic relations between data. [17] Information is organized in layers, from quantum level to atomic, molecular, cellular/organismic, social, and so on. Computation/information processing, involve data structure exchanges within informational networks, represented by Carl Hewitt's actor model [43]. Different types of computation emerge at different levels of organization in nature as exchanges of informational structures between the nodes (computational agents). [11]

The research in computing nature/natural computing is characterized by bi-directional knowledge exchanges, through the interactions between computing and natural sciences. While natural sciences are adopting tools, methodologies and ideas of information processing, computing is broadening the notion of computation, taking information processing found in nature as computation. [2][44] Based on that, Denning argues that computing today is a natural science. [45] Computation found in nature is a physical process, where nature computes with physical bodies as objects. Physical laws govern processes of computation which appear on many different levels of organization.

With its layered computational architecture, natural computation provides a basis for a unified understanding of phenomena of embodied cognition, intelligence and learning (knowledge generation), including meta-learning. [30][46] Natural computation can be modelled as a *process of exchange of information in a network of informational agents* [43], that is entities capable of acting on their own behalf.

One sort of computation is found on the quantum-mechanical level where agents are elementary particles, and messages (information carriers) are exchanged by force carriers, while different types of computation can be found on other levels of organization in nature. In biology, information processing is going on in cells, tissues, organs, organisms and eco-systems, with corresponding agents and message types. In biological computing the message carriers are chunks of information such as molecules, while in social computing they are sentences while the computational nodes (agents) are be molecules, cells, organisms in biological computing or groups/societies in social computing. [18]

5 INFO-COMPUTATIONAL LEARNING BY MORPHOLOGICAL COMPUTATION

The notion of computation in this framework refers to the most general concept of *intrinsic computation,* that is a spontaneous computation processes in the nature, and which is used as a basis of specific kinds of *designed computation* found in computing machinery [47]. Intrinsic natural computation includes quantum computation [47][48], processes of self-organization, self-assembly, developmental processes, gene regulation networks, gene assembly, protein-protein interaction networks, biological transport networks, and similar. It is both analog (such as found in dynamic systems) and digital. The majority of info-computational processes are sub-symbolic and some of them are symbolic (like languages).

Within info-computational framework, computation on a given level of organization of information presents a realization/actualization of the laws that govern interactions between its constituent parts. On the basic level, computation is manifestation of causation in the physical substrate. In every next layer of organization a set of rules governing the system switch to the new emergent regime. It remains yet to be established how this process exactly goes on in nature, and how emergent properties occur. Research on natural computing is expected to uncover those mechanisms. In words of Rozenberg and Kari: "(O)ur task is nothing less than to discover a new, broader, notion of computation, and to understand the world around us *in terms of information processing*." [2] From the research in complex dynamical systems, biology, neuroscience, cognitive science, networks, concurrency and more, new insights essential for the info-computational universe may be expected.

Turing 1952 paper [49] may be considered as a predecessor of natural computing. It addressed the process of morphogenesis proposing a chemical model as the explanation of the development of biological patterns such as the spots and stripes on animal skin. Turing did not claim that physical system producing patterns actually performed computation. From the perspective of computing nature we can now argue that morphogenesis is a process of morphological computation. Informational structure (as representation of a physical structure) presents a *program* that governs computational process [50] which in its turn changes that original informational structure obeying/ implementing/ realizing physical laws.

Morphology is the central idea in our understanding of the connection between computation and information. Morphological/morphogenetic computing on that informational structure leads to new informational structures via processes of self-organization of information. Evolution itself is a process of morphological computation on a long-term scale. It is also possible to study morphogenesis of morphogenesis (Meta-morphogenesis) as done by Aaron Sloman in [51].

Leslie Valiant [52] studies evolution by ecorithms – learning algorithms that perform "probably approximately correct" PAC computation. Unlike classical paradigm of Turing computing, the results are not perfect, but *good enough* (for an agent).

6 LEARNING FROM RAW DATA AND UP – AGENCY FROM SYSTEM 1 TO SYSTEM 2

Cognition is a result of a processes of morphological computation on informational structures of a cognitive agent in the interaction with the physical world, with processes going on at both sub-symbolic and symbolic levels. This morphological computation establishes connections between an agent's body, its nervous (control) system and its environment. Through the embodied interaction with the informational structures of the environment, via sensory-motor coordination, information structures are induced (stimulated, produced) in the sensory data of a cognitive agent, thus establishing perception, categorization and learning. Those processes result in constant updates of memory and other structures that support behaviour, particularly *anticipation*. *Embodied* and corresponding *induced* (in the Sloman's sense of virtual machine) [53] informational structures are the basis of all cognitive activities, including consciousness and language as a means of maintenance of "reality" or the representation of the world.

From the simplest cognizing agents such as bacteria to the complex biological organisms with nervous systems and brains, the basic informational structures undergo transformations through morphological computation (developmental and evolutionary form generation), develop and evolve.

Living organisms as complex agents inherit bodily structures resulting from a long evolutionary development of species. Those structures are embodied memory of the evolutionary past. They present the means for agents to interact with the world, get new information that induces memories, learn new patterns of behaviour and learn/construct knowledge. By Hebbian learning, world shapes human's (or an animal's) informational structures., Neural networks that "self-organize stable pattern recognition codes in real-time in response to arbitrary sequences of input patterns" are illustrative example. [54]

If we say that for something to be information there must exist an agent from whose perspective this structure is established, and we argue that the fabric of the world is informational, the question can be asked: *who/what is the agent*? An agent (an entity capable of acting on its own behalf) can be seen as interacting with the points of inhomogeneity (data), establishing the connections between those data and the data that constitute the agent itself (a particle, a system). There are myriads of agents for which information of the world makes differences – from elementary particles to molecules, cells, organisms, societies… - all of them interact and exchange information on different levels of scale and this information dynamics is natural computation.

On the fundamental level of quantum mechanical substrate, information processes represent actions of laws of physics. Physicists are already working on reformulating physics in terms of information [53]. This development can be related to the Wheeler's idea "it from bit". [55] and von Weizsäcker's ur-alternatives [56].

11 CONCLUSIONS AND FUTURE WORK

Contemporary Deep-Learning-Centered AI is developing from the present state System 1 coverage towards the System 2, with agency, causality, attention and consciousness as mechanisms of learning and meta-learning (learning to learn). In this process like in the past, deep learning is searching inspiration in nature, assimilating ideas from neuroscience, cognitive science, biology, and more. This approach to understanding, via decomposition and construction is close to other computational models of nature in that it seeks testable and applicable models, based on data and information processing.

At the same time, Computing nature approach models nature as consisting of physical structures that form levels of organization, on which computation processes develop. It has been argued that on the lower levels of organization finite automata or Turing machines might be an adequate model, while on the level of the whole-brain non-Turing computation is necessary, Ehresmann [57] and Ghosh et al. [58]

Within info-computational framework, cognition is synonymous with the process of life, which enables learning from life characteristics to cognitive properties within evolutionary process. As mentioned before, evolution is learning process where nature tests varieties of possibilities. Following Maturana and Varela [21], we understand the entire living word as possessing cognition of various degrees of complexity. In that sense bacteria possess rudimentary cognition expressed in quorum sensing and other collective phenomena based on information communication and information processing. Brain of a complex organism consists of neurons that are networked, communicational and computational units. Signalling and information processing modes of a brain are much more complex and consist of more info-computational layers than bacterial colony. Knowledge of the world for an agent is an informational structure that is established as a result of as well the interactions of the agent with the environment (System 1) as the information processes in agents own intrinsic structures – reasoning, anticipation, etc. (System 2).

For the future, work remains to be done on the connections between the low level and the high level cognitive processes. It is also important to find relations between cognition and consciousness as a mechanism helping to reduce number of variables that are manipulated by an agent (an organism) for the purpose of reasoning, decision-making, planning and acting in the world.

The goals of AI different from the goals of the computing nature framework. AI builds solutions for practical problems and in that it focus on (typically highest possible level of) intelligence (not yet emotional nor embodied intelligence at this stage of the development), while computing nature framework seeks to provide computational models of all kinds of natural systems, including living organisms and their evolution and development, with not only intelligence but also full scale of cognition with emotion and behaviours that are not always goal-oriented in the sense of AI. The priority of info-computational naturalism is understanding and connecting knowledge about nature, while for AI the priority is practical problem solving. Nevertheless, paths of the two are meeting in many cases and mutual exchange of ideas promises benefits for both.

REFERENCES

[1] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature. 2015.

[2] G. Rozenberg and L. Kari, "The many facets of natural computing," Commun. ACM, vol. 51, pp. 72–83, 2008.

[3] G. Dodig-Crnkovic, "Nature as a network of morphological infocomputational processes for cognitive agents," Eur. Phys. J. Spec. Top., 2017.

[4] G. Dodig-Crnkovic, "Cognition as Embodied Morphological Computation," in *Philosophy and Theory of Artificial Intelligence*, vol. 44, Studies in Applied Philosophy, Epistemology and Rational Ethics, 2018, pp. 19–23.

[5] Y. Bengio, "From System 1 Deep Learning to System 2 Deep Learning (NeurIPS 2019)," 2019. [Online]. Available: https://www.youtube.com/watch?v=T3sxeTgT4qc.

[6] Y. Bengio, "Scaling up deep learning," 2014.

[7] Y. Bengio, "The Consciousness Prior," *arXiv:1709.08568v2*, 2019.

[8] D. Kahneman, *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

[9] A. Clark, *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1989.

[10] B. Scellier and Y. Bengio, "Towards a Biologically Plausible Backprop," *Arxiv*, 2016.

[11] G. Dodig-Crnkovic, "Information, Computation, Cognition. Agency-Based Hierarchies of Levels," in *Fundamental Issues of Artificial Intelligence. Synthese Library, (Studies in Epistemology, Logic, Methodology, and Philosophy of Science), vol 376.*, V. Müller, Ed. Springer International Publishing, 2016, pp. 141–159.

[12] S. Kauffman, *Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.

[13] T. Deacon, *Incomplete Nature. How Mind Emerged from Matter*. New York. London: W. W. Norton & Company, 2011.

[14] H. Maturana, "Biology of Cognition," Defense Technical Information Center, Illinois, 1970.

[15] H. Maturana and F. Varela, *The Tree of Knowledge*. Shambala, 1992.

[16] J. Stewart, "Cognition = life: Implications for higher-level cognition," *Behav. Process.*, vol. 35, pp. 311-326., 1996.

[17] G. Dodig-Crnkovic and R. Giovagnoli, *COMPUTING NATURE*, vol. 7. Springer, 2013.

[18] G. Dodig-Crnkovic, "Physical computation as dynamics of form that glues everything together," *Inf.*, 2012.

[19] E. Ben-Jacob, "Bacterial Complexity: More Is Different on All

Levels," in *Systems Biology- The Challenge of Complexity*, S. Nakanishi, R. Kageyama, and D. Watanabe, Eds. Tokyo Berlin Heidelberg New York: Springer, 2009, pp. 25–35.
[20] E. Ben-Jacob, "Learning from Bacteria about Natural Information Processing," *Ann. N. Y. Acad. Sci.*, vol. 1178, pp. 78–90, 2009.
[21] H. Maturana and F. Varela, *Autopoiesis and cognition: the realization of the living*. Dordrecht Holland: D. Reidel Pub. Co., 1980.
[22] M. van Duijn, F. Keijzer, and D. Franken, "Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination," *Adapt. Behav.*, vol. 14, no. 2, pp. 157–170, 2006.
[23] P. Århem and H. Liljenström, "On the coevolution of cognition and consciousness," *J. Theor. Biol.*, 1997.
[24] H. Liljenström and P. Århem, *Consciousness Transitions: Phylogenetic, Ontogenetic and Physiological Aspects*. Amsterdam: Elsevier, 2011.
[25] G. Dodig-Crnkovic and von H. Rickard, "Reality Construction in Cognitive Agents through Processes of Info-Computation," in *Representation and Reality in Humans, Other Living Organisms and Intelligent Machines*, G. Dodig-Crnkovic and R. Giovagnoli, Eds. Basel: Springer International Publishing, 2017, pp. 211–235.
[26] H. Putnam, *Mathematics, Matter and Method*. Cambridge: Cambridge University Press, 1975.
[27] G. Dodig-Crnkovic and M. Schroeder, "Contemporary Natural Philosophy and Philosophies," *Philosophies*, vol. 3, no. 4, p. 42, Nov. 2018.
[28] G. Dodig-Crnkovic and M. Schroeder, *Contemporary Natural Philosophy and Philosophies - Part 1*. Basel: MDPI AG, 2019.
[29] G. Dodig-Crnkovic, *Investigations into Information Semantics and Ethics of Computing*. Västerås, Sweden: Mälardalen University Press, 2006.
[30] G. Dodig-Crnkovic and V. Müller, "A Dialogue Concerning Two World Systems: Info-Computational vs. Mechanistic," World Scientific Pub Co Inc, Singapore, 2009.
[31] G. Dodig-Crnkovic, "The info-computational nature of morphological computing," in *Studies in Applied Philosophy, Epistemology and Rational Ethics*, 2013.
[32] W.-L. Ng and B. L. Bassler, "Bacterial quorum-sensing network architectures," *Annu. Rev. Genet.*, vol. 43, pp. 197–222, 2009.
[33] G. Witzany, "Memory and Learning as Key Competences of Living Organisms," 2018.
[34] G. Witzany, "Introduction: Key Levels of Biocommunication of Bacteria," 2011.
[35] D. Dennett, *From Bacteria to Bach and Back: The Evolution of Minds*. W. W. Norton & Company, 2017.
[36] G. Witzany, *Viruses: Essential agents of life*. 2012.
[37] L. P. Villarreal and G. Witzany, "Viruses are essential agents within the roots and stem of the tree of life," *J. Theor. Biol.*, 2010.
[38] G. Witzany, *Biocommunication of archaea*. 2017.
[39] G. Witzany, "Bio-communication of Plants," *Nat. Preced.*, 2007.
[40] R. S. Pombo, O., Torres J.M., Symons J., Ed., *Special Sciences and the Unity of Science*, Logic, Epi. Berlin Heidelberg: Springer, 2012.
[41] R. Rosen, *Anticipatory Systems*. New York: Pergamon Press, 1985.
[42] K. Popper, *All Life Is Problem Solving*. London: Routledge, 1999.
[43] C. Hewitt, "What is computation? Actor Model versus Turing's Model," in *A Computable Universe, Understanding Computation & Exploring Nature As Computation*, H. Zenil, Ed. World Scientific Publishing Company/Imperial College Press, 2012.
[44] G. Rozenberg, T. Bäck, and J. N. Kok, Eds., *Handbook of Natural Computing*. Berlin Heidelberg: Springer, 2012.
[45] P. Denning, "Computing is a natural science," *Commun. ACM*, vol. 50, no. 7, pp. 13–18, 2007.
[46] Y. Wang, "On Abstract Intelligence: Toward a Unifying Theory of Natural, Artificial, Machinable, and Computational Intelligence," *Int. J. Softw. Sci. Comput. Intell.*, vol. 1, no. 1, pp. 1–17, 2009.
[47] S. Crutchfield, James P.; Ditto, William L.; Sinha, "Introduction to Focus Issue: Intrinsic and Designed Computation: Information Processing in Dynamical Systems-Beyond the Digital Hegemony," *Chaos*, vol. 20, no. 3, pp. 037101-037101–6, 2010.
[48] J. P. Crutchfield and K. Wiesner, "Intrinsic Quantum Computation," *Phys. Lett. A*, vol. 374, no. 4, pp. 375–380, 2008.
[49] A. M. Turing, "The Chemical Basis of Morphogenesis," *Philos. Trans. R. Soc. London*, vol. 237, no. 641, pp. 37–72, 1952.
[50] G. Kampis, *Self-modifying systems in biology and cognitive science: a new framework for dynamics, information, and complexity*. Amsterdam: Pergamon Press, 1991.
[51] A. Sloman, "Meta-Morphogenesis: Evolution and Development of Information-Processing Machinery p. 849.," in *Alan Turing: His Work and Impact*, S. B. Cooper and J. van Leeuwen, Eds. Amsterdam: Elsevier, 2013.
[52] L. Valiant, *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. New York: Basic Books, 2013.
[53] A. Sloman and R. Chrisley, "Virtual machines and consciousness," *J. Conscious. Stud.*, vol. 10, no. 4–5, pp. 113–172, 2003.
[54] G. A. Grossberg and S. Carpenter, "ART 2: self-organization of stable category recognition codes for analog input patterns," *Appl. Opt.*, vol. 26, no. 23, pp. 4919–4930, 1987.
[55] J. A. Wheeler, "Information, physics, quantum: The search for links," in *Complexity, Entropy, and the Physics of Information*, W. Zurek, Ed. Redwood City: Addison-Wesley, 1990.
[56] C. F. Weizcsäcker, "The Unity of Nature," in *Physical Sciences and History of Physics. Boston Studies in the Philosophy of Science. vol. 82.*, C. R.S. and M. W. Wartofsky, Eds. Springer, Dordrecht, 1984.
[57] A. C. Ehresmann, "MENS, an Info-Computational Model for (Neuro-)cognitive Systems Capable of Creativity," *Entropy*, vol. 14, pp. 1703-1716., 2012.
[58] S. Ghosh, K. Aswani, S. Singh, S. Sahu, D. Fujita, and A. Bandyopadhyay, "Design and Construction of a Brain-Like Computer: A New Class of Frequency-Fractal Computing Using Wireless Communication in a Supramolecular Organic, Inorganic System," *Information*, vol. 5, no. 1, pp. 28–100, Jan. 2014.

Is the Church Turing Thesis a Red Herring For Cognitive Science?

Dean Petters [1] and **Achim Jung** [2]

Abstract. This paper considers whether computational formalisms beyond the Church Turing Thesis (CTT) could be helpful in understanding the mind. We argue that they may be, and that the way that the CTT has been invoked in Cognitive Science may therefore act as a Red Herring. That is, the way the CTT is invoked in Cognitive Science may mislead and perhaps contribute to premature abandonment of possibly fruitful research directions in Cognitive Science. We do not suggest some sort of "hypercomputation'. Whilst it is possible to use a rich interactive machine to implement a simple function this does not lead to new computable functions. In other words, the CTT is valid even if more sophisticated machinery is employed. It is the other direction that is the core of this paper: When considering more sophisticated computational tasks, then standard Turing machines (and their mode of operation) are not sufficient to explore the range of possibilities. The CTT is commonly interpreted as stating that the intuitive concept of computability is fully captured by Turing machines or any equivalent formalism (such as recursive functions, the lamba calculus, Post production rules, and many others). The CTT implies that if a function is (intuitively) computable, then it can be computed by a Turing machine. Conversely, if a Turing machine cannot compute a function, it is not computable by any mechanism whatsoever. We suggest an inadvertent error that has been made which is the claim that relatively simple computational formalisms like Turing Machines can do anything that more complex computional formalisms can do. To show this we present the landscape of computability within and beyond the bounds covered by the mathematical CTT. This shows that in regions of the computational landscape beyond the CTT there may be hierarchies of increasingly powerful computational formalisms. Erroneously interpreting CTT as enforcing a 'one size fits all' interpretation to computational formalisms leads to extreme reductionism that means contemporary computationalism is viewed as inadequate to explaining many phenomena related to thought and mind in living systems. Once this Red Herring interpretation for CTT is avoided this leaves the way open to exploring how richer kinds of computation that may possess many shades of expressivity can form part of Cognitive Science explanations.

1 Introduction

This paper takes the position that there are physically implementable programs which are outside the scope of the Church-Turing thesis (CTT). That is, we refute the existing idea that all computation has a boundary between what are computable functions and non-computable functions that is clear and distinct boundary for all formalisms. We show why this finding is important for Cognitive Science. The central argument of this paper is that invoking a mathematical theorem to make inferences about real-time physically instantiated systems should be done with careful consideration of both the scope of the theorem and the properties and complexity of the physical system. Turing set out to solve the "Entscheidungsproblem" (decision problem) and for this purpose proposed a mathematical formalism that faithfully emulates the process of a human being following finitely specified instructions. It was soon found that other formalisms have the same expressive power in this specific setting, i.e., mathematical problem solving, and this then led to the CTT. Situations in contemporary computing are now so rich, they can no longer be said to be covered by a paradigm where the inputs are known in advance, the system is left alone to do its computation and then provides the answer. Critically, for richer kinds of computation, the empirical evidence suggests that there are many shades of expressivity, which is why no-one has ever postulated an analogue of the CTT for them. This has implications for Cognitive Science and Artificial Intelligence. This is because it means that there may be computational formalisms which are strictly beyond the existing CTT but nevertheless recognisably symbolic/representational (GOFAI) in approach. Thus showing that cognitive scientists do not need to 'go all the way' to invoke non-representational or non-computational approaches (so called nouvelle AI such as enactivist [5], embodied [10], or dynamical systems approaches [11]) when going beyond classic computational formalisms. Instead, to explore how recognisably computational (representational/symbolic) systems can model phenomena of interest differently to formalisms that are within the scope of the CTT they only need to go 'slightly' beyond CTT and keep within the realm of computationalism. In discussing formalisms beyond CTT we not proposing a form of hypercomputation. The CTT is still valid when more sophisticated machinery is employed when that machinery is used to do carry out computational tasks that can be carried out by a Turing Machine (or computationally equivalent formalism). We are instead considering more sophisticated computational tasks than standard Turing machines (and their mode of operation) are sufficient to explore. This is of critical relevance to Cognitive Science - which studies humans performance on such sophisticated tasks to discover what computational formalisms humans possess. These formalisms may be beyond the CTT but still be symbolic computation.

2 The role of the CTT in Computationalism in Cognitive Science

The CTT is closely linked to the historical origins of computationalist (cognitivist) account of cognition. For example, in his historical review Clarke [2] notes:

[1] Sheffield Hallam University, UK, email: d.petters@shu.ac.uk
[2] University of Birmingham, UK, email: A.Jung@cs.bham.ac.uk

"The next big development was the formalization (Turing, 1936) of the notion of computation itself. Turing's work, which predates the development of the digital computer, introduced the foundational notion of (what has since come to be known as) the Turing machine. This is an imaginary device consisting of an infinite tape, a simple processor (a "finite state machine"), and a read/write head. The tape acts as a data store, using some fixed set of symbols. The read/write head can read a symbol off the tape, move itself one square backward or forward on the tape, and write onto the tape. The finite state machine (a kind of central processor) has enough memory to recall what symbol was just read and what state it (the finite state machine) was in. These two facts together determine the next action, which is carried out by the read/write head, and determine also the next state of the finite state machine. What Turing showed was that some such device, performing a sequence of simple computations governed by the symbols on the tape, could compute the answer to any sufficiently well-specified problem. "

"We thus confront a quite marvelous confluence of ideas. Turing's work clearly suggested the notion of a physical machine whose syntax following properties would enable it to solve any well-specified problem." ([2], p. 11-13)

In this historical analysis Clarke suggests that the concept of Turing Machines and the CTT, along with ideas that had previously been formulated on logics and formal systems, led to a radical new computationalist approach in Cognitive Science. Clarke cites Pylyshyn, who made these same points in the 1970s:

"The work of Turing, in a sense, marked the beginnings of cognitive activity from an abstract point of view, divorced in principle from both biological and phenomenological foundations. It provides a reference point for the scientific ideal of a mechanistic process which could be understood without raising the spectre of vital forces or elusive homunculi but which at the same time was sufficiently rich to cover every conceivable formal notion of mechanism (that the Turing formulation does cover all such notions is, of course, not provable but is has stood all attempts to find exceptions. The belief that it does cover all possible cases of mechanism has become known as the Church-Turing thesis). It would be difficult to overestimate the importance of this development for psychology. It represents the emergence of a new level of analysis, which is independent of physics, yet it mechanistic in spirit. It makes possible a science of structure and function divorced from material substance, while at the same time it avoids the retreat to behavioralistic periperheralismm. It speaks the language of mental structures and of internal processes, thus lending itself to answering questinos traditionally posed by psychologists"

"While Turing and other mathematicians, logicians, and philosophers laid the foundations for the abstract study of cognition in the 30s and 40s it was only in the last twenty or so years thatthis idea begain to be articuated in a much more specific and detailed form: A form which lends itself more directly to attacking certain basic questions of cognitive psychology. The newer direction has grown with the continuing development of our understanding of the nature of computational process and of the digital computer as a general, symbol-processing system. It has led to the formation of a new intellectual discipline known as artificial intelligence, which attemnpts to understand the nature of intelligence by designing computational systems which exhibit it" ([8], 24-25)

What these quotes show is how the CTT led to promotion of multiple realisability and the stronger notion of medium independence as supporting foundations for cognitive science. However, gaining the notion of multiple realisability through invocation of the CTT brought with it the possibility of a limiting misconception - a Red Herring - as this misconception that all computational formalisms are equivalent has led to a mistaken view of computational approaches to the mind leading to extreme reductionism. This extreme reductionism follows from the misconception that very simple computational formalisms are computationally equivalent to more complex formalisms because they can produce the same set of functions. Turing's original machine is a very simple abstract concept. There is a control unit in a particular state, and finitely many alternative states. There is also an infinite tape, which acts as a memory and on which can be marked '0','1' or 'nothing'. There is also a read-write head which takes decisions and can change a '0' to a '1', change a '1' to a '0', or erase a '0' or '1'. There are even simpler computational formalisms like the *two-counter machine*. This can increment and decrement with branching. The extreme reductionism becomes apparent when we ask: Can this or a Turing Machine be programmed to be conscious? The line of reasoning that acts as a Red Herring is: If consciousness arises from computation, and the CTT is correct in stating that all computational frameworks are equivalent, then if any computational system can exhibit self-awareness these kind of simple machines will exhibit self-awareness. The widespread view (that we agree with) that Turing Machines or two-counter machines cannot be conscious simply by running the right program has led to the conclusion that psychological phenomena such as consciousness, agency or self-awareness are not computational in origin. We suggest that a different route out of this impasse is to accept that applying CTT to all forms of computation is a Red Herring. That it, it is an unhelpful misconception that misdirects research. Researchers looking for computational explanations for complex psychological phenomena that are not simply function computations should look beyond the CTT to more sophisticated computational formalisms.

Not all researchers have been misdirected by a misconception that the CTT applies to all forms of computation. Goldin and Wegner [3], examine this misconception and suggest that the operation of "batch processing" in the first generation of computing machines was so similar to Turing's mathematical concept of a (human) "computer" (i.e., his "Turing machines") that Turing machines were incorrectly adopted as a sole formal abstraction of computing practice. Goldin and Wegner point out the role of interactivity in processes that is so central to modern computing system is simply not covered by the CTT. Some researchers have been very aware that Turing machines are not appropriate for modelling interactive processes and have proposed alternative mathematical abstractions [4, 6]. A key issue is that when we consider computation from fixed input to single output (the "function view" of computation), then the equivalence of computational mechanisms is almost unavoidable. In contrast, mathematical models for interactive behaviour (the "process view" of computation) can be *quite different* in expressivity. A canonical, maximally expressive formalism for processes simply does not exist. We point the interested reader to Abramsky's [1] where this fact is highlighted and explored.

3 The landscape of computability in diagrams

The landscape of computability includes regions within and beyond the bounds of the classic Church-Turing thesis. Figure 1 shows that for batch style computation (all of the formalisms on both the top and bottom of the left half of the diagram) there is a 'one size fits' organisation of the landscape of computability due to the CTT. That is, functions are either computable or non-computable whatever formalim is used. Regarding the right-hand side of the diagram: the question mark signifies that we don't know what the situation is. Before we can distinguish computable from non-computable entities on the right hand side of the diagram, we first need to decide what "entity" is being computed by a distributed or probabilistic or other kind of program. Once we have a clear idea for that (unlikely in the case of distributed computing), we can then try to see whether we get an analogue of the CTT (with all reasonable formalisms being computationally equivalent), or whether the situation is more like that of the total functions in the bottom left square of the left hand side of the diagram. That is, for total functions that are computable, different formalisms cover different parts of the computable realm and none covers all of it. Therefore, the right hand side of the landscape of computability (for contemporary and future computation) might have an infinite tower of increasingly powerful computational formalisms. Or some other kind of hierarchy. Such as a finite tower of increasingly powerful formalisms. This is important for Cognitive Science because it means there may be computational formalisms for contemporary and future computational approaches that do not have a 'one size fits all' organisation. Therefore perhaps changing the current extreme reductionism which is currently justified by invocation of the CTT. The rationale for this extreme reductionism is that all formalisms within the scope of the CTT, however complex, can produce the same set of computations as very simple formalisms like Turing Machines).

4 From the Church Turing Thesis to the Chinese Room Argument

Figure 2 situates particular kinds of programs in the landscape of computability. In particular, situating the kind of batch program that Searle describes in his Chinese Room Argument (CRA) [9] and a class of adapted CRA program sketched by Petters and Jung [7] - with interruptions and interactivity, real-time processing, never-ending computation and parallel distributed control [1, 3, 4, 6]. This adapted CRA program will not lead to new computable functions, i.e., some sort of "hypercomputation". The CTT is still valid when more sophisticated machinery is employed to compute functions that could be computed by programs within simpler formalisms. Our claim is that when considering more sophisticated computational tasks, standard Turing machines (and their mode of operation) are not sufficient to explore the range of possibilities that can be produced with this kind of formalism.

5 Conclusion

This paper argues that due to the extreme reductionism enforced by the Church Turing Thesis contemporary computationalism is inadequate to explaining many phenomena related to thought and mind in living systems. This paper is not proposing a kind of hypercomputation. Whilst it is possible to use a rich interactive machine to implement a simple function this does not lead to new computable functions'. In other words, the CTT is valid even if more sophisticated machinery is employed. It is the other direction that is the core of this paper: When considering more sophisticated computational tasks, then standard Turing machines (and their mode of operation) are not sufficient to explore the range of possibilities. This paper suggests computational formalisms beyond the computational formalisms covered by the mathematical Church Turing are likely to be particularly valuable for explaining cognitive processes in living organisms that are not simply function computations.

REFERENCES

[1] S. Abramsky, 'Intensionality, definability and computation', in *Johan van Benthem on Logic and Information Dynamics. Outstanding Contributions to Logic, vol 5.*, eds., A. Baltag and S. Smets, 121–142, Springer, (2014).

[2] A. Clark, *Mindware*, Oxford University Press,, Oxford, 2014. 2nd edition.

[3] D.Q. Goldin and P. Wegner, 'The Church-Turing thesis: Breaking the myth', in *New Computational Paradigms*, eds., S. Barry Cooper, Benedikt Löwe, and Leen Torenvliet, pp. 152–168. Springer Berlin Heidelberg, (2005).

[4] C.A.R. Hoare, *Communicating Sequential Processes*, Prentice Hall International, 1985.

[5] D. Hutto and M. Myin, *Radicalizing Enactivism: Basic Minds without Content*, MIT Press, Cambridge, MA, 2013.

[6] R. Milner, *A Calculus for Communicating Systems*, volume 92 of *Lecture Notes in Computer Science*, Springer Verlag, 1980.

[7] D. Petters and A. Jung, 'From the Chinese Room Argument to the Church-Turing Thesis', in *Proceedings of the International Symposium on 'Philosophy after AI: Mind, Language and Action*, eds., G. Gallo and C. Stancati, 25–29, AISB Press, University of Liverpool, (2018).

[8] Z. Pylyshyn, 'Complexity and the Study of Human and Machine Intelligence', in *Mind Design*, ed., J. Haugeland, 23–56, Bradford Books/MIT Press, Cambridge, MA, (1979).

[9] J.R. Searle, 'Minds, brains, and programs', *The Behavioral and Brain Sciences*, **3**(3), (1980). (With commentaries and reply by Searle).

[10] L. Shapiro, *Embodied Cognition*, Routledge, London, UK, 2010.

[11] T. van Gelder, 'Dynamics and cognition', in *Mind Design II*, ed., J. Haugeland, MIT Press, Cambridge, MA, (1997).

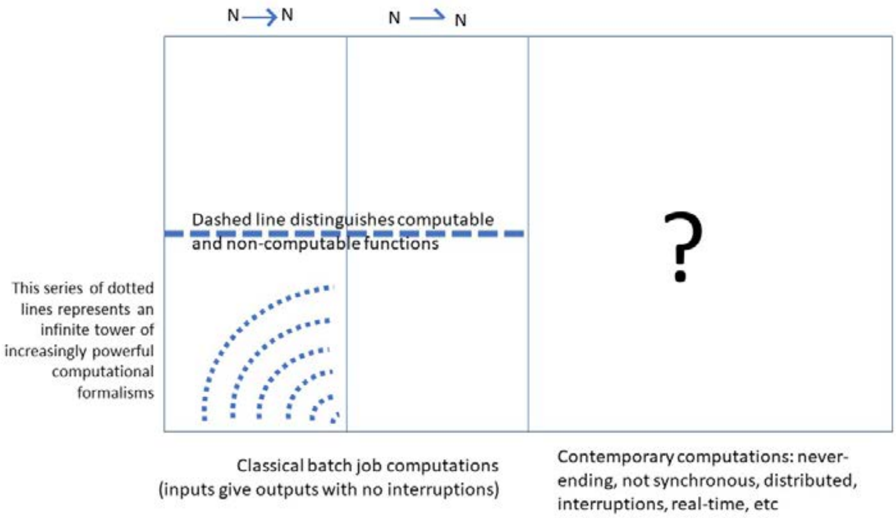

Figure 1. The landscape of computability. There are two vertical lines in the diagram. The left line distinguishes between total and partial functions. In the region of the landscape where total functions are computable it can be shown that for whatever formalism is considered, diagonalization can always be used to create a new function beyond the set computable by that formalism. So there is an infinite number of possible formalisms forming a tower of increasing computational power - represented by a series of curved dotted lines. The right hand split distinguishes classical batch computation or not. The right hand side of the right hand vertical line is therefore contemporary approaches like never-ending, not-synchronised, distributed, real-time, and probabilistic computation. As well as computations with other attributes we take for granted in 2020. Plus as yet undiscovered models possessing attributes that are beyond what can be possessed by computations in Turings and other formalisms for computation (like Churchs, Posts, Gödel's and many others). The dashed horizontal line through the left half of the figure distinguishes computable and non-computable functions. According to CTT, every reasonable formalism gives all lower left outputs below the dashed line. There is (as yet) no comparable thesis to the Church-Turing thesis for the right of the figure (contemporary computation). (Note: the areas are not to scale, below the horizontal split (computable) would actually be a tiny sliver compared to non-computable.)

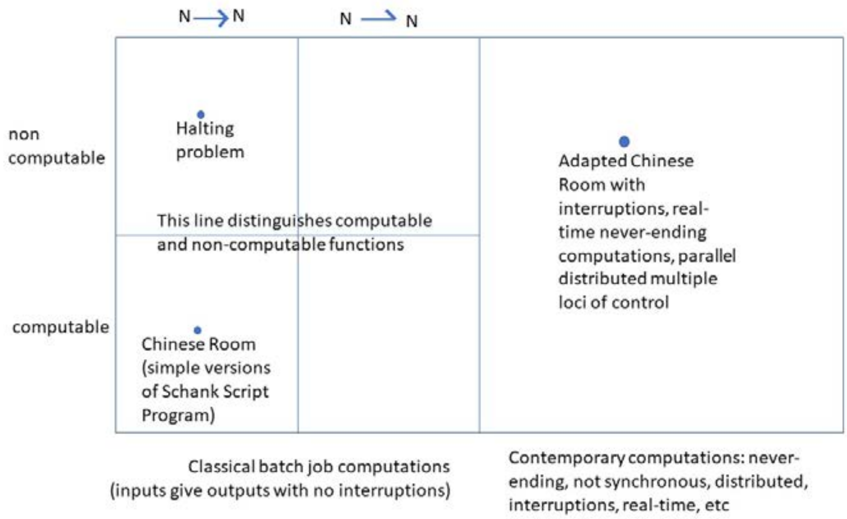

Figure 2. Situating particular kinds of programs in the landscape of computability. Turing showed that the halting problem was non-computable. Petters and Jung [7] show that Searle's original CRA argument through experiment described a batch job program that was computable. Petters and Jung [7] also briefly sketch an adapted CRA program with interruptions and interactivity, real-time processing, never-ending computation and parallel distributed control which is outside the scope of the traditional CTT and so on the right hand side of the landscape of computability. So this adapted CRA program may or may not be computable, and the formalism used to implement it may be more computationally powerful than the formalism used to implement Searle's original chinese room program. Therefore, conclusions from the original CRA may not apply to all implementable programs.

Creativity, Eco-Cognitive Openness, Human and Machine Inferential Capacities
Deep Learning and Locked Strategies

Lorenzo Magnani [1]

Abstract. Locked and unlocked strategies are at the center of this article, as ways of shedding new light on the cognitive aspects of deep learning machines. The character and the role of these cognitive strategies, which are occurring both in humans and in computational machines, is indeed strictly related to the generation of cognitive outputs, which range from weak to strong level of knowledge creativity. I maintain that these differences lead to important consequences when we analyze computational AI programs, such as AlphaGo, which aim at performing various kinds of abductive hypothetical reasoning. In these cases, the programs are characterized by *locked* abductive strategies: they deal with weak (even if sometimes amazing) kinds of hypothetical creative reasoning, because they are limited in what I call eco-cognitive openness, which instead qualifies human cognizers who are performing higher kinds of abductive creative reasoning, where cognitive strategies are instead *unlocked*.

1 ARE MACHINE COGNITIVE CAPACITIES LOCKED? AN ABDUCTIVE PERSPECTIVE

In 2015, Google DeepMind's program AlphaGo (able to perform the famous Go game, aslo called Baduk in Souther Korea) beat Fan Hui, the European Go champion and a 2 dan (out of 9 dan) professional, five times out of five with no handicap on a full size 19×19 board. Later on, in March 2016, Google also struggled againts Lee Sedol, a 9 dan player who was said to be the top world champion, to a five-game match. The famous DeepMind program won in four of the five games. It was seen that the program "generated" a new and eccentric move—never used by human beings—which was able to originate a new strategic path, considered as simulating actual "human" skillful capacities, better than the ones of the more expert humans. AlphaGo learnt its machine capacities to play the game by taking advantage of "seeing" data of thousands of games, perhaps also including those played by Lee Sedol, exploiting the so-called "reinforcement learning": the program in turn plays successively against itself to improve, enrich, and adjust once again further its own deep neural networks grounded on trial and error procedures.

Let us adopt the meaning of the term 'cognitive 'capacity" to refer—this seems particularly appropriate in the case of AI studies—to an expert mixture in reasoning of various strategic and heuristic devices. When we are referring to the non-computational and human case of game theory, the meaning of the word capacity more amply refers to the role of the agents in their relationships with other agents and to the various related contentious or collective cognitive acts. I contend that it is in the framework of abductive cognition [16] we can appropriately and usefully analyze the concept of cognitive capacities to the aim of seeing the distinction between *locked* and *unlocked* eco-cognitive settings. Indeed, I will contend that in AlphaGo only locked cognitive capacities are at play, and this fact seriously limits the type of creativity which is in general performed by deep learning machines.

In my studies on abduction, I have extendedly illustrated various kinds of human (but also computational) hypothetical cognition. I proposed the adoption of the distinction between *selective* abduction [14]—for example, in diagnosis (in which abduction is basically described as an inferential process of "selecting" from a "repository" of pre-stored hypotheses)—and *creative* abduction (abduction that produces *new* hypotheses). Furthermore, I have shown that abduction is not only related to the propositional aspect, i.e. when processed using human language (oral and written), but can also be "model-based" and "manipulative". In the first case, we deal with an abduction that is basically performed thanks to internal cognitive acts that take advantage of models such as simulations, visualizations, diagrams, etc.; in the second case the external dimension is at play: in this case, an eco-cognitive perspective is fundamental because we have to refer to those cognitive actions (embodied, situated, embedded, distributed and extended, and/or enacted, as recent cognitive research says) in which the role of external models (for example artifacts), props, and technological devices, is relevant, and in which the characters of the actions themselves are hidden and hard to be extracted. Action can give origin to otherwise unavailable data that grant to the agent the chance of solving problems by initiating and processing an appropriate abductive procedure of production and/or selection of hypotheses. As I say, manipulative abduction is occurring when we are thinking "through" doing and not only, in a pragmatic sense, about doing (cf. [16] chapter 1). It is clear that, when we are dealing with games such as Go, manipulative abduction is also at play, given the fact the reasoning is considerably intertwined with the manipulation of the stones and various embodied aspects are involved, together with the visualization of the whole scenario, the adversary, etc.

I have also to add that, the concept of abduction has been involved in AI at least since the beginnings of this young discipline. Already in 1988 Paul Thagard [26] described four types of abduction implemented in PI, a computational program devoted to perform some of the main cognitive capacities illustrated by philosophy of science: scientific discovery, explanation, evaluation, etc. The program explicitly executes the so-called simple, existential, rule-forming, and analogical abduction. In this case the use of computer simulation exhibited a first sophisticated new tool to increase knowledge about ab-

[1] Department of Humanities, Philosophy Section and Computational Philosophy Laboratory, University of Pavia, Italy, email: lmagnani@unipv.it

duction, illustrating how this kind of non deductive reasoning can be automatically rendered thanks to computational concrete artefacts. Early work on the so called *machine scientific discovery*, such as the well-known Logic Theorist (Newell et al. [24]), DENDRAL in chemistry (Lindsay et al. [13]), and AM in mathematics (Lenat [12]), demonstrated that *heuristic search* in combinatorial spaces represents an appropriate and general instrument for automating scientific discovery, and abduction was explicitly or implicitly categorized.[2] What about the new perspectives on hypothetical abductive reasoning offered by deep learning programs? To clarify the cognitive character of this program, the examination of the kinds of strategies that are at play is in my opinion central.

1.1 HUMAN AND MACHINE CAPACITIES AND THEIR ABDUCTIVE CHARACTER

I hope it is now patent that studies on abduction are very useful when we have to describe creative reasoning, and a simple, completely *new* and unexpected move of a human being who is playing a Go game surely represents a kind of creative reasoning. In this article, it will be the key concept of *knowledge-enhancing abduction* and the related one of *eco-cognitive openness* that will favor a deeper understanding of the logical and cognitive condition of those kinds of cognitive capacities I will describe and that I called *locked* and *unlocked abductive strategies* [21]. Locked and unlocked strategies are at the center of this article, as ways of shedding new light on the cognitive aspects of deep learning machine capacities. The character and the role of these cognitive strategies, which are occurring both in humans and in computational machines, is indeed strictly related to the generation of cognitive outputs, which range from weak to strong level of knowledge creativity. I maintain that these differences lead to important consequences when we analyze computational AI programs, such as AlphaGo, which aim at performing various kinds of abductive hypothetical reasoning.

We have to first of all to say that these programs are characterized by *locked* abductive strategies: they deal with weak (even if sometimes amazing) kinds of hypothetical creative reasoning, because they are limited in what I call eco-cognitive openness, which instead qualifies human cognizers who are performing higher kinds of abductive creative reasoning, where cognitive strategies are instead *unlocked*. An objection to the adoption of the concept of abduction to shed more cognitive light on the strong impact of deep learning programs such as AlphaGo in contemporary AI regards the fact that they are based on hierarchical neural networks that operate on a subconceptual level: abduction has been instead fundamentally investigated thanks to symbolic formal models related to the tradition of logic. I have indicated in the previous subsection that certainly the concept of abduction enters in the last part of the previous century traditional AI research thanks to the studies concerning automated scientific discovery (creative abduction) and medical diagnostic reasoning (selective abduction). The dominant representational tools were the symbolic ones such as classical logic programming, rule-based systems, probabilistic networks, etc. Can good abductive processes be modeled using representational tools and algorithms that operate on a subconceptual level? The answer is yes. For example, Bruza et al. [2] insisted that it would be misguided to adopt a simple traditional, symbolic perspective of an abductive logical system by assuming a propositional knowledge representation and proof-theoretic approaches for driving it, because this perspective seems conceptually incomplete insofar as it ignores what is going "down below" [6], which can be interpreted as the subconceptual level of cognition.

In sum, certainly the mainstream non-standard logical tradition which created models of abductive inferences was characterized by *symbols* but, from a more extended cognitive and philosophical perspective, also *multimodality* (that is cognition in terms of non propositional models, icons, thought experiments, simulations, etc.), and *implicit reasoning* appear to be important. Moreover, I have to remember that, from a wide cognitive and philosophical perspective, as I have illustrated in my own research [16, 19], the term abduction refers to all kinds of cognitive activities that lead to hypotheses, in human and non human animals. For example, humans often generate abductive hypotheses thanks to manipulative, embodied and unconscious endowments, and higher mammals surely do not take advantage of symbolic syntactic language but instead other multiple cognitive capacities. Analogously to what is happening in the case of humans, that can perform abductions in various ways, there is not in AI a unique method able to favor the development of programs able to reproduce abductive cognition to hypotheses. Various knowledge representation formalisms and algorithms can be adopted to implement an appropriate computational program.[3]

At this point, we can go ahead and try to analyze the specific kind of abductive performance (the generations of "moves") that characterizes the deep learning AI program AlphaGo.

2 NATURAL, ARTIFICIAL, AND COMPUTATIONAL GAMES

2.1 LOCKED AND UNLOCKED STRATEGIES IN NATURAL AND ARTIFICIAL COGNITIVE CAPACITIES

Go is a game played by human agents and AlphaGo is a computational deep learning program that can play the role of an automatic agent/player, so that a competition with humans can become partially computationally determined. Go is already an "artificial" game, as it is invented by human beings and, consequently, takes advantage of abstract rules and artifacts, such as the board and other material objects. AlphaGo is artificial too, but a more complicated fruit of the technological creativity of a more restricted and specialized group of human beings. However, we have to remember that also "natural cognitive games", so to speak, can be contemplated. For example, as I have already illustrated above when describing manipulative abduction, a strategic human cognition not only refers to propositional aspects concerning acts performed through written and spoken language, but it is also active in a distributed cognition environment, in a kind of "game" in which models, artifacts, internal and external representations, sensations, and manipulations play a central function: imagine the pre-linguistic cognitive "natural game" between humans and their surroundings, in which "unlocked" strategies (see below) are at play, such as the phenomenological tradition has illustrated, exactly involving embodied and distributed systems, and visual, kinesthetic, and motor sensations [20].

What counts here is that in the natural games the cognitive capacities are *unlocked* because, even if local constraints are always at play in the interaction humans/environments, no preset background is established. On the contrary, what happens in the case of the cognitive

[2] A more detailed illustration of the development, until the present times, of programs related to abduction and scientific discovery are illustrated in section 1.1 of [21].

[3] The need of a plurality of representations was already clear at the time of classical AI formalisms, when I was collaborating with AI researchers to implement a Knowledge-Based System (KBS) able to develop medical abductive reasoning [25].

capacities of human made "artificial games" such as Go, or in the case of their computational counterpart, such as AlphaGo? In these two last cases, the involved cognitive strategies are *locked*, as I will describe in the following paragraphs.

Let us abandon the problem (I have just sketched) of the prelinguistic cognitive abductive strategies which are at play in a natural interaction—natural game—between humans and their prepredicative surroundings (for example, philosophically studied by Husserl [9]) and let us concentrate on the cognitive abductive strategies that are at play in the artifactual case of the moves that are occurring in the adversarial game Go with two players and their respective changing surroundings, which in this case are basically formed by board, stones, and possible artifactual assisting accessories. In this game, analogously to the case of the natural processes, we still obviously find the role of visual, kinesthetic, and motor sensations and actions, but also the strong function of visual, iconic, and propositional artificial representations, anchored to the human made "meanings" (both internal and external) which gave birth to the game and which characterize its features and rules.

2.2 ANTICIPATION AS "READING AHEAD"

A fundamental strategy we immediately detect in artificial games such as Go, which is necessary for proficient and smart tactical play, is the capacity to *read ahead*,[4] as the Go players usually say. Reading ahead is a practice of generating groups of anticipations that aim at being very robust and complex (either serious-minded or intuitive) and that demand the consideration of

1. Clusters of moves to be adopted and their potential outcomes. The available scenario at time t_1, exhibited by the board, represents an adumbration[5] of a subsequent potential more profitable scenario at time t_2, which indeed is abductively credibly hypothesized: in turn, one more abduction is selected and actuated, which—consistently and believably—activates a particular move that can lead to an envisaged more fruitful scenario.
2. Possible countermoves to each move.
3. Further chances after each of those countermoves. It seems that some of the smarter players of the game can read up to 40 moves ahead even in hugely complex positions.

Further strategies that are usefully adopted by human players in the game Go are for instance related to "global influence, interaction between distant stones, keeping the whole board in mind during local fights, and other issues that involve the overall game. It is therefore possible to allow a tactical loss when it confers a strategic advantage".[6]

The material and external scenarios (which are composed by the sensible objects—stones and board) that characterize artificial games are the fruit of a cognition "sedimented"[7] in their embodiment, after the starting point of their creation and subsequent uses and modifications. The cognitive tools that are related to the application of both the game allowed rules and the individual inferential talents owned by the two players, strategies, tactics, heuristics, etc. are sedimented in those material objects (artifacts, in this case) that become *cognitive mediators*:[8] for example they orient players' inferences, transfer information, and provoke reasoning chances. Once represented internally, the external subsequent scenarios become object of mental manipulation and new ones are further made, to the aim of producing the next most successful move.

It is relevant to note again that these strategies, when actuated, are certainty characterized by an extended variety, but all are "locked", because the elements of each scenario are always the same (what changes is merely the number of seeable stones and their dispositions in the board), in a finite and stable framework (no new rules, no new objects, no new boards, etc.) These strategies are devoid of the following feature: they are not able to recur to reservoirs of information *different* from the ones available in the fixed given scenario. It is important to add a central remark: of course the "human" player can enrich and fecundate his strategies by referring to internal resources not necessarily directly related to the previous experience with Go, but with other preexistent skills belonging to disparate areas of cognition. This is the reason why we can say that the strategies of a "human" player present a less degree of closure with respect to the automatic player AlphaGo. In humans, strategies are locked with respect to the external rigid scenario, but more open with respect to the mental field of reference to previous wide strategic experiences; in AphaGo and in deep learning systems, the strategic reservoir cannot—at least currently—take advantage of that mental openness and flexibility typical of human beings: the repertoire is merely formed/learned to play the game by checking data of thousands of games, and no other sources.

I have also to say that the notion of cognitive locked strategy I am referring to here is not present in and it is unrelated to the usual technical categorizations of game theory. Fundamentally, in combinatorial game theory, Go can be technically illustrated as zero-sum (player choices do not increment resources available-colloquially), perfect-information, partisan, deterministic strategy game, belonging to the same class as chess, checkers (draughts) and Reversi (Othello). Moreover, Go is bounded (every game has to end with a victor within a finite number of moves and time), strategies are obviously associative (that is in function of board position), format is of course non-cooperative (no teams are allowed), positions are extensible (that is they can be represented by board position trees).[9]

3 LOCKING ABDUCTIVE COGNITIVE CAPACITIES JEOPARDIZES THE MAXIMIZATION OF ECO-COGNITIVE OPENNESS

As I have already anticipated above in Section 1, in my research I have recently emphasized ([19] chapter 7) the *knowledge enhancing* character of abduction. This means that in this case the abductive reasoning strategies grant successful and highly creative outcomes. The knowledge enhancing feature regards several kinds of new gen-

[4] In a book published in Japan, related to the description of various strategies that can be exploited in Go games, Davies emphasizes the role of "reading ahead" [4].

[5] The word belongs to the Husserlian philosophical lexicon [9] I have analyzed in its relationship with abduction in ([16] chapter 4).

[6] Cf. Wikipedia, entry Go (game) `https://en.wikipedia.org/wiki/Go\_(game)`.

[7] An expressive adjective still used by Husserl [10]. Translated by D. Carr and originally published in *The Crisis of European Sciences and Transcendental Phenomenology* [1954].

[8] This expression, I have extendedly used in [14], is derived from Hutchins, who introduced the expression "mediating structure", which regards external tools and props that can be constructed to cognitively enhance the activity of navigating. Written texts are trivial examples of a cognitive "mediating structure" with clear cognitive purposes, so mathematical symbols, simulations, and diagrams, which often become "epistemic mediators", because related to the production of scientific results [11]. that function as an enormous new source of information and knowledge.

[9] Cf. Wikipedia entry Go (game) `https://en.wikipedia.org/wiki/Go\_(game)`.

erated knowledge of various novelty degrees, from that new knowledge about a suffering patient we have abductively accomplished in medical diagnosis (a case of selective abduction, as no new biomedical knowledge is created, just new knowledge about a person) to the new knowledge developed in scientific discovery, which many epistemologists celebrated, for example Paul Feyerabend in *Against Method* [5]. In the case of an artificial game such as Go, the knowledge activated thanks to an intelligent choice of already available strategies or thanks to the invention of novel strategies and/or heuristics must also be considered a result of knowledge enhancing abduction.

I strongly contend that, to arrive to uberous selective or creative optimal abductive results, useful strategies must be applied, but it is also needed to be in presence of a cognitive environment marked by what I have called *optimization of eco-cognitive situatedness*, in which eco-cognitive openness is fundamental [18]. This feature of the cognitive environment is especially needed in the case of strong creative abduction, that is when the kind of novelty is not restricted to the case of a "simple" successful diagnosis. In Section 4, I will illustrate in more detail that, to favor good creative and selective abduction reasoning, cognitive capacities have to be *freed* thanks to inferential strategies which are not "locked" in an external restricted eco-cognitive environment, such as in a scenario characterized by fixed defining rules and finite material aspects, which would function as cognitive mediators able to constrain agents' reasoning.

At this point, it is valuable to furnish a short presentation of the concept of *eco-cognitive openness*. Surely an updated logic of abduction consists in what has been called "naturalization" of the well-known fallacy "affirming the consequent": in my recent research on abduction [17, 18, 19], I emphasized the importance in good abductive cognition to hypotheses of what has been called *optimization of situatedness*. Let us explain what is the meaning of the expression optimization of situatedness: abductive cognition is for example very important in scientific reasoning because it refers to that activity of creative hypothesis generation which characterizes one of the more valued aspects of rational knowledge. To get abductive results in science, the "situatedness" of the involved cognitive activities is strongly connected with eco-cognitive aspects, related to the contexts in which knowledge is "traveling": in the case of scientific abductive cognition (but also in other abductive cases, such as medical diagnosis) to favor the solution of an inferential process the situatedness also has to be characterized by the richness of the flux of information, which in many cases (surely in the case of scientific reasoning and discovery) has to be maximized. This maximization aims at a certain optimization of situatedness, which, as I quoted above, can only be made possible by a *maximization of changeability* of the basic starting data which inform the abductive cognitive process: inputs have to be maximally enriched, rebuilt, or modified and the same has to occur with respect to the knowledge applied during the hypothetical reasoning process. The aim is to have at disposal a favorable "cognitive environment" in which available data can become *optimally positioned*.[10]

In summary, abductive processes to hypotheses—in a considerable quantity of cases, for example in science—are highly *information-sensitive*, and face with a flow of information and data uninterrupted and appropriately promoted and enhanced when needed (of course also thanks to artefacts of various kinds). This means that also from the psychological perspectives of the individuals the epistemological openness in which knowledge channeling has to be favored is fundamental. [11]

4 LOCKED CAPACITIES BOUNDS CREATIVITY

Optimization of situatedness is related to cognitive capacities characterized by unlocked strategies. Instead locked strategies, such as the ones active in Go game, AlphaGo, and other computational AI systems and deep learning devices, do not favor the optimization of situatedness. Indeed, I have already contended above that, to obtain good creative and selective abductions, reasoning strategies must not be "locked" in bounded eco-cognitive surroundings (that is, in scenarios designed by fixed defining rules and finite material objects which would play the role of the so-called cognitive mediators). In this perspective, a poor scenario is certainly responsible for the minimization of the eco-cognitive openness and it is the structural consequence of the constitutive organization of the game Go (and also of Chess and other games), as I have already described in Subsection 2.2. I have said that in the game Go stones, board, and rules are rigid and so totally predetermined; what instead is undetermined are the strategies and connected heuristics that are adopted to defeat the adversary in their whole process of application. Of course, many of the strategies of a good player are already mentally present thanks to the experience of several previous games.

As I have already said, the available strategies and the adversary's ones are always *locked* in the fixed scenario: you cannot, during a Go game, play for few minutes Chess or adopt another rule or another unrelated cognitive process, affirming that that weird part of the game is still appropriate to the game you agreed to play. Your cognitive capacities are constrained. You cannot decide to change the environment at will so *unlocking* your strategic reasoning, for example because you think this will be an optimal way to defeat the adversary. Furthermore, your adversary cannot activate at his discretion a process of eco-cognitive transformation of that artificial game. On the contrary, in the example of scientific discovery, the scientist, or the community of scientists, frequently recur to disparate external models and change their reasoning strategies[12] to produce new analogies or to favor other cognitive useful procedures (prediction, simplification, confirmation, etc.) to enhance the abductive creative process.

The case of scenarios in human scientific discovery precisely represents the counterpart of the ones that are poor from the perspective of their eco-cognitive openness. Indeed, in these last cases, the reasoning strategies that can be endorsed (and also created for the first time), even if multiple and potentially infinite, are *locked* in a determined perspective where the components do not change (the stones can just diminish and put aside, the board does not change, etc.) I would say that in scenarios in which strategies are locked, in the sense I have explained, an *autoimmunization* [22, 1] is active, that constitutes the limitations that preclude the application of strategies

[10] I have furnished more cognitive and technical details to explain this result in [18, 19].

[11] A note on the history of philosophy can be added: already Aristotle provided a first fundamental study on abduction, which stresses the relevance, we can hazard, of *non-locked*, but highly open, cognition, in the celebrated (by Charles Sanders Peirce) passage of Chapter B25 of *Prior Analytics* regarding ἀπαγωγή (that is abduction, translated, in the English edition, with "leading away"). Indeed, it is exactly the idea of "leading away" which expresses that in smart abductions we have to integrate (or "unlock") the given components of the cognitive environment with the help of other cognitive tools and data that are away from them. I think that in Aristotle some of the current central aspects of abductive cognition are already present, and they are in tune with the EC-Model (Eco-Cognitive Model) of abduction I have introduced in [16, 17, 18, 19].

[12] Many interesting examples can be found in the recent [23].

that are not related to "pre-packaged" scenarios, strategies that would be foreigners to the ones that are strictly intertwined with the components of the given scenario. Remember I already said that these components play the role of *cognitive mediators*, which anchor and constrain the whole cognitive process of the game.

To summarize and further explain (by linking the problem of locked and unlocked strategies to the various cases of selective and creative abduction):

1. Contrarily to the case of those high level "human" capacities that are characterized by creative abductive inferences such as the ones expressed by scientific discovery or other examples of special exceptional intellectual results, the status of artificial games (and of their computational counterpart) is very poor from the point of view of the non-strategic knowledge involved. We are dealing with stones, a modest number of rules, and one board. When the game progresses, the shape of the scenario is stunningly modified but no unexpected cognitive mediators (objects) are appearing: for example, no diversely colored stones, or a strange hexagonal board. On the contrary, to continue with the example of high levels creative abductions in scientific discovery (for example, in empirical science), first of all the evidence is extremely rich and endowed with often unexpected novel features (not only due to modifications of aspects of the "same things", as in the case of artificial games). Secondly, the flux of knowledge at play is multifarious and is related to new analogies, thought experiments, models, imageries, mathematical structures, etc. that are rooted in heterogeneous disciplines and fields of intellectual research. In sum, in this exemplary case, we are facing with a real tendency to a status of optimal eco-cognitive situatedness (further details on this kind of creative abduction are furnished in [17, 18, 19]).
2. What happens when we are dealing with selective abduction (for example in medical diagnosis)? First of all, evidence freely and richly arrives from several empirical sources in terms of body symptoms and data mediated by sophisticated artifacts (which also change and improve thanks to new technological inventions). Second, the encyclopedia of biomedical hypotheses in which selective abduction can work is instead locked,[13] but the reference to possible new knowledge (locally created of externally available) is not prohibited, so the diagnostic inferences can be enhanced thanks to scientific advancements at a first sight not considered. Third, novel inferential strategies and linked heuristics can be created and old ones used in new surprising ways but, what is important, strategies are not locked in a fixed scenario. In sum, the creativity that is occurring in the case of human selective abduction is poorer than the one active in scientific discovery, but richer than the one related to the activity of the locked reasoning strategies of the Go game and AlphaGo, I have considered above.
3. In Go (and similar games) and in deep learning systems such as AlphaGo, in which strategies and heuristics are "locked", these are exactly the only part of the game that can be improved and rendered more fertile: strategies and related heuristics can be used in a novel way and new ones can be invented. Anticipations as abductions (which incarnate the activities of "reading ahead") just affect the modifications and re-grouping of the same elements. No other types of knowledge will increase; all the rest remains stable.[14] Of course, this dominance of the strategies is the quintessence of Go, Chess, and other games, and also reflects the spectacularity of the more expert moves of the human champions. However, it has to be said that this dominance is also the reason that explains the fact the creativity at stake is even more modest than the one involved in the higher cases of selective abduction (diagnosis). I will soon illustrate that this fact is also the reason that explains why the smart strategies of Go or Chess games can be more easily simulated, for example with respect to the inferences at play in scientific discovery, by recent artificial intelligence programs, such as the ones based on deep learning.[15]

The reader does not have to misunderstand me: I do not mean to minimize the relevance of creative heuristics as they work in Go and other board games. John Holland already clearly illustrated [7, 8] that board games such as checkers, as well as Go, are wonderful cases of "emerging" cognitive processes, where potentially infinite strategies favor exceptional games: even if simply thanks to a few rules regulating the moves of the pieces, games cannot be predicted starting from the initial configurations. While other cases of emerging cognitive processes (I have indicated the example of scientific discovery) characterize what can be called "vertical" creativity (that is, related to unlocked strategies), board games are examples of "horizontal" creativity: even if board games are circumscribed by locked strategies that constrain the game, "horizontal" creativity can show astonishing levels of creativity and skilfulness. We already said that these extraordinary humans skills have been notably appropriated by artificial intelligence software (see below the last paragraphs of this section): the example given in this article is the one of AI deep learning heuristics that were able to *learn from* human games. What are the remaining most important effects which derive from these computational AI programs equipped to concretize cognitive abductive inferences characterized by "locked" strategies?

I think humans with their biological brains do not have to feel mortified by these extraordinary skillful capacities of the AI programs. Unfortunately, given the present worldwide status of mass media, other magnificent human performances in various fields of creativity, much more creative than the ones related to locked strategic reasoning, are unable to reach the global echo AlphaGo gained. Indeed, human-more-skillful-abductive creative capacities, related to unlocked strategies, as I have tried to demonstrate in this article—still cognitively beautiful—are not sponsored by Google, which is a herculean corporation that can easily obtain the attention of not only the monocultural media of our age, but also of the social networks: many human beings are more easily impressionable by the "miracles" of AI, robotics, and information technologies, than by prodigious knowledge results of human beings-like-us, too often out of sight (after all—ça va sans dire—also AI traditional programs and AI deep learning systems have been created by humans. . .)

Google managers also think that AI deep learning programs similar to AlphaGo could be exploited to help science resolve important real-world problems in healthcare but also in other fields. This would be a good research program. Google seems to also expect to implement some business thanks to a commercialization of new deep learning AI powers to collect appropriate information and generate abductions in some advantageous fields. Simply checking the Wikipedia entry DeepMind (`https://en.wikipedia.org/wiki/DeepMind`),[16] [DeepMind is a British artificial intelligence company instituted in September 2010 and took by Google in 2014,

[13] It is necessary to select from pre-stored diagnostic hypotheses.

[14] Obviously, for example, new rules and new boards can be proposed, so realizing new types of game, but this chance does not jeopardize my argumentation.

[15] Some notes on the area of the so-called automated scientific discovery in AI cf. ([16] chapter 2, section 2.7 "Automatic Abductive Scientists").

[16] Date of access 10 of January, 2019.

the company created the AlphaGo program].

Indeed, even if based on what I called in this article locked strategies, and thus far from the highest levels of human creativity, AI deep learning system and various other programs can also offer chances for business and a good integration in the market. I think epistemologists and logicians have to monitor the use of these AI devices (of course, when less transparent than the natural and limpid—and so stupefying—performance of AlphaGo in games against humans). Recent research in the field of epistemology, cognitive science, and philosophy of technology[17] illustrate that good AI software, which surely furnishes a big new chance for opportunity and data analytics, can be transmuted in a tool that does not respect epistemological and/or ethical rigor. For example, in the case regarding the computational exploitation of big data, outcomes can inadvertently lead to epistemologically unacceptable computer-discovered correlations (instead possibly good from a commercial perspective), but these tools are sometimes—unfortunately—seriously illustrated as aiming at replacing tout-court human-based scientific research as a guide to understanding, prediction and action. Calude and Longo say: "Consequently, there will be no need to give scientific meaning to phenomena, by proposing, say, causal relations, since regularities in very large databases are enough: 'with enough data, the numbers speak for themselves' " ([3] p. 595). Unfortunately, some "correlations appear only due to the size, not the nature, of data. In 'randomly' generated, large enough databases too much information tends to behave like very little information" (*ibid.*). I agree with these authors: we cannot treat some spurious correlations as results of deep scientific creative abduction, but just as trivial generalizations, even if reached with the help of sophisticated artifacts.[18] I cannot further deepen the problems regarding issues connected to the impact of computational programs on ethics and society. In this article, I limit myself to deal with cognitive, logical, and epistemological aspects to the aim of introducing the distinction between human and machine capacities characterized by locked and unlocked strategies and its meaning with respect to intelligent computation.

5 CONCLUSION

In this article, with the help of the concepts of locked and unlocked strategies, abduction, and optimization of eco-cognitive openness, I have illustrated some central aspects of the character of cognitive capacities dominated by different reasoning strategies and related heuristics, to the aim of shedding new light on the epistemological aspects of deep learning machines. Taking advantage of my studies on abduction, I have contended that what I call *eco-cognitive openness* is weakened in the case of famous computational programs such as AlphaGo, because their cognitive capacities are governed by *locked abductive strategies*. Instead, unlocked abductive strategies, which are in tune with what eco-cognitive openness requires, qualify those high-level kinds of abductive creative reasoning that are typical of human-based cognition. Locked abductive reasoning strategies are much simpler than unlocked ones to be rendered at the computational level: they indeed take advantage of what I called a *autoimmunity* that grants the limitations that preclude the application of strategies that are not related to "pre-packaged" scenarios, strategies that would be foreign to the ones that are strictly intertwined with the components of a given scenario.

[17] Relatively recent bibliographic references can be found in my book [15].

[18] On this problem and other negative epistemological use of computational programs, cf. the recent [3].

ACKNOWLEDGEMENTS

Research for this article was supported by the PRIN 2017 Research 20173YP4N3—MIUR, Ministry of University and Research, Rome, Italy. Parts of this article are excerpted from section 1.2, 2-4 of L. Magnani, AlphaGo, Locked Strategies, and Eco-Cognitive Openness, *Philosophies* 2018, 4, 8, MDPI, Open Access Publication distributed under the Creative Commons Attribution License.

REFERENCES

[1] S. Arfini and L. Magnani, 'An eco-cognitive model of ignorance immunization', in *Philosophy and Cognitive Science II. Western & Eastern Studies*, eds., L. Magnani, P. Li, and W. Park, volume 20, 59–75, Springer, Switzerland, (2015).

[2] P. D. Bruza, R. J. Cole, D. Song, and Z. Bari, 'Towards operational abduction from a cognitive perspective', *Logic Journal of the IGPL*, **14(2)**, 161–179, (2006).

[3] C. S. Calude and G. Longo, 'The deluge of spurious correlations in big data', *Foundations of Science*, **22**(3), 595–612, (2017).

[4] J. Davies, *Tesuji. Elementary Go Series. 3.*, Kiseido Publishing Company, Tokyo, 1995.

[5] P. Feyerabend, *Against Method*, Verso, London-New York, 1975.

[6] P. Gärdenfors, *Conceptual Spaces: The Geometry of Thought*, The MIT Press, Cambridge, 2000.

[7] J. H. Holland, *Hidden Order*, Addison-Wesley, Reading, MA, 1995.

[8] J. H. Holland, *Emergence: From Chaos to Order*, Oxford University Press, Oxford, 1997.

[9] E. Husserl, *Ideas. General Introduction to Pure Phenomenology* [First book, 1913], Northwestern University Press, London and New York, 1931. Translated by W. R. Boyce Gibson.

[10] E. Husserl, *The Crisis of European Sciences and Transcendental Phenomenology* [1954], George Allen & Unwin and Humanities Press, London and New York, 1970. Translated by. D. Carr.

[11] E. Husserl, 'The Origin of Geometry (1939)', in *Edmund Husserl's "The Origin of Geometry"*, ed., J. Derrida, 157–180, Nicolas Hays, Stony Brooks, NY, (1978). Translated by D. Carr and originally published in [**?**], pp. 353-378.

[12] E. Hutchins, *Cognition in the Wild*, The MIT Press, Cambridge, MA, 1995.

[13] D. Lenat, 'Discovery in mathematics as heuristic search', in *Knowledge-Based Systems in Artificial Intelligence*, eds., R. Davis and D.B. Lenat, McGraw Hill, New York, (1982).

[14] R. K. Lindsay, B. Buchanan, E. Feingenbaum, and J. Lederberg, *Applications of Artificial Intelligence for Organic Chemistry: the Dendral Project*, McGraw Hill, New York, 1980.

[15] L. Magnani, *Abduction, Reason, and Science. Processes of Discovery and Explanation*, Kluwer Academic/Plenum Publishers, New York, 2001.

[16] L. Magnani, *Morality in a Technological World. Knowledge as Duty*, Cambridge University Press, Cambridge, 2007.

[17] L. Magnani, *Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*, Springer, Heidelberg/Berlin, 2009.

[18] L. Magnani, 'The eco-cognitive model of abduction. ᾿Απαγωγή now: Naturalizing the logic of abduction', *Journal of Applied Logic*, **13**, 285–315, (2015).

[19] L. Magnani, 'The eco-cognitive model of abduction. Irrelevance and implausibility exculpated', *Journal of Applied Logic*, **15**, 94–129, (2016).

[20] L. Magnani, *The Abductive Structure of Scientific Creativity. An Essay on the Ecology of Cognition*, Springer, Cham, Switzerland, 2017.

[21] L. Magnani, 'Playing with anticipations as abductions. strategic reasoning in an eco-cognitive perspective', *Journal of Applied Logic – IfColog Journal of Logics and their Applications*, **5**(5), 1061–1092, (2018). Special Issue on "Logical Foundations of Strategic Reasoning" (guest editors W. Park and J. Woods).

[22] L. Magnani, 'AlphaGo, locked strategies, and eco-cognitive openness', *Philosophies*, **4**(1), 8, (2019).

[23] L. Magnani and T. Bertolotti, 'Cognitive bubbles and firewalls: Epistemic immunizations in human reasoning', in *CogSci 2011, XXXIII Annual Conference of the Cognitive Science Society*, eds., L. Carlson,

C. Hölscher, and T. Shipley, Cognitive Science Society, Boston MA, (2011).
[24] *Handbook of Model-Based Science*, eds., L. Magnani and T. Bertolotti, Springer, Switzerland, 2017.
[25] A. Newell, J. C. Shaw, and H. A. Simon, 'Empirical explorations of the logic theory machine: a case study in heuristic', in *Proceedings of the Western Joint Computer Conference [JCC 11]*, pp. 218–239, Los Angeles February, (1957).
[26] M. Ramoni, M. Stefanelli, L. Magnani, and G. Barosi, 'An epistemological framework for medical knowledge-based systems', *IEEE Transactions on Systems, Man, and Cybernetics*, **22(6)**, 1361–1375, (1992).
[27] P. Thagard, *Computational Philosophy of Science*, The MIT Press, Cambridge, MA, 1988.

Typifications, Play, and Ritual

Michael Barber[1]

Abstract. Recent anthropological analyses (e.g., [4], [7], [13], [14]) have suggested similarities between play and rituals, and have even gone further to interpret ritual as derivative from play in accord with R. Bellah's claim that "Ritual is the primordial form of serious play in human evolutionary history" [3]. Perhaps, though, such an exploration of origins, instead of tracing back play and ritual to an evolutionary history that begins with other species, might rather undertake a genetic phenomenological investigation, excavating strata of meaning, which activities such as play or ritual both presuppose and which can be located in the everyday life-world. The phenomenological tradition of Edmund Husserl and Alfred Schutz has articulated the structures of that everyday world from which entire spheres of activity (referred to by Schutz as "finite provinces of meaning"), such as theory itself, phantasy, dreaming, play, religion, or art, arise and distinguish themselves. Typifications, that is, typical, regularized patterns of acting and interpreting, characterize everyday life and other provinces of meaning, and this paper will show how play constitutes itself on the basis of such typifications and how ritual also depends on typified patterns and modifications of the experience of play. Ritual and play reveal a pervasive creativity in the lives of animals and humans.

1 TYPIFICATIONS IN THE SCHUTZIAN LIFE-WORLD

In Schutz's account of the life-world, typifications, that is, typical patterns of acting, including the rules and classification systems of language by which objects are typified (e.g. as apples, dogs, etc.), interact with one's interests or values, more or less ranked and systemized and known as "relevances" (specifying what is of relevance to one) [19]. In everyday life, one is above all pragmatically engaged, seeking to master one's world to one's own satisfaction (and in more reflective cases in accord with the overall meaning of one's life in the face of the fundamental anxiety one faces about one's death). Even the child's haphazard movements, when they yield satisfactions (e.g., in breast-feeding), result in those satisfactions becoming relevant, and the infant has the sense that she "can do it again," that is, repeat the behavior (pursing one's lips and sucking) that has previously brought about such satisfactions. Hence the infant acquires in its stock of knowledge (in a non-discursive sense) the easily-revocable typification that this type of behavior enacted in the presence of one's mother's breast can yield such satisfaction. Husserl amplifies this point when he claims that when a child learns what a scissors is and how to use it one single time, she is able from then on to recognize any future scissors she encounters and put it to use—the typification of "scissors" becomes sedimented in her stock of knowledge and partially constitutive of it. Any future encounter with scissors can evoke this typification not through inference, reflection, or ratiocination, but simply through passive synthesis or association, immediately issuing in action.

Likewise, the infant might enjoy the warmth of being held in its parent's arms, but suppose it finds itself physically on the other side of the room from its parents, the space between it and its parents then becomes a kind of imposed relevance, that is, it now becomes relevant to come to terms with that distance, to work around it, in order to rejoin its parents (its ultimate intrinsic relevance). Then, perhaps, it might put into practice the typical pattern of crawling (which it must have already acquired through hit or miss practice or the imitation of others) in order to overcome this imposed relevance of distance. Indeed, such human creativity regularly appears on a higher level when someone experiences an imposed relevance upsetting its typical ways of action (e.g. in the case of a disability) and yet learns to deploy other typical behaviors to come to terms with the imposed relevance that impedes its reaching its ultimate goal (or relevance). It is significant that, for Schutz's account of the pragmatic world of everyday life, bodily engagement (as working) with the world defines its form of spontaneity, and one's understanding of reality depends on a bodily location, a 0-point, from which one accesses other strata of reality by physical locomotion, by the movement from here to there, in which one seems repeatedly and preeminently to have the sense "I can do it again" [19]. It is as though the adult in pragmatic everyday life still retains as a kind of deep-seated corporeal memory, as a deep layer within its genetic constitution, the bodily processes through which people first typify their world and achieve primitive mastery through typified action.

2 PLAY'S LIFE-WORLDLY ROOTS IN TYPIFICATIONS

In the enactment (or performance) of typifications, one can find the life-worldly roots of both play and ritual. For instance, insofar as play is concerned, having utilized a scissors already and upon encountering a new pair, the child (without discursively knowing it) actually engages in a kind of experiment, anticipating that the new scissors, although differing in some respects from the earlier pair, will function as the previous pair did. One's successful execution of such a typification reinforces and strengthens one's future employment of that typification in regard to scissors and thereby heightens one's sense of the ability to manage the environment (in this small detail). In a sense the very enactment of a typification consists in facing an object that never exactly coincides with one's typification, and this lack of coincidence, this not according with one's typification (based on previous experience), this exceeding what the typification encompasses, presents a small kind of imposed relevance that one brings under one's control by employing the typification and finding it effectively functional.

This indeterminism or uncertainty, even if momentary, in imple-

[1] Saint Louis University, United States, email: michael.barber@slu.edu

menting typifications, that is, in being passively stimulated to bring the typification in one's stock of knowledge to bear on an object like and yet different from previous ones, is also reflected in higher level play and games whose outcome may never be completely predictable [10], [17]. When one's typification proves effective, a child at some level acquires some mastery over its pragmatic environment, and, perhaps at some level feels satisfaction, increased confidence, or even inchoate joy. In like manner, when children engage in "pretend-play," such as acting as if a banana were a telephone receiver and holding it up to their ear, they typify the banana as a kind of telephone receiver (on the bases of their similar shapes), with a kind of anticipation and playfulness. This typification proves successful, demonstrates a kind of mastery, and confers a kind of joy insofar as this entire process seems to exhibit, through mimicry, the "same" (typical, like and yet different) behavior that their parents do when speaking on the phone.

Like children, animals, too, engage in pretend play, as when, for instance, young female chimpanzees cradle sticks, tree bark, or small logs (until they have real offspring), by projecting themselves into a typified type of behavior like (and unlike) that engaged in by actual chimpanzee mothers. Moreover, one might speculate that such chimpanzee play-behavior, a kind of "virtual" mothering behavior, satisfies the young chimps by providing them with a sense of environmental mastery insofar as such behavior conforms with that of "actual" chimpanzee mothers. In addition, animals repeatedly engage in play, pouncing on each other, and more powerful dogs or cats, for instance, allow themselves to be attacked by weaker animals, exchanging roles, biting each other softly, or wrestling with claws retracted. It is as though such animals project themselves into a *virtual* combat situation—and of course, when any animal encounters an *actual* attack by another animal, it, no doubt, faces an imposed relevance, interfering with any other relevances the attacked animal may be pursuing (e.g. finding food) and requiring it immediately to master its perilous situation. In playful animal conflict, it is as though the animals playfully generate a virtual battle, an imposed relevance, just for the sake of engaging in other typical behaviors (e.g. virtual biting, virtual scratching) and virtually coming to terms with that imposed relevance—as if the struggle and feeling of resultant success were of pleasure all by themselves. Animal and human enjoyment of pretending, then, demonstrates the playful dimensions of deploying typifications, namely, the passive creative synthesizing (without rational inference) by which animals and humans conjoin under a typification an object or event previously encompassed by it and something like that object or event, but also dissimilar to it and even exceeding it (e.g., a virtual combat). Such a playful deployment of typifications reenact both the challenge of coming to terms with what might seem to interrupt one's regularized schemes of interpretation or pragmatic mastery and also the enjoyment of striving to overcome such impediments and actually overcoming them—even if all this is done only virtually. Typifications pertain to the very structure of play [2], [4], [7], [20].

3 RITUAL'S LIFE-WORLDLY ROOTS IN TYPIFICATIONS

Ritual, by contrast with play, faces a different kind of imposed relevance, which interrupts regularized, typified patterns and which one must grapple with by means of a new level of typifications. For ritual, imposed relevances are not benign interruptions to be overcome pleasurably by playful forays, but rather painful disruptions in desperate need of being repaired or brought into some kind of equilibrium.

In examining how animals transform their ordinary behaviors into ritualized ones (like ritual dances by which animals demarcate their territories against invaders) in "uncertain or conflicted circumstances," Ellen Dissanayake identifies a similar, human ritualized behavior that is "biologically" based, universal, and cross-culturally observed and reparative: the gestural and vocal interactions between a mother and an anguished infant that are drawn from adult contexts of affinity and intimacy and that are rather typical, such as mutual gazes, soft and high-pitched sounds, sympathetic touching, pats, hugs, and kisses [7]. Likewise, Roy Rappaport refers to psychologist Erik Erikson's suggestion that the pre-verbal infant's experience of its mother resembles the worshiper's experience of God according to Rudolph Otto. While the mother is "mysterious, tremendous, overpowering, loving, and frightening," trust and calm is restored through regular stereotyped "daily rituals of nurturance and greeting" between mother and infant [16]. One can imagine an infant, who is typically experiencing comfort, being interrupted by an imposed relevance, such as hunger, physical pain (e.g., gas, or teething pains), or psychological terror about losing its mother. The imposed relevance here cannot be seen as opening an opportunity for a typically playful response, but rather there is a need for security, reassurance, calm—and one can imagine a mother's cooing or rubbing the child's back or kissing—all stereotypical, typified patterns that mothers universally deploy—restoring at a bodily level a tranquility that can also transform the infant's psyche.

Here the mother makes use of typical, tender physical responses to pacify and come to grips with an infant's terrifying, disruptive imposed relevance, reiterating thereby the structure in which a rupture upsets comfortable typical ways of proceeding only to be healed or removed through newly developed typified patterns, as when the child crawls to be with its distant parents or when someone comes to terms with a disability by new typified behaviors (e.g., by using a dictating machine to take notes when one's writing arm is incapacitated). Whether this generalized typification structure will be taken in the direction of play or toward ritual depends, though, on the type of the imposed relevance in question and how it is interpreted, whether it poses merely a challenge to be playfully overcome or whether it is experienced as a terrifying collapse that desperately requires typical behaviors to calm a distraught soul and/or to restore a lost order. Here the model of typical ritual gestures might be traced back to an almost universally experienced, intimate, foundational corporeal stratum, namely the typical soothing gestures of a mother allaying the infant anxieties that disruption imposes [11].

For instance, Dissanayake affirms that rituals "*relieve individual and group anxiety* by instilling confidence and fostering a sense of control over disturbing circumstances" [7]. Iain Morley also observes that rituals can reduce stress [13]. Regularized, physical ritual acts and movements, which are themselves typified patterns of action, precisely help produce such tranquility. Thus, the Manus, as Margaret Mead noted, chant monotonous tones together when cold and frightened, and Trobrianders produce singsong melodies during a terrifying storm, as Bronisław Malinowski observed. Furthermore, Dissanayake comments on how, because of ritual actions "performed in a coordinated fashion with others," those practicing such rituals "were psychologically comforted and felt relieved of tension," thereby drawing on "a 'behavioral reservoir' that existed in mother-infant interactions" [7]. Singing, dancing, even breathing in unison, rhythmic movements of any sort, the beating of drums approximating the physical tempo of heartbeats, the highly repeated and unchangeable prayers and movements—all typical of rituals—play a role in reducing tension in much the way that ritualized, repeated, smooth, physical caresses by a mother quiet a troubled infant [4], [16]. Kyriakidis compares ritual practice to what Mihály Csikszentmihalyi describes as "flow" [6] in which "one acts with a deep but effortless involvement that removes from awareness the worries and frustrations of everyday life. *Flow* assists attention and therefore learning by helping to focus effortlessly, whilst creating a deep sense of enjoyment" [10].

To locate such calming ritual behaviors in the pattern of the Schutzian life-world, that is, as typified responses to imposed relevances with which one must come to terms, it is important to consider examples of occurrences [impositions, interfering with the regular pursuit of (intrinsic) relevances and creating anxiety or even terror] in relation to which rituals are celebrated. Major annual calendarial changes, such as the disquieting transitions to a new year or passage into spring or fall, are marked by ritual celebrations, as are major, apprehensiveness-inducing transitions of life, such as passing from childhood to adulthood or from adolescence to motherhood. Victor Turner explains how liminal ritual celebrations accentuate such transitions with all their fearfulness as a prelude to facilitate a peaceful accommodation to a new life [21]. Rituals celebrate precarious tribal or national moments and diminish fears, as when a new chief is installed or a king crowned. Prehistoric Maltese groups celebrated rituals in relationship to the excess or scarcity of food supplies (in particular, of the island's animal population), and the Maring plant trees as part of the *rumbim* that celebrates the end of warfare, and they ritually sacrifice pigs when the supply of pigs becomes excessive and dangerous [12], [16]. Likewise, the Sioux engage in rituals to uphold Wakan-Tanka, that is, "the true, moral, eternal, harmonious, encompassing, unitary order" [16], and throughout history rituals are regularly used to domesticate impulses and to reinforce moral orders, perhaps always in danger of deteriorating or collapsing. Finally, as another example, rituals reassured the impoverished and threatened Bog Irish immigrants in England, according to Mary Douglas [8].

Clearly there are differences between the "imposed relevances" with which play and ritual deal: between relevances offering a challenge to be surmounted playfully (in which the play itself may be of more value than any outcome) and those that pose a cosmic, momentous, or communal threat to be addressed in ritual.

Several authors discuss how ritual counteracts the everyday life efforts to master one's environment through individualistic, rational, this-worldly strategies. Rappaport, for example, suggests that ritual usually involves an effort to re-establish a lost sense of *communitas* among participants. Further, rituals incorporate properties such as formality, invariance, canonicity, and perdurance to such an extent that ritual participants feel that they are partaking in activities that they do not spontaneously produce but that have been handed down to them by preceding generations and they thereby are part of a community much larger than themselves and their present [16]. Gregory Bateson points out further that purposive rationality cannot comprehend the wholeness of the world and so it finds itself incapable of addressing many of the uncontrollable imposed relevances that rituals handle (e.g. such as death, kinds of suffering and tragedy, needs for deep healing, irreconcilable differences), and, as such, it is prone, on its own, to become "pathogenic and destructive of life" [1]. Perhaps because of the limits of rationality to diminish anxieties produced by such large-sized, imposed relevances or even the prospect of them, ritual behaviors tend to dive beneath the level of rational thinking and to immerse themselves in corporeal rhythms, dance, and music to sooth the troubled spirit, just as a mother's caresses and sounds pacify the infant in a way that no ratiocination with the infant would be able to do (if it were even possible). Finally, when communities reach the end of their resources to come to terms via pragmatic and rational techniques with the everyday, this-worldly imposed relevances of cosmic proportions, they turn to the ritual sphere, in which they look to a power beyond themselves to entrust their fate to and to relieve anxieties, giving up (and entrusting over) to it their very effort to come to terms.

Whether such surrender entails fatalism, authoritarianism, or the suppression of individuality, or whether it relieves one or one's community from paralyzing anxieties and makes possible a renewed courageous engagement with those imposed relevances to make them somewhat manageable are further questions. It is of interest, though, that, in the cases of both play and ritual, when imposed relevances threaten the everyday typifications by which one masters life, both humans and animals seem able to revert to and indicate by signaling to each other entrance into an alternative reality that many anthropologists recognize has "different ontological status" (see [13], [15]) and that Schutz describes in terms of "finite provinces of meaning," distinctive from everyday life. Such provinces of meaning, though, have their own distinctive *epochés* by which one enters the province, tensions of consciousness, sense of self, and forms of spontaneity[2]—in sum, a distinctive type of activity with its own battery of distinctive typifications. Just as the mother resorts to typified actions (caresses, soft talk with certain accents and tones) to console the child, agitated by its ruptured expectations, so one resorts to typified behaviors in typified provinces of meaning (like play and ritual) in response to imposed relevances. When individuals or groups turn to the typifications (and relevances) of diverse provinces of meaning in this way—it appears as though it is impossible ever to escape typifications, whether they are everyday typifications disappointed by unforeseen impositions or whether they are being used to come to terms on a higher level with those disruptions.

4 IMAGINATION/CREATIVITY IN TYPIFICATION, PLAY, AND RITUAL

This effort to think via genetic phenomenology about how play and ritual can be traced back to everyday patterns of typification also reveals the creative, imaginative dimensions of human, and even animal, consciousness. In the case of typifications, it is as though the mere acquaintance with the pleasurable experience of breast feeding or with the handling of a scissors cannot just be stuck as just a one-time occurrence, inertly there in one's experience; but rather these experiences by being immediately typified equip their recipients with a creative potential to deal productively with future experiences of similar objects. Such objects of future experience are not taken just to be there like dumb objects, isolated from other objects and events, even if those objects and events are not exactly the same as what one already has experienced, but actors through typifying

[2] Patrick Bateson describes how dogs signal their readiness for play by dropping down on their forelegs and wagging their tales, how cats crouch their heads low, arch their backs, and paddle their back legs, and how chimpanzees put on a special "play face" [2]; and Dissanayake discusses how such movements indicate a desire to enter an "as if" or "other world" or "meta-reality" that differs from present reality [7]. These "announcements of a new reality" parallel the *epochés* by which, according to Schutz, one departs everyday reality and embarks upon an alternative province of meaning [19]. Of course, a question here is how animals engage in "provinces of meaning" that one might have thought were available only for human beings. In addition, one can find parallels between Schutz's descriptions of finite provinces of meaning in these discussions of play and ritual, whether one is talking about humans or animals. For instance, play and ritual both serve no functional purposes in contrast to the everyday life world and break with everyday life through a kind of *epoché*, as we have seen (see [4], [9], [10], [13], [16], [17]), and they engage in different goals than everyday life and hence have distinctive forms of spontaneity.

give them meaning and make possible future action. Even the passive synthesis by which one recognizes similarity displays creativity insofar as one brings one's previous experience to bear on a different but like individual. One not only refuses to see that new individual as just a distinct individual but also envisions it as more, as bearing a likeness to what one has previously experienced, as something with which an agent will be able to interact as he or she has done with previous experiences of like individuals, thereby enhancing future action. It is not surprising, then, that Dorion Cairns remarks that "the fundamental tendencies of mental life are tendencies to identify and to assimilate" [5]. In fact, this deployment of typifications exhibits a kind of experimental attitude toward life, in which like the sciences, one—perhaps without even being reflectively aware—takes one's typification as hypothetical, ready to see if it works out or if it is frustrated and does not seem to apply to the object or event at hand—in which case one will withdraw it or venture another. What is also of interest is that animals too exhibit such basic-level creativity by operating with typifications even though they do not reflect upon them.

Likewise, the mimicry of play, instanced when a child deploys a banana as a telephone receiver, effect an original synthesis that observers might not have expected between two quite separate objects sharing some level of similarity; one is reminded of poets whose inventive metaphors, artistically conjoining objects never thought together before, such as the "rosy-fingered dawn," evoke admiration. Likewise, animals producing virtual mothering or virtual conflict show themselves able to creatively transcend empirical givens, such as actual chimpanzee mothering or real life-and-death animal conflicts, transposing such experiences into phantasied, imagined domain, where a chimp carries a log or a "pseudo-fight" is enacted, without real biting or scratching and with claws retracted. That animals engage in such phantasied behaviors, in a way that seems purely for the sake of their enjoyment, indicates further a capacity for creative imitation, an ability to break out from the entire context of everyday life and the pragmatic motives governing it. Animals and humans appear then able to build a parallel reality and to not be bound by the seriousness of the everyday life world that we all start with.

And just as play involves even animals setting aside the everyday life-and-death struggle with another animal attacker and imaginatively replicating a phantasied imposed relevance of this kind of attack only for the seemingly pure enjoyment of a fictive struggle, so in ritual people, facing life-and-death imposed relevances of often cosmic proportions (such as the destructiveness of an earthquake or hurricane), can let go the misery resulting from such cataclysms and transpose themselves to a separate sphere, that of ritual. In ritual, the resilient spirit of the community is solidified, hope is provided, and space is made for positive, constructive responses. In a similar way, it is possible for mothers upon finding their children disturbed by physical or psychological distress from which they seem unable to escape to engage them with ritualized caressing or voice tones that remove misery and restore tranquility. In fact, ritual, as part of the finite province of meaning of religious experience runs parallel to many other non-pragmatic provinces of meaning, such as phantasy, dreaming, theoretical science, literature, and play—in all of which one or one's community creatively leap with freedom out of world of everyday life whose pressing pragmatic imperatives can easily hold us in bondage. From typifications to play to ritual, which pertain to a kind of continuum, one can see the great creativity and imaginative capacity of humans and animals to transcend the world that seems to allow no escape and to do so in a variety of ways that no one could have predicted. In this the non-pragmatic provinces of meaning resemble their prototype, Husserlian phenomenology, whose *epoché* broke open a novel, unexplored realm that the natural attitude never could have imagined.

REFERENCES

[1] G. Bateson, *Steps to an Ecology of Mind*, Ballentine, New York, 1972.

[2] P. Bateson, 'Play and creativity', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, University Press, 40–52, (2018).

[3] R. Bellah, *Religion in Human Evolution*, MIT Press, Cambridge, Massachusetts, 2011.

[4] G. Burghardt, 'The origins, evolution and interconnections of play and ritual: setting the stage', *Ritual, Play, Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 23–39, (2018).

[5] D. Cairns, 'Applications of the theory of sense-transfer', *Animism, Adumbration, Willing, and Wisdom: Studies in the Phenomenology of Dorion Cairns*, Zetabooks, Bucharest, 50–88, (2012).

[6] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*, Harper and Row, New York, 1990.

[7] E. Dissanayake, 'From play and ritualisation to ritual and its arts: sources of upper pleistocene ritual practices In lower middle pleistocene ritualized and play behaviors in ancestral hominins', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 87–98, (2018).

[8] M. Douglas, *Natural Symbols: Explorations of Cosmology*, Barrie and Jenkins, London, 1973

[9] Y. Garfinkel, 'Dancing with masks in the proto-historical near east', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 143–169, (2018).

[10] E. Kyriakidis, 'Rituals, games, and learning', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 302–308, (2018).

[11] L. Malafouris, 'Play and ritual: some thoughts from a material-culture perspective', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, Cambridge University Press, 311–315, (2018).

[12] C. Malone, 'Manipulating the bones: eating and augury in the Maltese temples', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 187–207, (2018).

[13] I. Morley, '"The pentagram of performance: ritual, play and social transformation', in *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 321–332, (2018).

[14] I. Morley, 'Pretend play, cognition and life-history in human evolution', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 66–86, (2018).

[15] R. Osborne, 'Believing in play and ritual', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 316–320, (2018).

[16] R. Rappaport, *Ritual and Religion in the Making of Humanity*, Cambridge University Press, Cambridge, 1999.

[17] C. Renfrew, 'Introduction: play as the precursor of ritual in early human societies', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 9–19, (2018).

[18] C. Renfrew, I. Morley, and M. Boyd, eds., *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 2018.

[19] A. Schutz, 'On multiple realities', *Collected Papers 1: The Problem of Social Reality*, Martinus Nijhoff, The Hague, 207–259, (1962).

[20] P. Smith, 'Pretend and socio-dramatic play in evolutionary and developmental perspective', *Ritual, Play, and Belief, in Evolution and Early Human Societies*, Cambridge University Press, Cambridge, 53–65, (2018).

[21] Victor Turner, *The Ritual Process: Structure and Anti-Structure*, Aldine Publishing Company, Chicago, 1969.

Habitual Behavior as a Bridge between I-intentionality and We-intentionality

Raffaela Giovagnoli [1]

Abstract. Habitual behavior represents a fundamental part of the nature of human beings in both the individual and social contexts. It presents two dimensions: "routine" and "goal-directed behavior" that organize human life and reduce its complexity. Habitual behavior could represent a plausible notion to bridge the gap between individual and joint intentions.

1 COLLECTIVE INTENTIONALITY AND THE "CENTRAL PROBLEM"

Intentionality is the propriety of the human mind to be directed at objects, state of affairs, goals and values. Collective Intentionality (CI) can be interpreted likewise and corresponds to that propriety of the human mind to be "jointly" directed at objects, states of affairs, goals and values. There are some important modes in which CI appears in everyday life: shared intention, joint attention, shared beliefs, collective acceptance, collective emotion. These topics are at the center of several cross-disciplinary researches. CI is the key-notion to understand the nature and structure of social reality and the very modalities that occur in human construction of the social world. Even though we can trace back accounts of social interactions, practices, social consciousness in the philosophical tradition, CI in the contemporary debate focuses on the conceptual and psychological features of joint or shared actions and attitudes i.e. actions and attitudes of groups or collectives, their relations to individual actions and attitudes, and their implications for the nature of social groups and their functioning. It addresses to the study of collective action, responsibility, reasoning, thought, intention, emotion, phenomenology, decision-making, knowledge, trust, rationality, cooperation, competition, and related issues, as well as their role in social practices, organizations, conventions, institutions, and ontology.

If I want to go to the cinema to see "The Wolf of Wall Street" tomorrow and you want to go to the cinema to see "The Wolf of Wall Street" tomorrow, does it mean that we have a collective intention? No, to have a collective intention does not mean to summate individual intentions. CI is irreducible to individual intentionality, and by virtue of this irreducibility CI can be attributed to participants *as a group*. Obviously, the fact that shared intentions are had by a group does not block attribution of the intentionality in question to the individuals. So, for instance, to say that a group intends to go for a walk is *the same as* saying that the participating individuals intend to go for a walk. Some philosophers criticize the Irreducibility Thesis and propose the Individual Ownership Thesis namely the basic claim that each individual has a mind of her own and has a sort of intentional autonomy that is incompatible with the view that individual minds are somehow fused when intentional states are shared.

Consequently, the central question in the field of CI is a plausible consideration of the ontology of individual agents and their psychological states and interactions. There are ontological (do group agents exist?), conceptual (how do we consider social concepts?), and psychological (how do we understand collective mental states?) dimensions that characterize the field of CI [1]. These questions are relevant to the traditional debate between methodological individualism and collectivism in the social sciences. We'll consider the role of habits in human individual and social ordinary life and we move from the fact that habitual behavior is fundamental to organize our activities in individual as well as in social contexts. Instead of considering classical and revised theories of intentionality, we prefer to focus on the notion of habit to understand the process reduction of the complexity of daily life, Habits play an important role and also in social life where we take part to informal joint practices as well as to institutionalized ones. We cooperate to create and to participate in social practices because we need to organize our life together with other people to create common spaces that have different functions and significance depending on the corresponding practice (for example, we all pay the ticket to take a train and many of us participate in religious rituals or similar activities).

---

[1] Faculty of Philosophy, Pontifical Lateran University, giovagnoli@pul.it

2 DIMENSIONS OF HABITUAL BEHAVIOR

The relationship between habits and rituals could provide the way to harmonise I-intentionality and We-intentionality. We begin with presenting a plausible sense for the notion of habit, which goes beyond the mere repetitive behavior or routine. Starting from Latin, there are two meanings for the English word "habit". The first is *Habitus*, that entails a deliberate disposition to act; the second is *Consuetudo*, that implies the constant repetition of an event or behavior without deliberation. The traditional philosophical sense of habit (*Habitus*) is introduced by Aristotle to characterize the notion of "virtue". Virtue can be considered as a habit in the sense of a disposition to deal with good or bad emotions and tendencies. Aristotle, conceived this notion of habit as a mechanism that is analogous to natural mechanisms, and somehow guarantees the uniform repetition of facts, acts, or behavior by eliminating or reducing effort and fatigue and so by making them pleasant.

We argue for a plausible account of the notion of habit that rests on some aristotelian thesis also by reference to researches in psychology and neuroscience. A habit is not only a mere automatism or a repetitive behavior, but also a stable disposition for action (practical skill), that implies the relationship between automatism and flexibility. The same process is involved in our participation and constitution of social informal and formal spaces [2] [3] [4] [5].

Habits have a very important function in individual life because they reduce the complexity of daily life; they make our daily life easier and pleasant. Naturally, we can control habits concerning the satisfaction of our basic natural needs. Depending from natural and social environment, we develop different habits that organize the way to satisfy our human needs. The difference between habits and automatism or simple routines is that the former give control over actions, while the latter don't [6]. According to this view, that crosses philosophy and neurobiology, the habit is a "stable disposition for self-development".

Graybiel observes a plausible relation between habits and goals because goals are explicitly present during action evolution and selection and they increasingly blur the more an action is repeated. Along this line, we find interesting studies in the ambit of neural-dynamic logic [7].We have examples of habits as fixed action patterns namely complex repetitive behavior in non-human animals and repetitive behavior and thoughts in human pathological conditions. She concludes that a habit completely disengaged from a goal becomes either a stimulus-response pair for a non-human animal or a pathological trait for human beings. Her theoretical contribution resides in the classification of habits as "neutral", "good" or "bad" where good habits seem to be those selected to guide our behavior and bad habits those that powerfully take control on our behavior. This categorization seems to make possible to include goals as drivers of habits. Graybiel also maintains that habits play an important role in social life; in this case they are "shaped" as mannerism and rituals.

The associationist view grounding William James's research seems far from explaining the complexity of human habits. Consequently, Bernacer and Murillo [8] underscore three important results from a deep study of the Aristotelian analysis of habits in *Nicomachean Ethics.* An acquired habit is an acquired disposition to perform certain types of actions; this disposition, usually acquired by means of repetition of one or more actions, makes the exception of these actions prompter, more spontaneous and autonomous from continuous supervision, all of which generally leads to a better performance. If the habit increases cognitive control of actions it can be termed a habit-as-learning; on the contrary, if it increases their rigidity it is a habit-as-routine. Habits-as-routines are fundamental for the cognitive enrichment of actions entailed by a variable amount of practice (efforts are required to engage in activities and performances). Differently, habits-as-learning are not merely acquisition of a way of acting; they entail a cognitive capacity connected to the habit that can flexibly used in different contexts. Habits-as-routines and habits-as-learning have a different relation to consciousness. Habits-as-routines represent a fully unconscious performance. Habits-as-learning reduce or eliminate consciousness of basic elements of the action in order to concentrate on higher goals, while preserving at all times the possibility of recovering them for conscious attention. It is worthy to underline the contribution of the Aristotelian distinction between good and bad habits, that intends good habits as those enhancing the agent's control to reach certain goals. Consequently, we can clarify the relation between habits and emotions. The habits-as-learning entail control and for this reason they are fundamental to reach personal goals. This is the process that favors the agent's pleasure and happiness.

Some authors intend the idea of "habit learning" as the performance of an action, previously learned after many repetitions namely in an unconscious manner, and whose execution is inflexible and independent to the outcome [9]. This perspective requires an integration with other perspectives that recognize the importance of the sensitivity to the outcome and of different levels of flexibility and feedback. According to Lombo and Giménez Amaya, a neurobiological view of “habit learning” and recent experimental contributions (especially those of Graybiel) are consistent with the Aristotelian notion of “habit”. Human habits are essentially based on two aspects: (a) the stable character of an acquired quality; and (b) the capacity for new actions that arises from that quality.

3 RITUALS AS SOCIAL HABITS

Recent studies from cognitive neuroscience, biology and psychology show converging perspectives on the organization of goal-directed, intentional action in terms of (brain, computational) structures and mechanisms. They conclude that several cognitive capabilities across the individual and social domains, including action planning and execution, understanding others’ intentions, cooperation and imitation are essentially goal-directed [10] [11] [12] [13]. To form habits we need goal representations both in the individual and social contexts. They have a crucial role in planning and control of action; moreover, action understanding and imitation are performed at the goal rather than the movement level. It seems that the motor system is highly engaged in anticipatory, simulative and generative processes. Some authors introduce an interesting speculative perspective, and make the case that the same predictive mechanisms provide both a “linkage with the future” required for taking goal-directed action, and a “linkage with others” required to act socially. We can observe a significative reformulation of basic concepts in cognitive and behavioral sciences, and a common theoretical view—a motor-based (or action-based) view of cognition — is emerging across disciplines. They provide a description of the abilities of action execution, its planning, and understanding of others’ intentions as essentially goal-directed and served by the same representations, which are action- oriented and involve deeply the motor apparatus.

Routines and goal-directed behavior characterize habits both in the case of individual and social behavior. We create our own habits while fulfilling our basic needs and desires. But, we are social beings and we need to organize our activities also to participate in different social practices. For example, rituals have the important function to create social spaces in which individuals can share emotions, experiences, values, norms and knowledge. The function to share experiences is fulfilled when there exist a social space created by cooperation for reaching a certain goal. If we want to get a positive result about the extension of habits in the social dimension we need to move from a sort of goal-directed activity that we can perform together. We create social habits in the form of rituals by using the “status function”, which is a peculiar kind of function from which we constitute the social world.. Rituals are characterized by two special features: (1) “collective intentionality” that expresses our social nature and (2) collective imposition and recognition of a status that deserve to concretely create institutions.

The “constitutive rule” is essential to the process of constitution of institutions in general [14]. The canonical form introduced by Searle is:

Status Function = X counts as Y in C

For instance, a certain expression counts as promise in a certain context C. So, it is fundamental to assign functions to objects and persons. We use ordinary language to represent state of affairs and norms, namely to understand what are the conditions of satisfaction of different speech acts (assertions, commands, promises etc.). Beyond the classical dimensions of syntax, compositionality and generativity, there is a fundamental dimension which generates public norms i.e. “deontology”, which is characterized by the speech act of “declaration”. For example, if we say “This is my house” or “This is my coach”, we not only represent a state of affairs, but we create a deontology which manifests itself in rights, obligations and duties as well as in the acceptation of the corresponding speech acts from the part of the interlocutors.

We pointed out the fundamental process of assigning functions to objects or to some non-physical entities, which is a form of symbolization aiming at creating institutional reality. This process is at the basis of the institutionalization of rituals and works in every community even though social practices in general are culturally characterized. Status Function apart, there are other two basic notions that occur in the explanation of successful functioning and stability of social institutions. The first is “cooperation” as a “strong” form of CI and the second is “collective recognition” as a “weak form” of it. We think that these two forms of intentionality correspond to the notion of “flexibility”, which imply the voluntary control over our actions and to the notion of “rigidity”, which characterize the mere following rules in the sense of routinely behavior.

A very famous example of a ritual (Searle often refers to) is “marriage”. First, we need to be moved to act in a certain way. We-Intentionality works when we want to do something together (we have a collective intention) so that we can cooperate to achieve our common goal. As we already anticipated, CI presents a weak form (collective recognition) and a strong form (cooperation). Both are crucial for rituals, in our case marriage. Now we can see how a social transition from one status to another is performed through an institutionalized ritual:

•We have "collective recognition", which means that the couple simply accepts the institution of marriage prior to actually getting married.

•But, the actual marriage ceremony is an example of active cooperation, in which the couple enters in a new social situation acquiring new social statuses consequently.

• This fact obtains by the performance of the speech act of promise.

•The social context requires also the speech act of declaration from the part of the institutional figure who has the suitable deontic powers to celebrate the rite and to ascribe the new status to the couple.

4 CONCLUSION

We propose to consider the notion of "habitual behavior" and its dimensions of "routine" and "goal-directed behavior" as a bridge between individual and social intentions. A research that crosses philosophy and neuroscience/neurobiology could explain the functioning of habits and can extend to the social sphere of rituals and their function in individual and interpersonal contexts. The example of a ritual like marriage shows that we need to routinely follow the procedure and, at the same time, to actively cooperate in a joint activity to create a new social situation.

REFERENCES

[1] M. Jankovic and K. Ludwig, *The Routledge Handbook of Collective Intentionality,* Introduction, Published online on: 31 Oct 2017.

[2] R. Giovagnoli, Habits and Rituals, in *Proceedings MDPI* of the IS4SI 2017 Summit, Gothenburg, 2017.

[3] R. Giovagnoli, From Habits to Rituals: Rituals as Social Habits, in *Open Information Science De Gruyter*, v.2, Issue 1 (2018).

[4] R. Giovagnoli, Habits, We-intentionality and Rituals in *Proceedings MDPI* of the IS4SI 2019 Summit, Berkeley 2019.

[5] R. Giovagnoli., *From Habits to Rituals: Rituals as Social Habits* in R. Giovagnoli and R. Lowe (Eds.), *The Logic of Social Practices*, Springer, Sapere, Cham, 2020, pp. 185-199.

[6] J. Lombo and J. Gimenez-Amaya, The unity and stability of human behavior. An interdisciplinary approach to habits between philosophy and neuroscience, Frontiers in *Human Neuroscience*, (8), 2017.

[7] A. Graybiel, Habits, Rituals and the Evaluative Brain, *Annual Review of Neuroscience*, /31), 2008, pp. 359-87.

[8] J. Bernacer and J.I. Murillo, The Aristotelian Conception of Habit and Its Contribution to Human Neuroscience. *Frontiers in Human Neuroscience*, (8), 2014.

[9] R.Lowe, Learning and Adaptation: Neural and Behavioral Mechanisms behind Behavior Change, *Connection Science*, vol. 50, Issue 1, 2018.

[10] C. Castelfranchi and G. Pezzulo, Thinking as the Control of Imagination: a Conceptual Framework for Goal-directed Systems., *Psychological Research* (73), (4), 559-577, 2009.

[11] M. Iacoboni et al., *Grasping the Intentions of Others with One's Own Mirror neuron system*. PLOS Biology, 2005.

[12] G. Pezzulo et al, *The Challenge of Anticipation*, Springer, Basel, 2008.

[13] G. Pezzulo and C. Castelfranchi, The Symbol Detachment Problem, *Cognitive Processing*, 8 (2), 2007.

[14] J. Searle, *Making the Social World*, Oxford, Oxford University Press, 2010.